Mehran University Research Journal of Engineering and Technology

https://doi.org/10.22581/muet1982.2401.2806

2024, 43(1) 192-205

IoT and machine learning solutions for monitoring agricultural water quality: a robust framework

Mushtaque Ahmed Rahu [a, *], Muhammad Mujtaba Shaikh [b], Sarang Karim [b], Abdul Fattah Chandio [a], Safia Amir Dahri [b], Sarfraz Ahmed Soomro [b], Sayed Mazhar Ali [c]

[a] *Department of Electronic Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah 67450 Sindh Pakistan*

[b] *Department of Telecommunication Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah 67450 Sindh Pakistan*

[c] *Department of Electrical Engineering, Mehran University of Engineering and Technology, SZAB Campus Khairpur*

[*] Corresponding author: Mushtaque Ahmed Rahu, Email: rahumushtaque@gmail.com

Received: 14 April 2023, Accepted: 20 December 2023, Published: 01 January 2024

KEYWORDS

Data Collection

Water Quality Monitoring

Agriculture

IoT

Cloud

Machine Learning

ABSTRACT

All living things, comprising animals, plants, and people require water to survive. The world is covered in water, just 1 percent of it is fresh and functional. The importance and value of freshwater have increased due to population growth and rising water demands. Approximately more than 70 percent of the world's freshwater is used for agriculture. Agricultural employees are the least productive, inefficient, and heavily subsidized water users in the world. They also utilize the most water overall. Irrigation consumes a considerable amount of water. The field's water supply needs to be safeguarded. A critical stage in estimating agricultural production is crop irrigation. The global shortage of fresh water is a serious issue, and it will only get worse in the years to come. Precision agriculture and intelligent irrigation are the only solutions that will solve the aforementioned issues. Smart irrigation systems and other modern technologies must be used to improve the quantity of high-quality water used for agricultural irrigation. Such a system has the potential to be quite accurate, but it requires data about the climate and water quality of the region where it will be used. This study examines the smart irrigation system using the Internet of Things (IoT) and cloud-based architecture. The water's temperature, pH, total dissolved solids (TDS), and turbidity are all measured by this device before the data is processed in a cloud using the range of machine learning (ML) approaches. Regarding water content limits, farmers are given accurate information. Farmers can increase production and water quality by using effective irrigation techniques. ML methods comprising support vector machines (SVM), random forests (RF), linear regression, Naive Bayes, and decision trees (DT) are used to categorize pre-processed data sets. Performance metrics like accuracy, precision, recall, and f1-score are used to calculate the performance of ML algorithms.

1. Introduction

Everyone in the country needs food, hence agriculture is essential to the economy of the country. It is related to one of the most important occurrences in American history. If the country has a sizable farming population, it is seen as being both socially and economically prosperous. The key sector that generates employment in the majority of countries is agriculture. Help with planting and animal care is typically required on a large farm with numerous occupants. Big farms can use neighboring handling conveniences to enhance and finish their agricultural properties. The role of agriculture in human civilization has evolved significantly over time. Variations have made it probable to use fewer assets and accomplish less effort. Although, there has never been a demand and supply balance because of the high population density [1]. By 2050, it is expected that there will be 9.8 billion folks on earth. Most population growth is probably to be found in developing countries. The percentage of people who live in cities is predicted to increase from 49% to 70% by 2050 [2]. Also, especially in developing countries, the need for food will rise as wages rise. People in these nations will thus be more attentive to the quality of their nutrition and food. Customers might start to choose beans and eventually meat over grains and cereals as a result. Water is an essential usual reserve for irrigation even if it is inadequate. In the country, irrigation consumes a lot of water. Crop irrigation has a substantial impact on crop output since it is impacted by several ecological influences, including soil and air temperature, soil moisture, and humidity. During reaping lands, farmers mostly depend on mankind's direction and knowledge. The field's water supply needs to be sustained. The lack of water is the main issue in the modern world. Globally, the public currently encounters such scarcity. In the upcoming years, the situation could get worse. A recognized and superior approach to farm management that has grown in popularity in contemporary agriculture is referred to as "smart farming." Agriculture and information technology are used to track the health and productivity of crops. Monitoring the condition of field crops and other indicators is necessary for this. The ultimate goal of intelligent farming is to lower input costs while keeping the same level of output quality. When a large amount of pesticide or fertilizer is administered at once, the entire field is treated as one unit [3]. Moreover, fresh water is a crucial natural resource for all ecosystems to survive. Unfortunately, only 2.53% of a whole water body is currently existing as fresh water. As a result, fresh water is in short supply in the majority of countries around the world. For all ecosystems to survive, fresh water needs to be accessible. As per the World Resources Institute (WRI), almost all nations will soon face a water shortage [4]. The fact that industry and agriculture use an excessive amount of freshwater has a big impact on ecosystems downstream. It is crucial to use fresh water in a way that avoids having a detrimental influence on future generations due to its shortage.

