Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2022)
IEEE Xplore Part Number: CFP22N67-ART; ISBN: 978-1-6654-9707-7

2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA) | 978-1-6654-9707-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIRCA54612.2022.9985728

IoT based Ocean Acidification monitoring system with ML based Edge Analytics

Gopika K V*, Nitin Kumar M*, Maneesha Vinodini Ramesh*,
Center for Wireless Networks & Applications (WNA), Amrita Vishwa Vidyapeetham, Amritapuri,India
*gopikakadengal@gmail.com
*nitinkm@am.amrita.edu
*maneesha@amrita.edu

***Abstract*—Ocean Acidification (OA) is often referred to as "an evil twin of climate change". Increased global atmospheric emission of CO_2 has increased its concentration in the ocean due to dissolution and absorption of atmospheric CO_2, leading to ocean acidification. This further affects the coral reefs and other calcifiers such as Pteropods (marine snails), shellfish (clams, etc.), crustaceans (shrimp and lobsters), starfish, sea urchins, and their kin (echinoderms) and interferes with the marine food web and leads to food shortages thereby affecting the socio-economic lives of the people living in the coastal area. To quantify the above impacts and to study other biochemical variations in the ocean, it is necessary to monitor ocean acidification for a longer time period. Current methods for monitoring ocean acidification employ gliders, drifters, buoys, mooring, and periodic testing of samples collected during research voyages. These methods are often not cost-effective, time-efficient, requires a lot of manpower, and provides inferior spatio-temporal resolution. This research work analyzes the existing ocean acidification data collected from the ALOHA station at the North Pacific Ocean and propose a predictive model at the edge level by using Regression to reduce the effective cost of the IoT system. The proposed regression model is also used to design a novel IoT architecture for designing an energy-efficient, near real-time ocean acidification monitoring model. The proposed IoT hardware can be deployed by using buoys and fishing boats to further reduce the monitoring cost and increase the spatio-temporal resolution of the data compared to the existing monitoring systems.**

***Index Terms*—Machine learning,Internet of Things(IoT), Edge Computing, Ocean acidification, Near-Real-time monitoring, Blue economy, Sustainable development Goal(SDG)**

I. Introduction

Unsustainable practices, such as rapid urbanization, have led to a rapid increase in atmospheric CO_2. Out of the total emissions into the atmosphere, roughly 30% of carbon dioxide (CO_2) is absorbed by the ocean [1] [2]. This dissolution of atmospheric CO_2 into the ocean increases the concentration of H+ ions, thereby decreasing the pH level. As the pH level decreases which is called as ocean acidification(OA) [3], it will be difficult for calcifying organisms to build their shells due to the reducing amount of calcium carbonate saturation in oceanic water.

Ocean acidification will also affect the plants and other living organisms such as zooplankton and phytoplankton in the ocean. This will interrupt the food web interface such that some of the species will become abundant and others will even go extinct. For example, seagrasses will grow faster, while the oyster population may reduce. In that way, the entire life cycle or food chain will be affected, causing huge impacts on the marine ecosystem. This will further affect the people living in the coastal area who depend upon the ocean as a source of income and the "Blue economy" as a whole.

The Sustainable Development Goal (SDG) set by United Nations in 2015, in particular, SDG 14 "Life below water", specifically addresses the problem of ocean acidification under target 14.3 [4]. To meet this target, we need more and more sampling and observation centers in ocean. Considering the cost of building and running such stations, it might not be feasible to build and collect such large spatial data for long-term continuous monitoring.

So in this paper, we are going to address the drawbacks of these existing methods of ocean acidification data collection and propose a design for a near real-time, energy-efficient IoT based monitoring station for continuous ocean acidification data collection. This system also features edge intelligence to optimise the number of sensors used and to reduce the overall cost of the system.

II. Related Work

In this section we review the existing methods for collection of ocean acidification data and we also enumerate their drawbacks.

A. Existing Methods for Ocean Acidification Monitoring

- Laboratory Testing [5]: Spectrophotometric method is used for measuring pH of oceanic water in laboratory. In this method, water samples collected from the ocean using Niskin bottles are analyzed for pH by observing color changes after adding reagent. Spectrophotometric pH is determined using m-cresol purple or phenol red as an indicator. The resulting hue (absorbance) is measured at certain wavelengths. The main drawbacks of laboratory testing are the sample collection is needed to be performed manually, therefore there is a chance of associated human errors and real time data collection is not possible.
- Research Cruises [6]: Research Cruises are dedicated to data collection over relatively shorter periods hence data collection for a longer duration will be difficult. Victoire M.C. Re rolle et al [6] proposed a system where a simple microfluidic design (Semiconductor based fluorescence

method) is integrated into a shipboard instrument. A photodetector and a CMOS color camera were used for this carbonate system measurements. The results were highly precise, low consumption of power and reagents is also an advantage of this system. Still this system is complicated and expensive, especially in the case of optical components. Also long-term precision of this microfluidic pH systems has not yet been evaluated.

