Engineering Innovations
ISSN: 2813-1002, Vol. 14, pp 85-98
doi:10.4028/p-XUQf31
© 2025 The Author(s). Published by Trans Tech Publications Ltd, Switzerland.
Submitted: 2024-11-29
Revised: 2025-01-14
Accepted: 2025-01-22
Online: 2025-04-04

Implementation of an IOT Sensor Network and Machine Learning to Measure the Water Quality

Jose Fernando Alarcón Madrid[1,a*], Owen Josue Paz Quintanilla[1,b*], and Martin Martinez-Rangel[2,c]

[1]Universidad Tecnológica Centroamericana, San Pedro Sula, Honduras

[2]Universidad Autonoma de Morelos, Cuernavaca, México
[a]fer.alarcon002@unitec.edu, [b]owenjosue09@unitec.edu, [c]mmtzr@uaem.mx

Keywords: Internet of Things, Machine Learning, Sensor Network, Water Quality

Abstract. Lagoons have a great importance for society, and activities such as fishing or tourism are essential for these areas, for this reason it is important to have a monitoring system in terms of water quality. The central axis of this project was the design and implementation of a sensor network based on the Internet of Things, collecting data using an ESP32 and the Thingspeak platform for data visualization and storage. Data is analyzed using MATLAB, allowing to obtain an estimation of the water quality index of Laguna Jucutuma indicating an average rating of 40, as well as using Machine Learning techniques to obtain a models with an error margin below 3%.

Introduction

Laguna Jucutuma is one of the most important wetland ecosystems for the region and for the country in general, being habitat for different species of flora and fauna, as well as economic activities such as fishing or tourism. However, in recent years it has been strongly affected by the pollution that surrounds it, either by construction in nearby areas, sewage drainage, forest fires or the cultivation of African palm.

Currently, it is common knowledge that lakes and lagoons are polluted daily. In contrast to this problem, it is important to use a network of sensors to measure variables such as: temperature, turbidity, pH, TDS or amount of dissolved solids and oxidation potential. These variables are sensed and sent to an online platform called ThingSpeak, where they will be stored and historical records can be observed in the form of graphs, in order to have an easier data analysis. Subsequently, these data will be analyzed using MATLAB and where a water quality index will be obtained to finish with the use of a form of Artificial Intelligence, for the prediction of useful data within the network.

One of the differential factors of the project is its low power consumption. Being designed and tested so that a 6W power panel is enough to energize the whole electronic circuit; using sensors that can be energized between 3.3V and 5V. In this way the prototype and subsequent buoy will be self-sufficient to have a continuous data measurement. With the implementation of long-range communication modules (LoRa), data collection can be done at any point within the range of the module, allowing an easy connection to the Internet for data visualization on the ThingSpeak platform.

Theorical Framework

Water quality analytics is essential for defining and monitoring the safety and health of water sources, requiring evaluation of chemical, biological, and physical variables to meet standards for drinking, recreation, and ecosystem preservation [1]. Maintaining water quality is challenging, but using sensors and technologies like AI and computer vision can help collect essential data and improve natural resource management techniques [2]. Exploiting the potential of Machine Learning and the Internet of Things (IoT) to solve environmental problems has gained popularity in recent years. Water quality analysis and forecasting, a crucial area for human wellbeing, agriculture, and ecological sustainability, exemplifies this trend [3]. Water and air are vital for Earth's quality of life. Monitoring pH, temperature, and pollution in aquatic resources is crucial to reduce their impact.

This article is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license
(https://creativecommons.org/licenses/by/4.0)

Sensor network

Wireless sensor network (WSN) systems are highly regarded for water monitoring due to their real-time, wide-area monitoring capabilities. The deployment of numerous sensor nodes is facilitated by the current affordability and compactness of WSN sensors and wireless communication modules [4]. These devices enable grab sampling reporting and facilitate quantitative analytical approaches at specific times and locations, addressing cost concerns by establishing a cost-effective network that offers valuable qualitative water quality information, thereby supporting informed decision-making [5]. The system's applications extend to live water quality assessment with minimal complexity, energy consumption, and cost, utilizing RF frequencies for direct data transmission without local power supply or node-level signal processing. The system's design and evaluation leverage low-cost printed circuit board (PCB) technology and traditional microwave architecture [6]. The integration of sensor networks is crucial for effective monitoring systems, particularly for collecting data on variables necessary to establish a water quality index for specific areas of Laguna Jucutuma, which also informs predictive modeling efforts.

Water quality

Water quality indicators are recognized for their ability to promptly and accurately detect annual cycles, regional and temporal variations, and current trends in water quality, even at low concentrations [7]. A comprehensive detection system capable of monitoring pH, free chlorine (Cl), temperature, and micro-organic contaminants necessitates a novel design approach to develop an integrated, cost-effective system that can assess multiple water quality parameters simultaneously [8]. In addition to physicochemical parameters like pH, turbidity, temperature, oxidation-reduction potential (ORP), electrical conductivity, and flow rate, microbiological measurements offer a faster and more economical means to evaluate water quality [9]. Traditional monitoring methods reliant on sample collection and laboratory analysis are costly and lack real-time data capture capabilities. Addressing these limitations can significantly enhance water quality monitoring and management capabilities.

Internet of Things

The Internet of Things (IoT), also referred to as the Industrial Internet, Internet of Everything, or simply IoT, represents a transformative paradigm that envisions a globally interconnected network of devices and objects [10]. An IoT-based water quality testing network addresses this concept by continuously monitoring and evaluating water properties using IoT solutions tailored for water quality monitoring [11]. IoT enables real-time access to information from physical objects, fostering innovative services that enhance productivity and efficiency. Key technologies associated with IoT include ubiquitous computing, RFID, wireless sensor networks, and cloud computing [12]. IoT proves crucial for monitoring systems due to its inherent advantages, particularly when integrated with sensor networks that provide continuous data collection. This setup ensures ongoing monitoring of water parameters, accessible globally, which is beneficial for both public awareness and scientific research on water quality management, including treatment plants.