Many soil kinds, comprising clay, salty, and sandy soils, are present in the soil. Every kind of soil has unique pros and cons. A notable instance of this is sandal soil, which has a huge volume for drainage. On another hand, drainage quickly eliminates soil nutrients. The qualities of soil have a significant impact on how much water the plant wants [5]. Using a variety of data collection and storage techniques, IoT has also made smart farming potential. Current irrigation systems with smart sensor networks gather field data for prime plant irrigation. ML is used in a variety of real-life uses, like precision farming, smart agriculture, smart healthcare, smart water quality, smart manufacturing, and smart logistics. IoT expands effectiveness, lowers costs, maximizes energy use, retains forecast accuracy, and offers excessive compact of convenience to the overall population. Security issues are becoming more prevalent as data and systems processing turn into further variation. Privacy and security issues are key roadblocks to the growth of IoT. According to a recent study by ecologists [6], almost the world's population may confront water calamity by 2025. Contrarily, most freshwater is used for industrial and agricultural purposes, which has a considerable effect on ecosystems lower down the watershed. Therefore, careful management of freshwater use is necessary to prevent adverse effects on the availability of water for future generations. Figure 1 illustrates that in Pakistan 81% of fresh water is used for irrigation, 13% is used for production, and 6% is used for home purposes, according to Shahmir Janjua [7]. By 2025, it's anticipated that human water usage will rise by up to 26%.

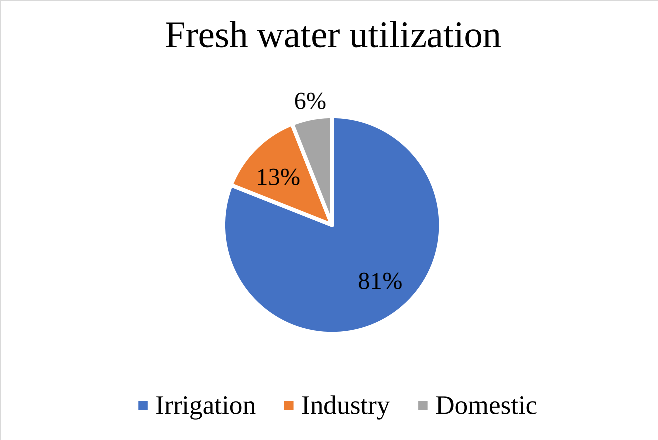

Fig. 1. Freshwater utilization

Natural resources are being depleted at an accelerated rate due to population growth and demand. Since irrigation uses at least 75% of the water used globally, it is crucial to the productivity and expansion of agriculture, making water quality the main issue. Since farmers adopt irrigation, which is the world's biggest use of water, irrigation needs to be improved, and because it is done, irrigation is not an efficient use of water. Farmers must monitor data such as soil variety, climatic situations, existing water resources, soil humidity, soil nutrients, and soil pH to improve irrigation management and make conclusions that reduce or eliminate agricultural complications.

To address the complex issues facing agriculture, irrigation, a data-driven technology, needs to be integrated with new technologies and cutting-edge approaches. This study provides a summary of current IoT-enabled technologies that can improve agricultural water quality and irrigation management [8].

This article discusses how irrigation and IoT have developed, what needs to be taken into account for optimal irrigation, why effective irrigation optimization is necessary, and how vibrant irrigation optimization could decrease water waste. Usage of IoT, arrangement of models, controllers, and sensors in agriculture, platforms of cloud IoT, water prediction, and ML models for irrigation are all demonstrated in the study. The creation of more effective irrigation management applications benefits from the convergence of tools, technology, and methodologies. For the creation of efficient irrigation management software, access to real-time data must be improved, like water quality metrics, and weather, plant, and soil data.

This study makes the case that agriculture can benefit from an Internet of Networks paradigm. Privacy and security are essential components of Internet of Apps, in addition to IoT networks tied to agriculture. This method takes a set of data as input. All symbolic features are translated to numeric features for pre-processing, and all numeric features are subsequently converted back to symbolic features. To extract features, the prime constituent investigation is employed. Following pre-processing, the pre-processed data set is categorized using ML methods like SVM, linear regression, RF, DT, and Naive Bayes. When assessing the effectiveness of ML systems, metrics accuracy, precision, recall, and f1-score are taken into attention. [9]