- Ships of Opportunity (SOPs) [7]: Ships of Opportunity/Ferrybox is a combination of volunteer commercial and research vessels to collect data on physical, chemical, and biological oceanography. W. Petersen et al [6] have used a ferry box for water quality measurements where the system is controlled by an internal computer using GPS locations. The ferry box is connected to the shore station with the use of GSM for data transfer and remote control. A ferry box is a cost-effective solution. The system enables high-frequency monitoring of oceanographic parameters including temperature, salinity, and turbidity as well as biologically relevant factors like chlorophyll, nutrients, oxygen, and pH along a transect. As maintenance of inline senors can be performed whenever required, data collected using this technique is highly reliable. But for long-term monitoring using Ferrybox, the changes in ship's route, makes it impossible to continuously monitor the required study area.
  Moreover on ferries, it's easier to handle and operate the system but in the case of cargo ships, changes in routes of ships, requires transfer of instruments between the ships which increases the cost of operation. Skilled manpower is required for initial installation, which also increases the cost of these systems.
- Wave Glider [8]: Measurement systems can be mounted over the remotely controlled surfboard whose motion is driven by wave energy, this surfboard is called Wave Glider. Compared to SOPs/ferrybox systems, Wave Gliders allow continuous monitoring of the study area, as they are dedicated devices for collection of oceanic data. But, this measurement scheme does not allow monitoring at different depth and cost of the instruments are normally high in comparison to other methods like laboratory testing or ferrybox systems.
- Underwater Glider [9]: The underwater glider is an autonomous underwater vehicle that uses a buoyancy engine. By deploying instruments on an underwater glider, data collection over V-shaped depth profiles can be performed up to 1500 meters. However, the cost of the monitoring system can be very high in this method compared to previous methods. Grace K. Saba et al. [9] used a SlocumG2 Glider capable of diving to 1000 meters with Deep-Sea ISFET (Ion Sensitive Field Effect Transistor) pH sensor coupled with CTD(Conductivity, Temperature, Depth) sensor and optical sensors like spectral backscatter, chlorophyll fluorescence, and dissolved oxygen (DO)sensor. The data collected was then used to map pH parameters against other parameters. The Gliders can also be used in challenging environments, including upwelling zones, bays, areas with high riverine and/or eutrophication influence, and near ice shelves in the Antarctic, beneath hurricanes, coastal storms, and on river-dominated continental shelves. The comparison of the true pH measured using the spectrophotometric method against the measurements made by the glider shows that sampling below specific target depth causes a significant mismatch between glider pH and spectrophotometric pH measurements. Yuichiro Takeshita et al. [10] deployed a Spray underwater glider integrated with the Deep-Sea-DuraFET(DSD) for pH data collection. The spatial accuracy of both the gliders used in [9] and [10] deteriorates once it dives below water due to a lack of GPS positioning. This result in erroneous spatial coordinates associated with the pH readings.
- Measurement schemes used generally for monitoring ocean acidification are discussed above. Other monitoring schemes that can be considered for studying the ocean acidification are listed below: -
  - ∗ Buoys [11] [12]: It is a floating platform that can be anchored to the seabed or allowed to drift with ocean currents. Instruments can be mounted over this and used to monitor ocean data with low spatial resolution. Buoys can be affected by biofouling.
  - ∗ Vertical Profilers [13] [14] [15]:It is an instrument that can monitor the entire water column at a specific position but providing wide-area monitoring can be costly due to the large number of profiles required.
  - ∗ Mooring: The moored profiler is attached to a subsurface mooring cable that can run from 50 meters down to the seafloor at 5,000 meters or more. Typical mooring instrumentation is usually located at discrete depths and subject to fouling. This can generate time-series data sets that are incomplete due to a limited number of sensors or erroneous data from fouled sensors.
  - ∗ Drifters: Freely-floating drifting buoys. Sensor accuracy can be compromised due to fouling.

B. *Drawback of Previous Method*

The drawbacks for the previous methods for monitoring of ocean acidification, as evident from the above literature are:-

- ∗ The lack of enough data on ocean acidification from continuous monitoring systems.
- ∗ The lack of data with high enough temporal and spatial resolution.
- ∗ Most ocean data collection methods such as buoys are often affected by biofouling.
- ∗ Operational cost of ships used in research cruises are very high
- ∗ Ferrybox systems can only be fitted on boats that follow a fixed path.
- ∗ Long term evaluation of certain new sensors used for ocean acidification measurements such as micro-fluidic system is yet to be performed.

* Gliders used for the collection of ocean acidification data is costly and hence not accessible to everyone.
* Sensor on the glider shows a variation in the accuracy when measured over the targeted depth.

III. Data Collection

For developing a new IoT system for monitoring of ocean acidification, we have collected existing ocean acidification data from the North Pacific Station (ALOHA). This station have been measuring the chemical and biological parameters of seawater once a month since October 1988, and the data collection spans 30 years [16]. The temperature and salinity measurements are taken from a calibrated Seabird CTD sensor. Phosphate and silicate were measured colorimetrically via an auto analyzer. Dissolved inorganic carbon (DIC) is equivalent to total CO_2 and total alkalinity (TA). Coulometry and open-cell acid titration were used to measure these parameters. During the early years of the time series, a few missing DIC and TA data points were replaced using values from C. D. Keeling's parallel Station ALOHA CO_2 measuring experiment.

IV. Statistical Analysis of Data

Data collected from the previous study need to be preprocessed prior to analysis. Preprocessing involve the following steps:-

- Conversion of data types in collected data to standard data types required by the data analysis tool.
- Collected datasets has data points with missing value, which can have a significant impact during data analysis. Hence, these data points were filled with linear interpolation.

A. Inter-relationship study

Inter-relationship study explains the distribution of data and reveals the underlying correlations between various types of data. This study also helps in finding the confounding data that can be selected as the best features for a specific Machine Learning (ML) application development. For doing this, the parameters in the collected data were plotted using seaborn library in python as shown in figure1. The plots along with the Pearson Correlation matrix shown in figure2 convey that pH and CO_2 have a strong negative relationship with a coefficient of -0.99 ,then to the temperature with a correlation of -0.6 then, with salinity of 0.13, and then to DIC with -0.41. Here correlation coefficient is a valuable numerical measure of association between two variables, which varies in value between -1 and 1. Here we are considering pH, pCO_2, temperature, and salinity for the further investigation. Fitting a regression model to the data will be very useful. Here pH, pCO_2 and DIC shows a Normal distribution, while the distribution of salinity and DIC is right-skewed. The summary of all the 18 parameters that are related to ocean acidification and their patterns are also summarized in the correlation matrix.