LoRa

For Internet of Things applications requiring low-power, battery-operated embedded devices that must transmit small amounts of data over great distances quickly, the LoRa (Long Range) family of wireless communication is gaining popularity [13]. LoRa is an attractive option for smart sensing in industrial and civil infrastructure applications such as environmental monitoring, because to its extended range and low power consumption [14]. LoRa can communicate reliably over long distances using unlicensed ISM frequency channels. This physical layer contains FEC and CSS. LoRa has emerged as the most popular physical layer for Low Power Wide Area Networks due to its versatility and strength. As a result, LoRaWan was designed to work at higher layers [15]. Continuing with the technology for the internet of things, LoRa communication allows to have that distance factor. It is important that the buoy does not connect directly to the ThingSpeak platform to avoid possible coverage or connectivity problems. LoRa allows to send data from the buoy to a receiving device, which is in charge of uploading the data.

Machine learning

Machine Learning is increasingly indispensable for analyzing data, predicting outcomes, and classifying patterns, especially given the burgeoning volume of aquatic environmental data [16]. Compared to traditional models used in water research, machine learning-based data-driven models excel in addressing complex nonlinear challenges. Leveraging statistical analysis and current data, machine learning plays a critical role in predicting water quality, contamination levels, and implementing preventive measures to maintain cleanliness [17, 18]. Machine learning models can effectively simulate pollutant movements and hydrological processes when sufficient large-scale datasets are available, facilitating the detection and monitoring of dissolved contaminants in water using sensors that measure changes in capacitance values [19]. When a data set with high capabilities is available, Machine Learning technologies can efficiently capture the movement of contaminants and hydrological processes with the integration of sensors can provide important data to identify and monitor a variety of dissolved contaminants in water.

Methodology

This work consists of the design, integration and implementation of a water sensor network with two significant approaches, the first one is the IoT approach consisting of data collection and visualization in real time, and the second one is Machine Learning where different techniques were used to get the best predictive model of water quality index.

Approach

A quantitative research based on data collection of the variables as potential of hydrogen, total dissolved solids, turbidity, temperature and oxidation-reduction potential collected by the sensor network. This data is analyzed with the use of MATLAB and using the weighted arithmetic water quality index method to estimate the quality of water. The approach focuses on the use of the water quality index to evaluate how the variables affect water quality.

Study area

Laguna Jucutuma is situated in San Pedro Sula City, Cortés Department, Honduras, with geographical coordinates at approximately 15°31'12.32"N latitude and 87°54'22.27"W longitude. The lagoon is located in the northern part of the country, near the urban center of San Pedro Sula.

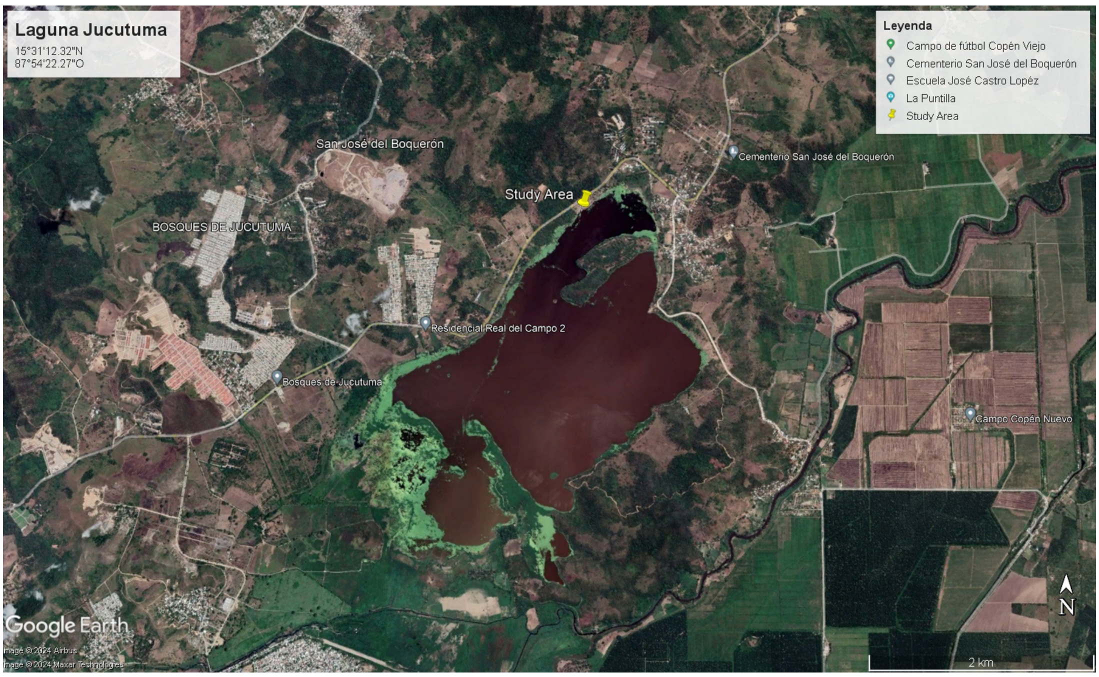

Fig. 1: Satellite Image of Laguna Jucutuma

Applied techniques and tools

- Arduino: The use of the Arduino IDE allowed us to create a code in C++ for the microcontroller to allow data collection and integration to the cloud layer. In addition, it was used to perform AT commands to program the LoRa modules.
- ThingSpeak: It is an online platform which allows users to store data in the cloud, visualize the data and perform data analysis within the online site.
- MATLAB: Is a powerful tool that is widely used to analyze data. It was used to determine the water quality index and the use of Machine Learning techniques to create different linear regression predictive models.
- Proteus: It is a useful CAD program that is used to design electronic circuits, simulate them and create PCB schematics and layouts to make electronic boards.
- SolidWorks: Is a CAD that is used for 2D and 3D modeling for manufacturing, it was used to design a prototype.