2. Problem Statement

Monitoring water quality is a crucial duty that ensures the safety and usability of water resources. Traditional water quality monitoring techniques, however, take a long time, cost a lot of money, are inaccurate, and frequently don't produce data in real-time. IoT [10, 39, 43] and ML [40] have recently come to light as viable technologies for tracking water quality. IoT uses networked sensors and other gadgets to gather and send data from the real world to the digital one. Water quality monitoring can be automated using IoT and offers real-time data that is accessible from a distance. This can significantly cut down on the time and expense of using conventional water quality monitoring techniques, and effective, accurate system performance for ML. IoT sensor data can be analyzed using ML [11] to find patterns that could point to changes in the quality of water. Based on the current situation and previous data, ML algorithms can forecast the water quality of the future. This can aid in spotting possible problems before they develop into larger ones. A thorough framework for IoT and ML-based water quality monitoring is needed to overcome these difficulties. The creation of reliable IoT sensors, data processing and analysis methods, and ML algorithms that can provide precise and dependable forecasts regarding water quality should all be part of this framework. Ultimately, the monitoring and management of our water resources could be completely changed by the application of IoT and ML for water quality monitoring. These technologies can help to ensure that our water supplies are safe and useable for future generations by giving real-time data and prediction insights.

3. Objectives

To ensure that irrigation water is safe and healthy for crops and does not harm soil or plants, the following are a few of the specific goals for irrigation water quality:

1. Ensure that the water used for irrigation is free from harmful chemicals, and impurities that negatively affect crops.

2. Maintain an appropriate pH level in the irrigation water. This helps ensure that the soil maintains the right nutrient balance for plant growth.

3. Control TDS levels in irrigation water. A high range of TDS can damage crops and deplete the soil of nutrients. TDS monitoring is essential to determine the overall quality of water used for irrigation. However, TDS levels can vary significantly depending on the type of soil, fertilizer used, and other factors, making it challenging to establish a baseline TDS level. United Nations (UN) [38] Quote: "TDS monitoring is essential to ensure that irrigation water is within safe limits and does not negatively impact crop yields."

4. Maintain appropriate levels of turbidity in the irrigation water. This is crucial for maintaining healthy root systems in plants. Turbidity monitoring is critical to determine the quality of water for irrigation purposes, and it is necessary to ensure that the water is free from suspended solids and other contaminants.

5. Effective monitoring of water temperature is essential to manage water resources sustainably and support agricultural productivity.

Water quality control for irrigation's overall goal is to guarantee that water is used responsibly and effectively to maximize crop yields. The necessity of monitoring these factors for effective management of water resources and agricultural output is acknowledged by the UN [38].

4. Related Work

4.1 Review of IoT in Agricultural Water Quality

Monitoring water quality is crucial in agriculture to guarantee safe and responsible water use for crop production. Due to its potential for real-time monitoring and effective data collecting, the use of IoT technology in water quality monitoring has attracted interest recently. We will examine recent works on IoT-based quality of water monitoring in agriculture in this literature review [22]. The creation of less price IoT-based water quality monitoring system for precision agriculture was the focus of one study by [12] authors. The system was made up of pH, dissolved oxygen, and temperature sensors that were wirelessly networked to a central computer. The study showed the system's potential for real-time water quality parameter monitoring in precision agriculture. Another study by [13] monitored the water quality in a greenhouse tomato production system using an IoT-based system. The system was made up of temperature, electrical conductivity, dissolved oxygen, pH, and dissolved oxygen sensors that were linked to a central database. According to the study, IoT-based systems can monitor and analyze water quality data in real-time, allowing for effective control of greenhouse systems. In a Chinese rice paddy area, a study by [14] used IoT-based water quality monitoring. A system had a network of temperature, electrical conductivity, and water level sensors that were linked to a central computer. The study showed that IoT-based systems were capable of providing precise and actual monitoring of quality water in rice paddy fields. In conclusion, recent studies have demonstrated encouraging results for the application of IoT-based systems for quality water monitoring in agriculture. These schemes' actual monitoring and effective data collection capabilities can facilitate the effective management of water resources in agriculture, resulting in the sustainable crop production of high-quality crops.

4.2 Review of ML in Agricultural Quality

Monitoring water quality in agriculture is crucial for making sure that agricultural methods are sustainable and that the environment is protected. The accuracy and effectiveness of water quality monitoring in agriculture have greatly improved in recent years thanks to the widespread use of ML techniques. Several researches have been conducted to investigate the use of ML learning algorithms in agricultural water quality monitoring. A study by [15] employed ML techniques, for instance, to forecast the water quality indicators of Chinese rivers and lakes. A deep learning method was employed in a different study by [16] to forecast the nitrate content of groundwater. The research revealed that deep learning algorithms outperformed conventional regression models in terms of accuracy. Data on water quality has also been classified using ML methods. For instance, the work by [17] classified water quality data into various categories based on the water quality indicators using ML algorithms. ML techniques have been utilized for anomaly detection in water quality data in addition to prediction and categorization. In the work by [18], inconsistencies in water quality data gathered from a river in Pakistan were found using ML algorithms. The majority of the research points to ML algorithms as having enormous promise for monitoring water quality in agriculture. These algorithms can assist in decreasing expenses, increasing the precision and effectiveness of water quality monitoring, and increasing the sustainability of agricultural methods.