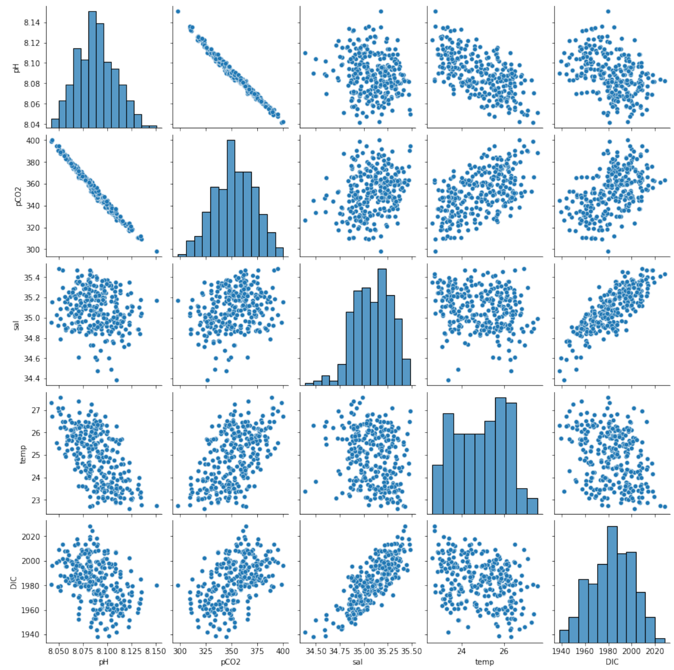

Fig. 1. Pairplot between different parameters

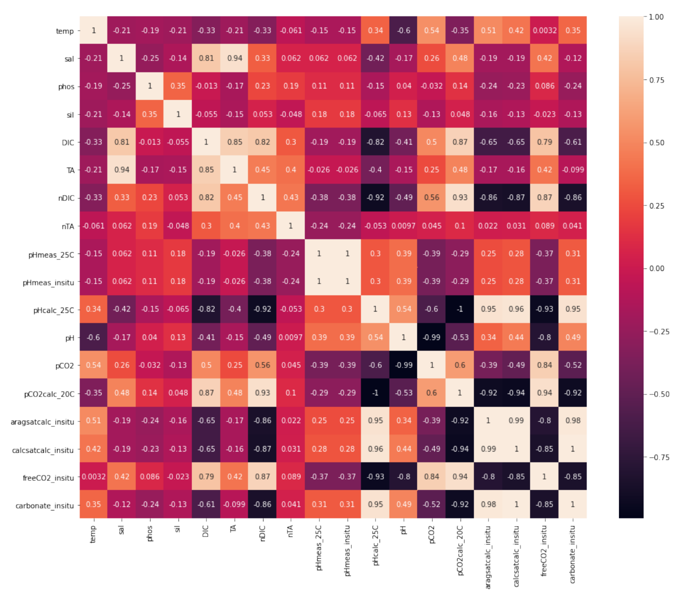

Fig. 2. Correlation matrix

B. Machine learning model for Predicting pCO_2

Machine learning(ML) model, polynomial regression, is used to correlate the output with input features. Polynomial regression involves converting the original characteristics into polynomial features with the necessary degree, after which a linear model is used to model the data [17]. The steps followed for deriving this ML model is shown in fig 3.

For the purpose of training, the dataset has been split into two: the *training set* which is used for training and the *test set* which is used for testing, in the ratio $\frac{2}{3}$: $\frac{1}{3}$. The ML model

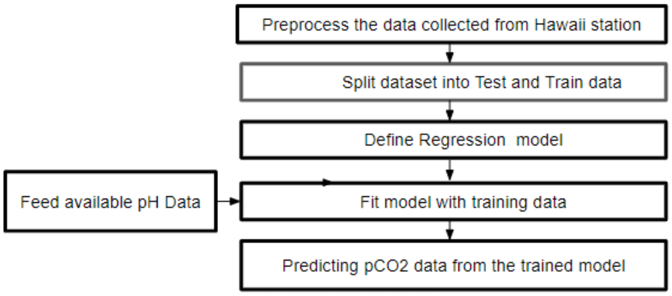

Fig. 3. Regression model

was then trained for various degrees of polynomial regression and also by selecting various input features.

C. Performance Measures

Performance evaluation, often known as evaluation metrics, is a metric for determining how effectively a model performs. These methods allow us to determine the accuracy and precision with which the ML model makes predictions. Here we have used MSE(Mean Squared Error) and MAE(Mean Absolute Error) for assessment of the ML model as shown below 1 and 2. Here y_i and $\hat{y}_i$ are original and predicted values in data set.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(\hat{y}_i - y_i)| \tag{2}$$

TABLE I
PERFORMANCE OF REGRESSION MODEL ON DIFFERENT FEATURES

Parameters used as input	R2 Score	MSE Score
pH	0.9832095191633299	5.49248221181907
pH and Temperature	0.9919623240742125	2.6450053668477693
pH, Temperature and DIC	0.9989694630532812	0.33912486394992264
pH ,Temperature ,DIC and Salinity	0.9989705934231068	0.3387528864923934

We can observe that as we increase the number of input parameters for the ML model, the prediction accuracy increases. This is observed using the changes in R2 score and MSE values in table I. Highest R2 score and lowest MSE score is observed, when we use all four input parameters, i.e, pH, Temperature, DIC and Salinity.

Further the degree of polynomial regression is also varied from 1^{st} order to 10^{th} order and the results of the experiments are shown in table II, III, IV and V.