Materials

The materials used in the sensor network required to be waterproof and submergible, different interfaces were used to isolate the electronic board with the sensors. Also, the interfaces and micro-controller required a voltage function of 3.3V or 5V to be powered by a solar panel.

- ESP32: Is the main microcontroller programmed to make data collection via ADC and command the LoRa module to send data to other ESP32 which serves as the access point to the cloud.
- pH sensor: Is a sensor that detects ions of hydrogen and by electrical potential measures the pH of a body of water, determining if the water is alkaline or acid.
- TDS sensor: Total suspended solids sensors that indicate how many soluble solids are dissolved in water.
- Turbidity sensor: A sensor which uses light to detect suspended solids in water, measured to indicate the opaqueness of water.
- DS18B20: A digital temperature sensor that is suited to work in the water.
- ORP sensor: Submergible sensor used to measure the oxidation-reduction potential of water.
- - Solar panel: A 5V/1.2A/6W panel was used as the main power supply of the sensor network.
- 3.7V LiPo Battery: Is the secondary power supply of this project, the battery supply is 3.7V and its capacity is 2000mAh.
- Solar Power Manager: Is a high efficiency solar power management module that gives us controllable 5V 1A outputs and provides a space to connect the solar panel and connect a LiPo battery to supply the board and charge while the solar panel is active.
- RYLR993 Lite: Is a LoRa module which acts as a transceiver, two of them were used, one as transmitter and other as receiver, it is low-consumption, and the communication range is up to 20km.
- 868/915 MHz Antenna: Is an antenna that operates with radiofrequency usually in a range of 868 to 915 MHz.

Data collection and sampling

For this project the analogue-digital converter was used to convert voltage readings to understandable parameters except for the DS18B20 temperature sensor, which is digital. The ESP32 ADC converter is of 12-bit resolution and with coding provided by the manufacturer which includes voltage conversion, sensor calibration and temperature compensation to get accurate values from the respective sensors.

Two LoRa modules were used to transmit and receive data, with the use of UART communication with the microcontroller and the transmitter, AT commands are used to communicate with the receiver via radiofrequency, in Laguna Jucutuma it was used a frequency of 915MHz which is the frequency band set aside for ISM (Industrial, Scientific, and Medicine) in America. The receiver was connected to another microcontroller located in a Wi-Fi access point available to upload to the ThingSpeak server.

The data sampling for this study was conducted over a period from May 1st, 2024, to June 3rd, 2024. During this timeframe, Laguna Jucutuma was systematically monitored to gather comprehensive water quality data. To ensure a robust dataset representing various conditions within the specified period, data collection occurred on 20 different days. On each of these days, sampling was performed between 11:00 AM and 1:00 PM GMT-6. This specific time window was chosen to maintain consistency and reduce variability caused by diurnal changes in water quality.

Data samples were taken at intervals of 40 seconds, resulting in approximately 100 data samples being collected each day. This high-frequency sampling approach yielded a total of around 2000 samples over the entire period. The consistent timing and frequent sampling intervals were designed to capture detailed and reliable measurements, providing a comprehensive understanding of the water quality dynamics in Laguna Jucutuma during the study period.

Data processing

The first step in data processing is the pre-processing which consists in the preparation of data, with ThingSpeak we download a .csv file that contained the whole dataset, in MATLAB a code was developed to clean data and normalize to perform better Machine Learning models.

Water Quality Index

Water Quality Index (WQI) is a valuable tool in water resource management, helping to monitor and improve water quality through systematic evaluation and informed decision-making by converting different parameters into a single value. This simplification makes it possible to compare the water quality in various locations and times by condensing vast amounts of data into an understandable manner. WQI assists in clearly communicating the state of water quality to management and the public.

The Weighted Arithmetic Water Quality Index (WAWQI) is a widely used method to assess overall water quality by providing a consolidated view of water quality, considering the relative importance of each parameter through assigned weights. In order to calculate the water quality index, the next formula was used [20]:

$$WQI = \sum W_i Q_i \tag{1}$$

where W_i is the unit weightage for each parameter denoted by the following formula:

$$W_i = k/S_i \tag{2}$$

S_i is the recommended standard of each parameter and k is a proportional constant calculated by:

$$k = 1/\sum(1/S_i) \tag{3}$$

following Eq. 1, Q_i is the subindex of each parameter which is determined by:

$$Q_i = 100(V_i - V_o)/(S_i - V_o) \tag{4}$$

V_i is the monitored value of each parameter and V_o is the ideal value of each parameter values can be visualized in [20]. After calculating the result can be classified according to Table 1:

Table 1: WQI Status

WQI	Status
0-25	Excellent
26-50	Good
51-74	Poor
76-100	Very Poor
>100	Unfit

Machine learning techniques

Support Vector Machine (SVM) Regression

SVM regression leverages the same core principle as SVM classification - finding a hyperplane that maximizes the margin between the predicted values and a margin of error. In regression, this margin of error represents the allowed deviation between predicted and actual values. Unlike linear regression, SVM regression can handle non-linear relationships between features and the target variable by employing kernel functions that map the data into higher-dimensional spaces. However, SVM regression can be computationally expensive for large datasets and requires careful parameter tuning for optimal performance.

Gaussian Process Regression

Gaussian Process (GP) regression is a probabilistic approach that assumes the data follows a Gaussian distribution. It utilizes a kernel function to model the covariance between data points, enabling predictions with uncertainty quantification. This provides information about the model's confidence in its predictions. GP regression is flexible and can handle complex relationships but can also be computationally expensive for large datasets and may require expertise for effective parameter selection.

Kernel Approximation

Kernel Approximation methods represent a family of techniques that employ kernel functions to capture non-linear relationships in the data. These methods often involve approximating a complex function using a combination of simpler basis functions. While offering more flexibility than linear regression, they can be computationally expensive for high-dimensional data and require careful selection of the kernel function and its hyperparameters.

Neural Networks

Artificial Neural Networks (ANNs) are powerful tools capable of modeling complex relationships between features and the target variable. In regression tasks, ANNs learn these relationships through a series of interconnected layers of processing units (neurons). By adjusting the weights and biases of these connections during training, ANNs can capture intricate patterns in the data. However, ANNs require significant training data and computational resources, and their interpretability can be challenging due to their complex internal structure.

Model performance

The evaluation of a machine learning model's performance is a crucial step in the development process. This section explores various metrics employed to assess the efficacy of the proposed models: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-Squared (Coefficient of Determination) [21].

R-Squared

Shows how much the dependent variable's variance that the model's independent variables can account for. A higher number indicates a better fit. Its range is 0 to 1. It is important to note that R-Squared does not directly reflect the magnitude of errors but rather the model's ability to capture the overall trend in the data.

$$R^2 = 1 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{5}$$

MSE

Calculates the average squared difference between the predicted values and the actual values, lower MSE values signify a better model fit, with a value of 0 indicating a perfect fit. However, MSE is sensitive to outliers, as large errors are squared and contribute more significantly to the overall error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

RMSE

Is easier to interpret in the same units as the predicted values, making it a more intuitive metric. However, it inherits the limitations of MSE regarding sensitivity to outliers.

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}} \tag{7}$$

MAE

Determines the mean of the absolute variations between the values that were anticipated and those that were observed. Unlike MSE, MAE is less sensitive to outliers, providing a clearer picture of the average magnitude of errors.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

MAPE

Is beneficial for comparing errors across different data ranges, particularly when dealing with percentages or proportions. However, it is not ideal for scenarios with zero actual values, as it can lead to division by zero errors.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \tag{9}$$

In these formulas, n is the total number of observations in the dataset, y_i is the monitored value of the dependent variable, $\hat{y}$ is the predicted value of the dependent variable and $\hat{y}$ is the mean value of y_i.

Study methodology

In this study, we employed a hierarchical method to systematically evaluate and optimize a three-layered water quality sensor network system. This system architecture is designed to enhance the monitoring and management of water quality parameters across various environments. The three distinct layers—device, edge, and cloud—each contribute uniquely to the overall functionality and performance of the sensor network.

A. Device Layer

The device layer forms the foundational level of the sensor network, encompassing physical sensors, power supply mechanisms, and microcontroller units. These components are deployed directly within the water bodies or monitoring points to capture real-time data on crucial parameters such as pH levels, total dissolved solids, turbidity, temperature and oxidation reduction potential. Calibration processes within this layer ensure the accuracy and reliability of sensor readings, crucial for generating precise environmental data.

B. Edge Layer

Situated between the device and cloud layers, the edge layer acts as a pivotal hub for data aggregation and preliminary processing. Here, data from multiple sensors is consolidated and formatted before transmission to the cloud for further analysis and storage. Integration of technologies like

Long-Range (LoRa) communication modules enables efficient and low-power data transmission over extended distances, facilitating seamless connectivity between remote sensor nodes and the central cloud infrastructure.

C. Cloud Layer

At the apex of the hierarchy lies the cloud layer, where vast volumes of sensor data are stored, managed, and analyzed. Cloud-based storage solutions provide scalability and flexibility, accommodating the continuous influx of real-time data streams from distributed sensor nodes. These insights can inform timely decision-making processes related to water quality management, resource allocation, and environmental remediation strategies.

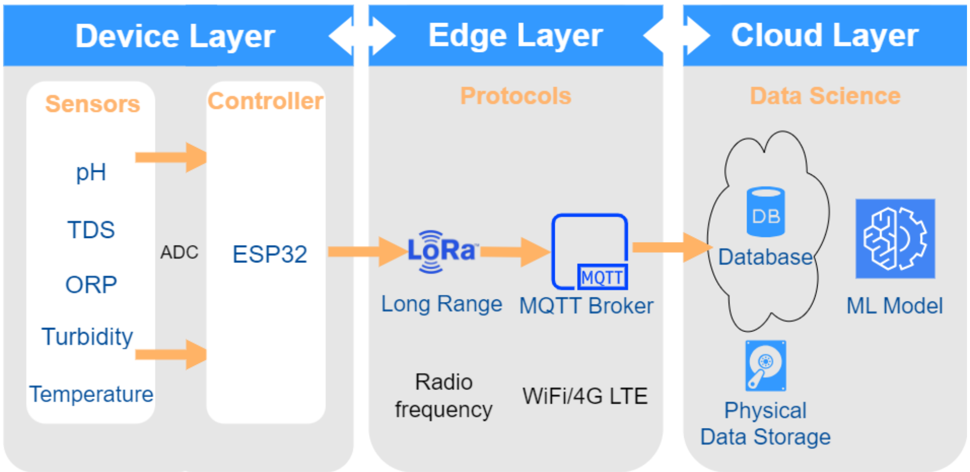

Fig. 2: Hierarchical Method Framework

Validation methodology

To validate the effectiveness and reliability of the three-layered water quality sensor network system, encompassing the device layer, edge layer, and cloud layer, we implemented a rigorous validation methodology. This approach aimed to assess the accuracy of sensor readings, the efficiency of data transmission, and the robustness of data analysis and decision-making processes facilitated by the cloud-based infrastructure. *A. Sensor Calibration and Validation*

- Calibration Procedures: Sensors at the device layer underwent meticulous calibration procedures to ensure consistency and accuracy in measuring water quality parameters. Calibration standards were established based on laboratory benchmarks and manufacturers information.