TABLE II
PERFORMANCE OF REGRESSION MODEL ON TEST DATA SET USING PH ONLY

Regression Using pH	MSE	MAE
Linear	5.451120	1.901273
Degree 2	4.793997	1.726850
Degree 3	4.925111	1.784921
Degree 4	4.942551	1.788387
Degree 5	4.977883	1.802906
Degree 8	4.977103	1.802923
Degree 10	4.976967	1.802843

TABLE III
PERFORMANCE OF REGRESSION MODEL ON TEST DATA SET USING PH AND TEMPERATURE ONLY

Regression using pH and Temperature	MSE	MAE
Linear	2.603810	1.238336
Degree 2	2.349839	1.195376
Degree 3	2.214801	1.155596
Degree 4	2.238442	1.159221
Degree 5	2.135292	1.146120
Degree 8	2.413379	1.234840
Degree 10	2.194053	1.175982

TABLE IV
PERFORMANCE OF REGRESSION MODEL ON TEST DATA SET USING PH ,TEMPERATURE AND DIC

Regression using pH, Temperature and DIC	MSE	MAE
Linear	0.332504	0.449670
Degree 2	0.026554	0.128685
Degree 3	0.029315	0.135195
Degree 4	0.030989	0.141657
Degree 5	0.031010	0.142418
Degree 8	0.031193	0.142776
Degree 10	0.035902	0.151447

TABLE V
PERFORMANCE OF REGRESSION MODEL ON TEST DATA SET USING PH ,TEMPERATURE AND DIC, SALINITY

Regression using pH, Temperature , DIC and Salinity	MSE	MAE
Linear	0.306344	0.4412620
Degree 2	0.010433	0.049662
Degree 3	0.010456	0.049663
Degree 4	0.0107126	0.0526833
Degree 5	0.0125793	0.0616495
Degree 8	0.0130005	0.0623807
Degree 10	0.01399313	0.0666271

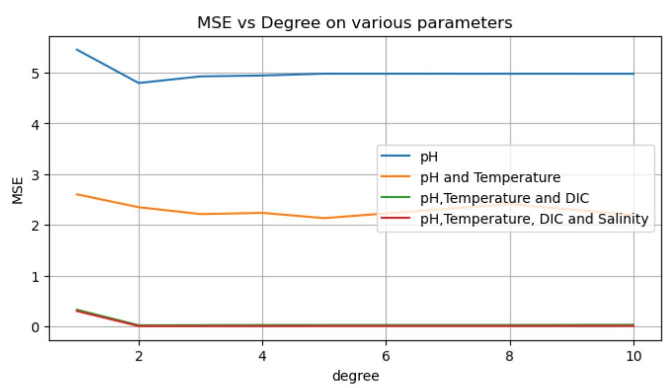

Fig. 4. Performance analysis on model using MSE Score

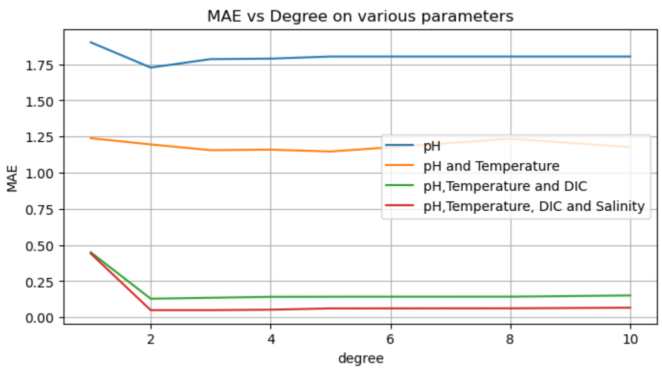

Fig. 5. Performance analysis on model using MAE Score

D. Selected ML model for developing IoT system

From the above polynomial regression models with various degree and using various parameter, the model with best prediction accuracy is obtaining when all four features are used at degree two which has least MSE and MAE values compared to other models. Hence this model is selected for development of the IoT system from the above tables II, III, IV, V, fig.4 and fig.5

V. Proposed IoT Architecture

The proposed IoT system for monitoring/collection of Ocean Acidification data can be sub-divided into various functional layers. These layers are **(a)***Sensing Layer*, **(b)***Network Layer*, **(c)***Middleware Layer*, **(d)***Application Layer*, and **(e)***Business Layer*. These layers have been arranged as per the flow of data between the various layers and are shown in fig.6.

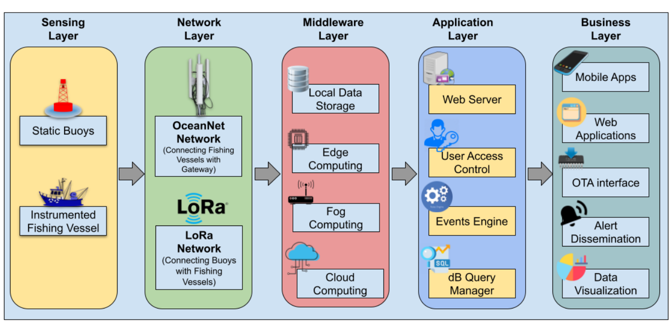

Fig. 6. Various Layers of IoT Architecture for Ocean Acidification Monitoring

(a) *Sensing Layer*:- The sensing layer consists of a network of *static buoys* and dynamically moving *instrumented fishing vessels*. The main function of the sensing layer is to sense and gather information about the environment. The key sensors employed in the sensing layer to collect the data are pH, Temperature, DIC and Conductivity. The data collected by this layer is further augmented with temporal and spatial references using RTC and GPS data. Some boats also have pCO_2 sensor to validate the output of the ML model that generate the pCO_2 value on other boats and buoys .The sensor in the sensing layer also has a mechanism to prevent biofouling to obtain lower maintenance cost. The biofouling mechanism used in these sensors can be found in [8], [12]