- Validation Tests: Validation tests were conducted under varying environmental conditions to verify sensor accuracy and reliability that included comparison against standardized reference measurements.

B. Edge Layer Data Transmission Evaluation

Resilience tests were conducted to simulate scenarios of network congestion or interference, ensuring the robustness of data transmission mechanisms in adverse conditions.

C. Cloud Layer Data Collection and Machine Learning Evaluation

- Data Collection: We ensured that data was sent to ThingSpeak by guaranteeing the microcontroller an access point to upload data to the cloud.

- Analytical Accuracy: Machine learning models and analytical algorithms deployed in the cloud layer were subjected to validation tests to assess their predictive accuracy and reliability.

Design

For the correct operation, a design was made around three important parts or stages for the correct operation: Electronics, Communication and Data Processing.

Electronics

After selecting the variables to be measured and selecting the corresponding sensors, an electronic circuit is designed together with a microcontroller for data processing. The information is collected by the five sensors and then processed by the ESP32 microcontroller, which is then sent to another microcontroller through the REYAX RLYR 993 Lite module. The emitter device will be powered with 5V from the solar panel, having a maximum load capacity of 1.2A, being more than enough for the 147.5mA consumed by the emitter buoy.

In turn, the receiving device has the same module and controller, as well as five LED lights that indicate the receipt of the information. Once the two circuits are designed, the PCB design was done in the Proteus software.

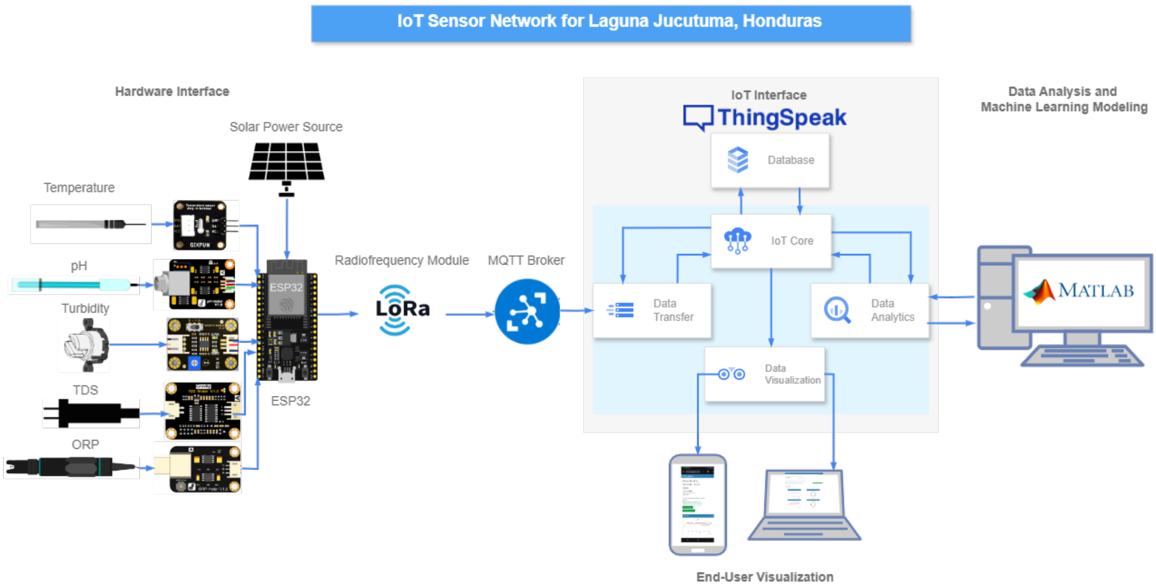

Fig. 3: IoT Architecture

Communication

To send data, both the transmitting and receiving devices use the REYAX RLYR 993 Lite LoRa modules. These modules use LoRa (Long Range) technology and communicate by AT commands through the UART serial protocol and radio frequency, using the air as a means of transport, so in this section the devices are wireless. Subsequently, the transmitter device, upon receiving the data obtained by the sensors, the micro-controller is responsible for sending the data to the ThingSpeak platform, using the MQTT protocol (Message Queuing Telemetry Transport), a protocol that allows the sending of light messages, which travel through reduced bandwidths and designed for devices that occupy small codes and communication between machines, M2M, and IOT.

Data processing

For this stage, ThingSpeak is in charge of graphing and saving the data; this allows to have a historical count of the information, as well as the ease of visualization of such data. MATLAB oversees using these data so that, by using Eq.1, a response can be generated, and, in turn, predictive models can be generated from the data obtained.

Results

This section shows the results obtained during the whole period, as well as the manufacture of elements described in the design part. Starting with the electronic part, as shown in Fig. 7, a PCB board was manufactured where the 5 sensors, the microchip, the radio frequency module and the power supply from the solar panel are concentrated. This part was fundamental since in this way there is a better conduction of the information, this due to the medium through which the energy circulates, leaving aside the cables and working on copper tracks. This also avoids loss of information, so the measurements are more accurate.

Continuing with the manufacturing section in the electronic part, as can be seen in Fig. 8, a PCB board was also made for the receiver device; device that is responsible for sending data through IoT, so it is crucial to maintain a stable and adequate current and voltage.

Fig. 4: Final Prototype Deployment

For the communication part, Fig. 8 shows an antenna which is used to transport the radio frequency information from the transmitting device located in the buoy to the receiving device. For data processing, the data is sent from the module to the ESP32 with serial communication (UART). Subsequently, as shown in Fig. 9, these data are sent to ThingSpeak, where they can be displayed in the form of graphs, facilitating a study of these same. This platform also offers the benefit of saving these values, to be later downloaded and used by MATLAB.

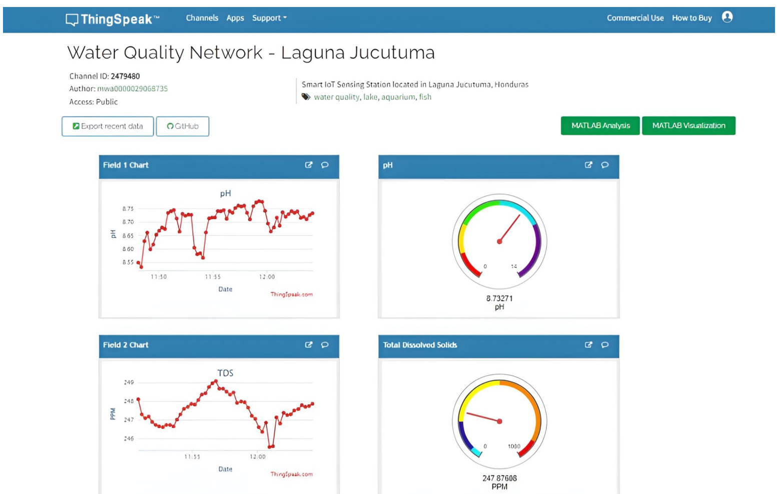

Fig. 5: ThingSpeak Interface

In the final stage of the project, in the data processing, all the collected information is used in MATLAB, where it shows us different results, in the first place it shows statistical information of the collected data of each variable, such as the average, the median or the standard deviation. Different machine learning models were used to find the most suitable model. As can be seen in Table 2 the model with better fit and less magnitude of errors was the Gaussian Process, SVM showed very good efficacy but not better than Gaussian Process, both ANN and Kernel were good, but they were susceptible to errors.

Table 2: Model Performance Results

Model	RMSE	R-Squared	MSE	MAE	MAPE
Gaussian Process	0.78544	0.98	0.61692	0.33159	0.80%
Support Vector Machines	0.91823	0.98	0.84315	0.55446	1.30%
Neural Networks	1.133	0.97	1.2836	0.68593	1.70%
Kernel Approximation	1.3339	0.95	1.7792	0.88599	2.10%

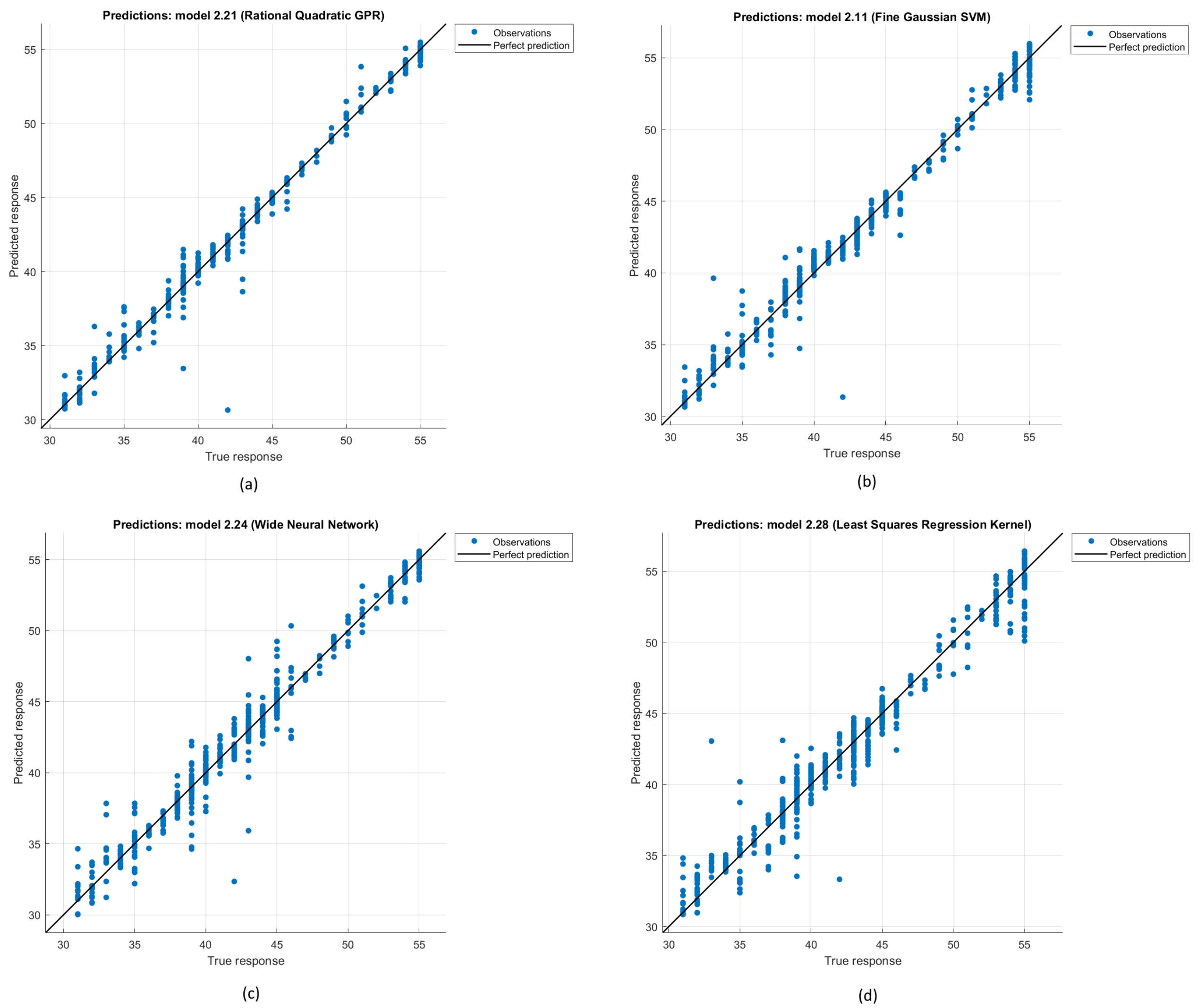

Fig. 6: (a) Rational Quadratic GPR Model, (b)Fine Gaussian Model, (c)Wide Neural Network Model, (d)Least Squares Regression Kernel Model