(b) *Network Layer*:- This layer is responsible for connecting static buoys, instrumented fishing vessels and servers. Network layer uses *LoRa* and *long-range WiFi* to accomplish the connectivity. The long-range Wifi connectivity service is provided by the OceanNet project. OceanNet is a cost-effective communication system for marine Internet connectivity that provides a guaranteed 45+ km range from the shore, which can be further extended opportunistically by using a multi-level point-to-multi-point (P2MP) network using long-range Wi-Fi technology. Fishing vessels will be connected to this oceanNet network [18]. In the network layer buoys are stationary units and fishing vessels are roving units. The data transfer from buoys to fishing vessels happens through low power LoRa network and data transfer from fishing vessel to gateway (shore) happens through OceanNet network. Both the units perform temporal data collection. The use of LoRa facilitates data transfer with less propagation effects since the ocean environment has the least obstacles and line of sight communication. The nodes in the network can be leaf nodes, relay nodes or cluster heads, where leaf nodes will collect data, transfer the data to the Relay node or Cluster head (CH) and cluster head (CH) will send it to the Base station as we can see in Fig 7. This is a delay-tolerant and opportunistic network. The various modes of data transfer from leaf node to cloud server it depended upon its spatial location and availability of opportunities to connect, as shown below:-

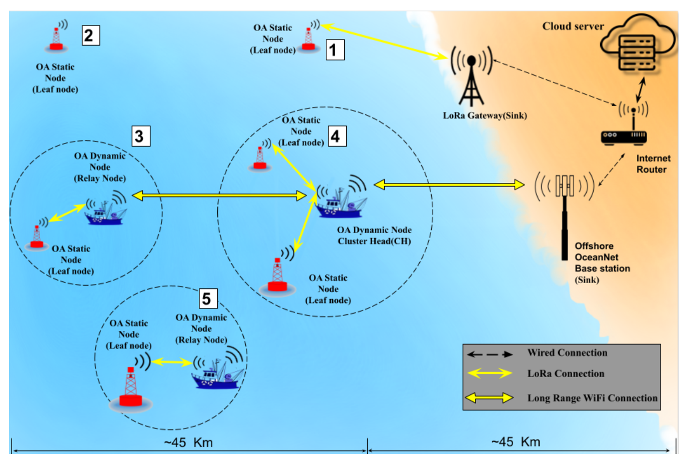

Fig. 7. Proposed Delay-Tolerant IoT Network Topology.

1) Mode 1: Direct transfer of data from leaf nodes(buoy) to LoRa gateway.
2) Mode 2: Leaf node(buoy) has collected data and is waiting for an opportunity to forward the data to cloud server based on availability of nearby relay node(boat).
3) Mode 3: Leaf node(buoy) is forwarding all collected data via a relay node(boat) to a cluster head(boat) and then to cloud server via offshore base station.

4) Mode 4: Cluster Head collecting the data from nearby leaf nodes(buoys) and sending it to cloud server.
5) Mode 5: Leaf node(buoy) is transferring collected data to relay node(boat) through LoRa connection but relay node(boat) is not in the coverage of Long range Wi-Fi connection. Hence data are stored in the relay node(boat) itself and forward to the server whenever connectivity is obtained.

Antenna Toolbox in MATLAB R2021a was used to calculate the coverage of the Gateway antenna used to collect data from the Buoys and Fishing Vessels. During the MATLAB Simulation, following parameters were selected for transmitter gateway: dipole Yagi-Uda antenna design, transmitting power of 21 dBm, antenna height of 56 meter and transmitter frequency of 2402.5 MHz. While the parameters selected for receiver node: dipole antenna design, antenna height of 9 meter, receiver sensitivity of -105 dBm. The output of this simulation is shown in fig 8.

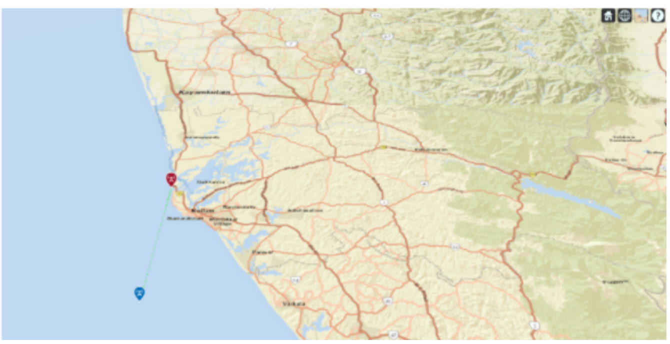

Fig. 8. Coverage of Communication module using MATLAB R2021a – Antenna Toolbox

(c) *Middleware Layer*:- The middleware layer consist of three sublayers, i.e Edge Computing layer, Fog Computing layer and Cloud computing layer. All the three layers also have option for data storage.

- The Edge Computing Sub-Layer is responsible for real-time data processing on device premises. Basic data analytics such as data filtering, data packet size Optimization and other algorithms that require low computing power come under this sublayer [19]. This will save network bandwidth and extend operational time of the edge node by conserving the battery power [20]. This layer is also responsible for generation of pCO2 value in leaf nodes using the above discussed ML model.
- The Fog Computing Sub-Layer is responsible for Data caching, Buffering, and data filtering. This sublayer also performs data aggregation, data summarization and fog level data analytics such as health monitoring of leaf nodes and cluster heads and checking of data thresholds such as pH and temperatures. The fog computing sublayer is physically implemented on fishing vessels and gateway hardware. The local storage of data in fog layer can reduce the frequent communication with cloud and also reduce data transmission overhead [21].
- The Cloud Computing Sub-Layer is responsible for central control, and central processing of the data. It also facilitates massive and scalable long-term storage and computing power, to meet the demands of the future. The cloud computing infrastructure is hosted on massive servers either self-hosted or leased as per requirement.

(d) *Application Layer*: The application layer consist of applications and services to support the business layer. The application layer host the following: (a) web server-used to provide browser access to data in the business layer as HTML pages. (b)user access control-stores the user privileges to data for various stakeholders in the business layer. (c) events engine -contains the set of actions to be takes under various scenarios such as high pH, high temperature, failure of cluster node etc., (d) dB-query manager-runs all the data related queries on the cloud database, keeps a record of all historic queries, runs scheduled data queries for data summarization etc.