Around the 20 days the range of the WQI was between 29 and 52, at the beginning of May, water quality index was at 40 and started increasing due to the rise of air pollution and smog, the highest value was registered in May 15. At the beginning of June, WQI index was low because air pollution turned better. Other studies implement statistically based methodologies with which a good level of accuracy is achieved but are dependent on people-based processes [22, 23]. In this case we propose a device capable of monitoring water quality by applying artificial intelligence techniques capable of operating in site. This project is low cost and can be developed in different lakes and rivers where water quality is having adverse effects on nature.

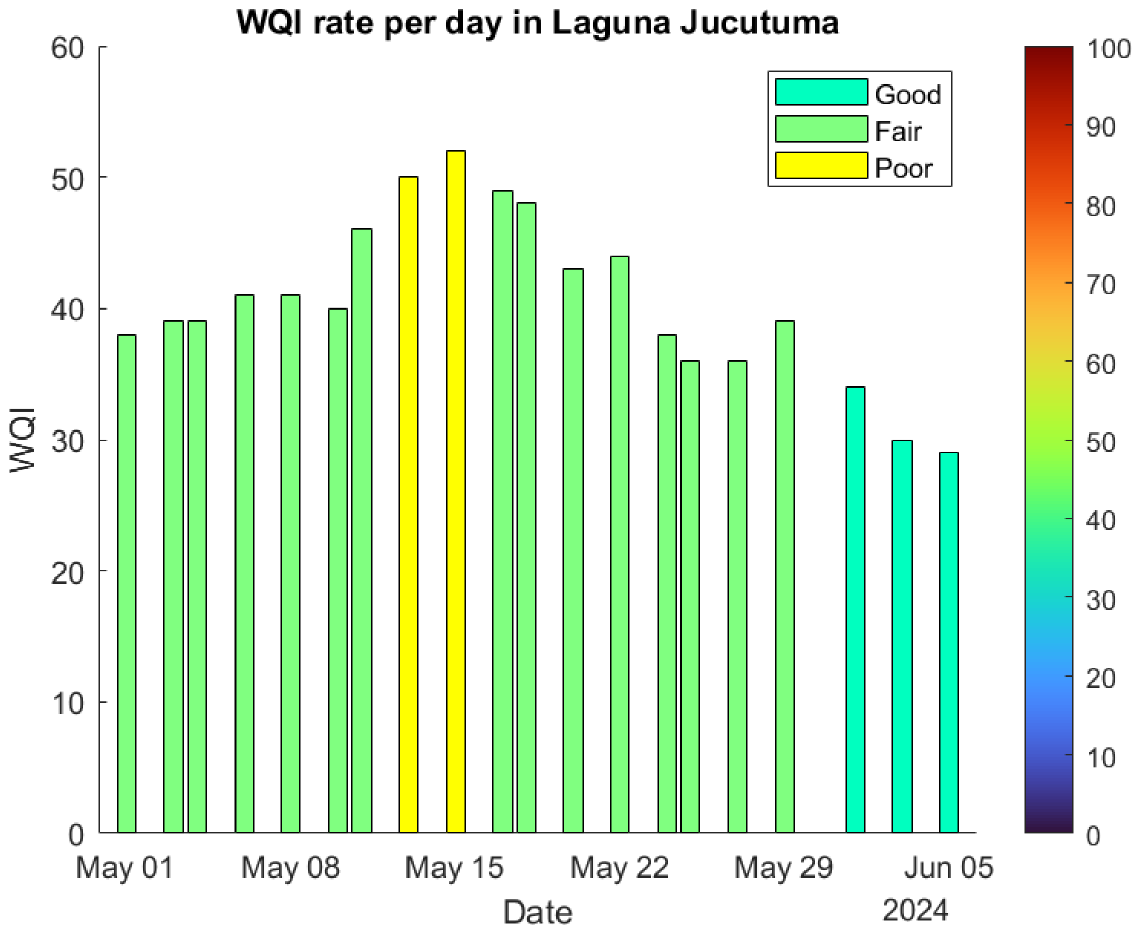

Fig. 7: Water Quality Index rate per day

Conclusions

The implementation of IoT sensors and Machine Learning allows the collection and analysis of real-time data on water quality, improving environmental management in Laguna Jucutuma, facilitating the protection and prevention of pollution problems in the lagoon.

The maximum value of the water quality index was calculated as 52 on May 15 when the air pollution was at its highest point and the lowest index of 29 was recorded on June 6.

The Gaussian Process regression learner model was the most fitted model with an RMSE of 0.78544, R-squared of 0.98, MSE of 0.61692, MAE of 0.33159 and MAPE of 0.80%.

References

[1] Poornima Jayaraman, Kothalam Krishnan Nagarajan, Pachaivannan Partheeban, and Vallidevi Krishnamurthy. Critical review on water quality analysis using IoT and machine learning models. *International Journal of Information Management Data Insights*, 4(1):100210, 4 2024.

[2] Omar Capetillo-Contreras, Francisco David Pérez-Reynoso, Marco Antonio Zamora-Antuñano, José Manuel Álvarez-Alvarado, and Juvenal Rodríguez-Reséndiz. Artificial Intelligence-Based Aquaculture System for Optimizing the Quality of Water: A Systematic Analysis. *Journal of Marine Science and Engineering*, 12(1):161, jan 13 2024.