(e) *Business Layer*: The business layer consist of applications responsible for accessing the collected data and visualize the data. The business layer also contain the mobile application and web application that run using the services provided by the application layer. All other layers of the IoT architecture are accessed via business layer for purposes such as system wide management, Edge device reprogramming via OTA(over-the-air), alert dissemination to various stakeholders of events(failure of leaf node, high pH detection, high temperature detection, etc.).

VI. Subsystems of Edge Device

This section elaborates on the various subsystems inside the edge devices. These subsystems are enumerated below:-

A. Sensor Subsystem

Sensor Subsystem consist of various sensors installed inside the buoys and fishing vessels to collect the data. The sensor used in the current design are CO2-Pro™ Submersible CO_2 Sensor, Deep SeapHOx™ V2 Ocean CT(D)-pH-DO Sensor and DIC sensor. The sensor help the system to measure pH, CO_2, DO, temperature, conductivity, salinity and dissolved-inorganic-carbon. GPS and RTC are also part of the Sensor subsystem and adds spatial and temporal metadata. The various modules in the sensor system are interconnected as shown in fig 9. The proposed solution on the fishing vessel has a scope for depthwise vertical profiling of oceanic water during night time as fishing vessels are anchored at that time.

B. Power Subsystem

Power Subsystem consist of various entities used to energize other subsystems in the leaf node. Solar cells are used as the main power source in leaf nodes. Solar power was selected to energize the leaf node due to it's abundant availability in leaf node locations. Additional solar power is also stored in a lead acid batteries for reliable power supply. A charge controller is used to recharge the batteries using solar power. Voltage regulator with regulated output of 3.3V and 12V is utilized to power the sensors and communication subsystems. The various

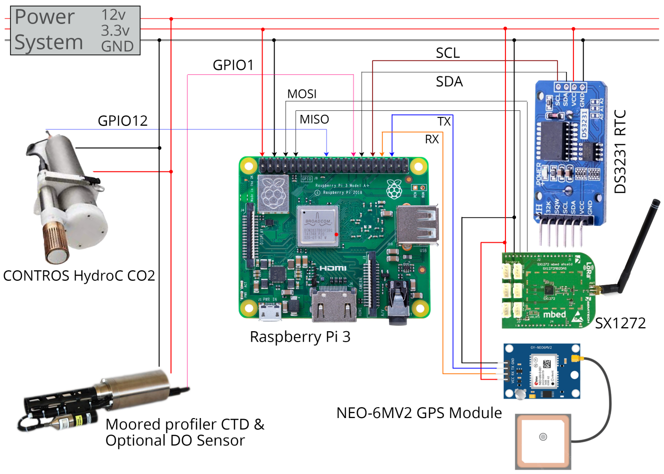

Fig. 9. Sensor subsystem

modules in the power system are interconnected as shown in fig 10.

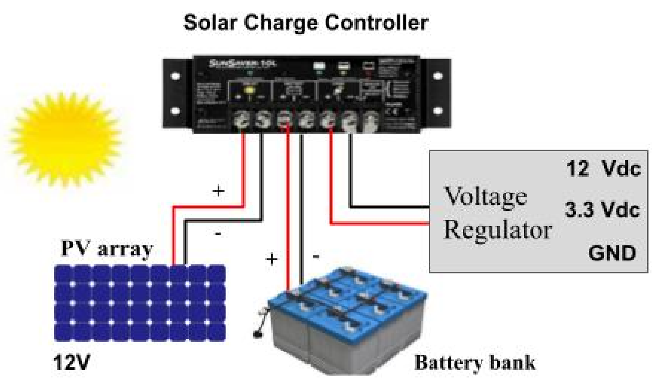

Fig. 10. Power subsystem

C. Communication Subsystem

The communication Subsystem consist of LoRa module and Long range WiFi module. On buoys only the LoRa module is installed, where as on the boats both the LoRa and long range WiFi module is implemented. The buoys transmit the data to the nearby boats as per there availability over LoRa network. The data transmission is based on the *store and forward algorithm*. This algorithm stores all the sampled data and forwards it into the cloud based on availability of boats.

D. Data-Driven Decision Making, pCO_2 prediction, Data-aware Sensing Using Edge Computing

The flow of data between the various blocks inside the edge node is shown in 11. The *Sensors* block is responsible for collection of various parameters, i.e., pH, Temperature, pCO_2 Salinity and DIC. The data collecting from *Sensors* block needs some level of processing, which is performed in the *Processing* block. This processing helps us to remove less quality data and redundant data to achieve lower processing power, lower storage space and save bandwidth [22]. The data is then forwarded into the *Storage* block and *Analytics* block. The *Storage* block is responsible for local data storage, data packet preparation and data packet optimization. The *Analytics* block is responsible for decision making and contains sub-blocks such as pCO2 prediction, threshold verification and data classification. The *communication* blocks consist of Long range Wi-Fi and LoRa , this will finally sent the prepared data packets to the cloud server.

Decision-making system is implemented within data classification sub-block. This system can decide which is relevant and which is irrelevant data. Therefore we can discard irrelevant ones and save bandwidth as well as storage space, that will help for long-term monitoring. The decision-making system can decide whenever the sensors should be ON according to temperature data therefore we can achieve energy efficiency also it can predict value of pCO_2.

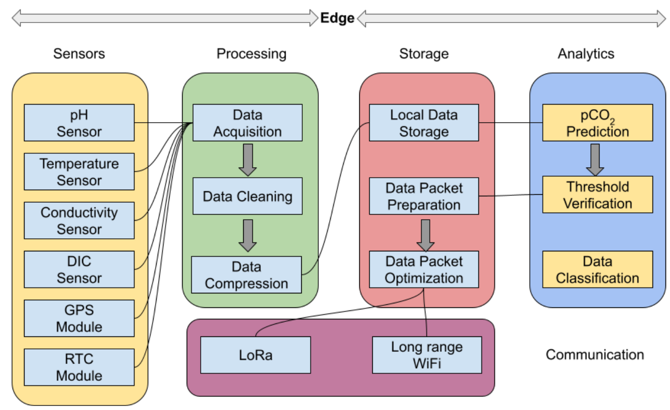

Fig. 11. Data Flow Diagram and Decision-Making at Edge

VII. FEATURES OF PROPOSED MONITORING SCHEME

Features of proposed system are listed below.

- The proposed system uses lesser number of sensor by implementing edge analytics, thereby reducing overall system cost.
- The proposed system uses delay tolerant opportunistic network architecture for collection of data.
- The proposed systems design has low energy consumption due to feature such as context aware energy management, delay-tolerant network, edge data analytics(removal of redundant data and lower quality data), data packet optimization, etc.
- The proposed system architecture can withstand biofouling and other ocean conditions that can affect accurate measurements of OA data
- The system uses hybrid network architecture using LoRa and Long range Wi-Fi rather than depending upon costly satellite networks such iridium
- The proposed monitoring system is based on IoT architecture and has features such as scalability and repeatablity
- The low cost feature of the proposed system helps in better monitoring of shelf water ecosystem which is

dominated by river discharge. This also helps in implementation of marine protected areas(MPA)
- The proposed system also achieves better spatio-temporal resolution using a network of buoys and instrumented boats

TABLE VI
COMPARING EXISTING MODEL WITH PROPOSED MODEL

Existing Monitoring system	Proposed Monitoring System
Short term monitoring	Long term monitoring
The data are collected in a difference of seconds or hours	Delay tolerant
Existing systems have not designed a system with energy efficiency	Energy efficient
Sensors and other related equipments are affected by biofouling	Sensors are less affected by biofouling
The cost of the system is high	Cost is comparatively low

In table VI we compare existing monitoring system with proposed monitoring system. In the existing monitoring systems, the experiments are of short time period, because of factors such as capital and operational cost, damages of system as part of the bio-fouling etc. In existing OA monitoring systems, data sampling is very frequent, it collects data with an interval of one second or one hour. But the rate of change of pH value is not that frequent. Hence we don't need a real time system that requires a high operational cost. Considering these, we have selected a delay tolerant architecture in the proposed system. Since the existing systems perform real time operations, the energy consumption is high. Since our proposed architecture uses a delay tolerant network mechanism and also context aware energy management, we can reduce the overall energy consumption of the system. Biofouling will cause surface growth of algae and growth of barnacles around the buoy. In the proposed system we have included the sensors that have self cleaning mechanism to prevent growth of algae and also special coatings are provided to prevent growth of barnacles. Since the current monitoring systems are based on satellite communication, the overall system is not at all cost effective in comparison to proposed system.

VIII. Conclusion and Future works

In this paper we have shown that monitoring of ocean acidification using new emergent technologies such as smart sensors, wireless network and IoT have many advantages over traditional methods. Moreover we have also shown that we can use the existing OA data to train ML/AI models that are able to make predictions on useful parameters such as pCO2 value using other parameters, thereby reducing the requirement of additional sensors. The paper explores the possibility of deployment of such models at the edge node to reduce the cost of OA monitoring system. The paper also propose and showcase an IoT system with various features that are specifically adaptable to the monitoring of OA, such as extending the battery life using adaptive data sampling and delay-tolerant, opportunistic networks. Though system for monitoring of global ocean acidification are available, they are not sufficient in monitoring of continental shelf-water that shows high variability in oceanic parameters. Hence the low cost OA monitoring system proposed in this paper can ensure proper data collection and protection of organisms in shelf-water. This would also ensure that the economy and lives of the people dependent on these marine resources are also safeguarded.

As future works we will be implementing the proposed system towards Kerala in India where the OceanNet network is already installed . During this implementation we will be conducting extensive sea trials to bring improvements to the proposed OA monitoring system.

IX. Abbreviations in the paper

∗ temp : The mean in situ seawater temperature, in °C.
∗ sal: The mean seawater salinity, in practical salinity units. phos The mean seawater phosphate concentration, in μmol kg-1. sil The mean seawater silicate concentration, in μmol kg-1.
∗ DIC: The mean seawater dissolved inorganic carbon (= total CO2) concentration, in μmol kg-1.
∗ TA: The mean seawater total alkalinity, in μeq kg-1.
∗ nDIC: The mean seawater salinity-normalized DIC, in μmol kg-1 at salinity = 35.
∗ nTA: The mean seawater salinity-normalized TA, in μeq kg-1 at salinity = 35.
∗ pHmeas_25C: The mean measured seawater pH at 25 °C, on the total scale.
∗ pHmeas_insitu: The mean measured seawater pH, adjusted to in situ temperature, on the total scale.
∗ pHcalc_25C: The mean seawater pH, calculated from DIC and TA at 25 °C, on the total scale.
∗ pHcalc_insitu: The mean seawater pH, calculated from DIC and TA at in situ temperature, on the total scale
∗ pCO_2calc_insitu The mean seawater CO2 partial pressure, in μatm, calculated from DIC and TA at in situ temperature.
∗ pCO_2calc_20C: The mean seawater CO2 partial pressure, in μatm, calculated from DIC and TA at 20 °C.
∗ aragsatcalc_insitu: The mean seawater aragonite saturation state (solubility ratio).
∗ calcsatcalc_insitu: The mean seawater calcite saturation state (solubility ratio).
∗ $freeCO_2$_insitu: The mean seawater free CO2 concentration, in μmol kg-1.
∗ carbonate_insitu: The mean seawater carbonate ion concentration, in μmol kg-1.