[3] Mushtaque Ahmed Rahu, Abdul Fattah Chandio, Khursheed Aurangzeb, Sarang Karim, Musaed Alhussein, and Muhammad Shahid Anwar. Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality. *IEEE Access*, 11:101055–101086, 2023.

[4] Tomoaki Kageyama, Masashi Miura, Akihiro Maeda, Akihiro Mori, and Sang-Seok Lee. A wireless sensor network platform for water quality monitoring. In *2016 IEEE SENSORS*, pages 1–3, Orlando, FL, USA, 10 2016. IEEE. [Online; accessed 2024-06-26].

[5] Kevin Murphy, Brendan Heery, Timothy Sullivan, Dian Zhang, Lizandra Paludetti, King Tong Lau, Dermot Diamond, Ernane Costa, Noel OConnor, and Fiona Regan. A low-cost autonomous optical sensor for water quality monitoring. *Talanta*, 132:520–527, 1 2015.

[6] Rashid Mirzavand, Mohammad Honari, Bahareh Laribi, Behnam Khorshidi, Mohtada Sadrzadeh, and Pedram Mousavi. An Unpowered Sensor Node for Real-Time Water Quality Assessment (Humic Acid Detection). *Electronics*, 7(10):231, oct 3 2018.

[7] Shweta Tyagi, Bhavtosh Sharma, Prashant Singh, and Rajendra Dobhal. Water Quality Assessment in Terms of Water Quality Index. *American Journal of Water Resources*, 1(3):34–38, oct 29 2020.

[8] Arif Ul Alam, Dennis Clyne, and M. Jamal Deen. A Low-Cost Multi-Parameter Water Quality Monitoring System. *Sensors*, 21(11):3775, may 29 2021.

[9] Niel Andre Cloete, Reza Malekian, and Lakshmi Nair. Design of Smart Sensors for Real-Time Water Quality Monitoring. *IEEE Access*, 4:3975–3990, 2016.

[10] In Lee and Kyoochun Lee. The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, 58(4):431–440, 7 2015.

[11] Halima Haque, Kashshaf Labeeb, Rabea Basri Riha, and Md. Nasfikur R. Khan. Iot Based Water Quality Monitoring System By Using Zigbee Protocol. In *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pages 619–622, Pune, India, mar 5 2021. IEEE. [Online; accessed 2024-02-14].

[12] N Vijayakumar and R Ramya. The real time monitoring of water quality in IoT environment. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pages 1–5, Coimbatore, India, 3 2015. IEEE. [Online; accessed 2024-06-26].

[13] Shilpa Devalal and A. Karthikeyan. Lora Technology - An Overview. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 284–290, Coimbatore, 3 2018. IEEE. [Online; accessed 2024-06-26].

[14] Aloÿs Augustin, Jiazi Yi, Thomas Clausen, and William Townsley. A Study of LoRa: Long Range & Low Power Networks for the Internet of Things. *Sensors*, 16(9):1466, sep 9 2016.

[15] Jetmir Haxhibeqiri, Eli De Poorter, Ingrid Moerman, and Jeroen Hoebeke. A Survey of LoRaWAN for IoT: From Technology to Application. *Sensors*, 18(11):3995, nov 16 2018.

[16] Mengyuan Zhu, Jiawei Wang, Xiao Yang, Yu Zhang, Linyu Zhang, Hongqiang Ren, Bing Wu, and Lin Ye. A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health*, 1(2):107–116, 6 2022.

[17] Sakshi Khullar and Nanhey Singh. Machine learning techniques in river water quality modelling: a research travelogue. *Water Supply*, 21(1):1–13, feb 1 2021.

[18] Maria Elena Perdomo, Manuel Cardona, Denisse Melisa Castro, and Wendy Melany Mejia. Literature review on artificial intelligence implementation in the honduran agricultural sector. In *2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*, pages 1–5, 2022.

[19] Nida Nasir, Afreen Kansal, Omar Alshaltone, Feras Barneih, Mustafa Sameer, Abdallah Shanableh, and Ahmed Al-Shamma'a. Water quality classification using machine learning algorithms. *Journal of Water Process Engineering*, 48:102920, 8 2022.

[20] Razvan Bogdan, Camelia Paliuc, Mihaela Crisan-Vida, Sergiu Nimara, and Darius Barmayoun. Low-Cost Internet-of-Things Water-Quality Monitoring System for Rural Areas. *Sensors*, 23(8):3919, apr 12 2023.

[21] Mosleh Hmoud Al-Adhaileh and Fawaz Waselallah Alsaade. Modelling and Prediction of Water Quality by Using Artificial Intelligence. *Sustainability*, 13(8):4259, apr 12 2021.

[22] Aldo Fernando Zavala Jones, Ivanna Maria Caballero, and Maria Elena Perdomo Perdomo. DEVELOPMENT OF METHODOLOGY FOR THE ANALYSIS OF WATER QUALITY IN RIVERS. August 2023.

[23] Arlene Celeste Alarcon Madrid, Allan Francisco Garcia, Andrea Yanara Hernandez Martinez, Maria Elena Perdomo Perdomo, and Lessy Marisol Salmeron Sierra. Statistical Analysis of The Effectiveness of An Electrocoagulator In The Process Of Cleaning Wastewater From The Rio Blanco River. In *Proceedings of the 22nd LACCEI International Multi-Conference for Engineering, Education and Technology (LACCEI 2024): "Sustainable Engineering for a Diverse, Equitable, and Inclusive Future at the Service of Education, Research, and Industry for a Society 5.0."*. Latin American and Caribbean Consortium of Engineering Institutions, 2024.