X. ACKNOWLEDGMENT

I would like to express gratitude for the immense amount of motivation provided by our chancellor Sri. Mata Amritanandamayi Devi, Amrita University and we also thank the faculty and staff members of our research center (Center for Wireless Networks &, Applications, Amrita Vishwa Vidyapeetham) for their support in carrying out this work. Further we would like to thank Dore, J.E., R. Lukas, D.W. Sadler, M.J. Church, and D.M. Karl of Hawaii Ocean Time-series (HOT) project in the ALOHA station for openly sharing the data required for this study.

REFERENCES

[1] G. N. Plass, “The carbon dioxide theory of climatic change,” *Tellus*, vol. 8, no. 2, pp. 140–154, 1956.

[2] R. A. Feely, C. L. Sabine, T. Takahashi, R. Wanninkhof *et al.*, “Uptake and storage of carbon dioxide in the ocean: The global co˜ 2 survey,” *OCEANOGRAPHY-WASHINGTON DC-OCEANOGRAPHY SOCIETY-*, vol. 14, no. 4, pp. 18–32, 2001.

[3] P. Pillai and M. Supriya, “Real time co 2 monitoring and alert system based on wireless sensor networks.”

[4] A. Sturesson, N. Weitz, and Ã. Persson, “Sdg 14: Life below water,” *A Review of Research Needs Technical annex to the Format report ForskningfÃ¶ r Agenda*, vol. 2030, 2018.

[5] C. M. McGraw, C. E. Cornwall, M. R. Reid, K. I. Currie, C. D. Hepburn, P. Boyd, C. L. Hurd, and K. A. Hunter, “An automated ph-controlled culture system for laboratory-based ocean acidification experiments,” *Limnology and Oceanography: Methods*, vol. 8, no. 12, pp. 686–694, 2010.

[6] V. M. Rérolle, C. F. Floquet, M. C. Mowlem, D. P. Connelly, E. P. Achterberg, and R. R. Bellerby, “Seawater-ph measurements for ocean-acidification observations,” *TrAC Trends in Analytical Chemistry*, vol. 40, pp. 146–157, 2012.

[7] W. Petersen, M. Gehrung, and F. Schroeder, “Ferrybox-continuous and automatic water quality observations along transects in the north sea,” in *2008 IEEE/OES US/EU-Baltic International Symposium*. IEEE, 2008, pp. 1–4.

[8] “Deep SeapHOxTM Ocean CT(D)-pH-DO Sensor,” http://www.ems-ocean.com/catalogue2/SBE/SeapHOx(deep).pdf, note = Accessed: 2022-6-30.

[9] G. K. Saba, E. Wright-Fairbanks, T. N. Miles, B. Chen, W.-J. Cai, K. Wang, A. H. Barnard, C. W. Branham, and C. P. Jones, “Developing a profiling glider ph sensor for high resolution coastal ocean acidification monitoring,” in *OCEANS 2018 MTS/IEEE Charleston*. IEEE, 2018, pp. 1–8.

[10] Y. Takeshita, B. D. Jones, K. S. Johnson, F. P. Chavez, D. L. Rudnick, M. Blum, K. Conner, S. Jensen, J. S. Long, T. Maughan *et al.*, “Accurate ph and o 2 measurements from spray underwater gliders,” *Journal of Atmospheric and Oceanic Technology*, vol. 38, no. 2, pp. 181–195, 2021.

[11] A. Mukhopadhyay, S. Karunakaran, and K. Nidhi, “Transmission of multimedia data in underwater terrain using acoustic waves,” in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2021, pp. 597–602.

[12] “Contros HydroC CO_2,” https://www.4h-jena.de/en/maritime-technologies/sensors/hydrocrco2/, note = Accessed: 2022-6-30.

[13] G. Surendran, G. Udupa, and G. Nair, “Design and modelling of cable suspended sonde for water quality monitoring,” in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*. IEEE, 2017, pp. 1260–1265.

[14] Y.-G. Park, S. Seo, D. G. Kim, J. Noh, and H. M. Park, “Coastal observation using a vertical profiling system at the southern coast of korea,” *Frontiers in Marine Science*, vol. 8, p. 668733, 2021.

[15] S. Dong, J. Sprintall, S. T. Gille, and L. Talley, “Southern ocean mixed-layer depth from argo float profiles,” *Journal of Geophysical Research: Oceans*, vol. 113, no. C6, 2008.

[16] J. E. Dore, R. Lukas, D. W. Sadler, M. J. Church, and D. M. Karl, “Physical and biogeochemical modulation of ocean acidification in the central north pacific,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 30, pp. 12 235–12 240, 2009.

[17] G. Bonaccorso, *Machine Learning Algorithms: Popular algorithms for data science and machine learning*. Packt Publishing Ltd, 2018.

[18] A. Karthik, D. G. Koshy, L. Rajagopal, A. Luke, M. Meera, and N. S. Shibu, “Study and analysis of oceannet—marine internet service for fishermen,” pp. 1–8, 2017.

[19] S. Anand and M. V. Ramesh, “Multi-layer architecture and routing for internet of everything (ioe) in smart cities,” in *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. IEEE, 2021, pp. 411–416.

[20] N. Bhalaji, “Efficient and secure data utilization in mobile edge computing by data replication,” *Journal of ISMAC*, vol. 2, no. 01, pp. 1–12, 2020.

[21] S. Sreeraj, A. Unnikrishnan, K. Vishnu, N. E. Kennith, S. Anand, and M. V. Ramesh, “Empowerment of women self help groups: human centered design of a participatory iot solution,” in *2020 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 2020, pp. 1–8.

[22] D. Sivaganesan *et al.*, “Design and development ai-enabled edge computing for intelligent-iot applications,” *Journal of trends in Computer Science and Smart technology (TCSST)*, vol. 1, no. 02, pp. 84–94, 2019.