4.3 Reviews of Water Quality

In many industries, such as agriculture, water quality, and defense applications in daily life, wireless sensor networks (WSNs) are essential.

The lifespan property of WSN, or the amount of time it takes for a sensor to run out of power, was investigated by authors in [19]. Also, they looked at the WSN lifespan problem using a metaheuristic approach. They provide a metaheuristic algorithm with three steps: transition, evaluation, and determination, which aid in determining the best solution to the issue.

Their rationale makes it evident that mastery of the numerous field characteristics and subject-specific information about longevity problems are prerequisites for applying a metaheuristic technique. A lot of researchers have investigated a variety of metaheuristic algorithms. Even if these tactics are used and the performance of the WSN improves, there are still some unsolved problems. When the possibility is measured, numeral sensors or cluster heads, for instance, might be decreased. Moreover, even though the bulk of metaheuristic techniques are considered with optimization in mind, they might not be effective when used to address longevity issues [20].

If farmers want to boost production, they must prioritize water quality. In these situations, using WSN may be very helpful in guiding farmers and other agricultural actors, such as management irrigation organizations, in making appropriate judgments that support irrigation necessities and crop output predictions [21].

Authors in [23] based on principal pivot irrigation schemes, created the self-contained agricultural accuracy system. The method uses an underground sensor network to observe area variables including soil temperature and humidity. This confirms that power settings are adjusted for input configuration and standard energy preservation and that sensors are configured.

Authors in [24] explored an application of wireless sensors in irrigation planning. They use a sensor array system and sensor network-based precision knowledge to calculate soil moisture and temperature to estimate watering needs in real time.

Authors in [25] presented a new ontologically guided wireless sensor or actuator-based separate zone watering system. The PA used in this technique is based on speaking plants to protect further water. The performance of a scheme is enhanced by combining several ML approaches to identify network node issues. Several end-user apps have been formed earlier to develop automation and usability of irrigation responsibilities.

Authors in [26] conducted studies for sessional exploration on dry and rainy times and investigated the effects of five water harvesting procedures. In three periods of sowing, midseason, and after reap, their model took into account the moisture content of soil at four dissimilar complexities. They confirmed that their approach is real in reducing water use and increasing output.

Authors in [27] shaped a microcontroller software to accurately manage the soil temperature of plants. For communication connectivity, it uses a solar cell and a cellular Internet interface. According to the results of their 136-day studies, the authors suggested irrigation approach may utilize up to 90% less water when compared to standard farming practices.

Authors in [28] devised the method for cotton crop irrigation. Their method was utilized to calculate the soil water balance using data sets from multiple cotton-cultivating areas. Using data sets, researchers developed an Android app. Their program was also intended to collect climate data from climate locations near sites where it was utilized. Using both internal data and downloaded meteorological data, the application assessed the irrigation requirements and autonomously scheduled the irrigation systems to increase cotton yield.

Authors in [29] solve challenges including farming reserve optimization, decision provision, and field monitoring, an environmental WSN clarification known as a smart greenhouse monitoring scheme was developed. Their plan boosts crop productivity while also making the most use of water and fertilizer. They discussed how the environment influences how plants grow.

Authors in [30] indicated that assessing the global water situation to make wise water use decisions is crucial. Farmers in Pakistan, notably those in the Nawab Shah district, frequently over-irrigate their fields because they are unaware of the crop water necessities and believe that using extra water will increase the harvest.

Authors in [31] demonstrated how to calculate operational rainwater, crop and reference evapotranspiration, net and gross irrigation water demand, crop growth, and irrigation scheduling using crop water requirement simulation models. Using modeling techniques, the ideal quantity of irrigation water has been determined globally. But generally speaking, numerous studies in previous work highlight the significance and pressing need for an exploration of crop water requirements in light of climate change, taking into account the significance of approximating crop water necessities, particularly for major crops, like cotton, wheat, bananas, and sugarcane in Pakistan. However, precise crop water requirements for these crops in the context of climate change are still unknown.

5. Methodology

The framework for a smart irrigation system for an IoT network for agricultural fields is presented in this section and may be seen in Figure 2. Real-time data sets, water pH, TDS, temperature, and turbidity sensors, as well as Arduino, centralized cloud storage, ML methods, and mobile applications, make up the framework's major elements.

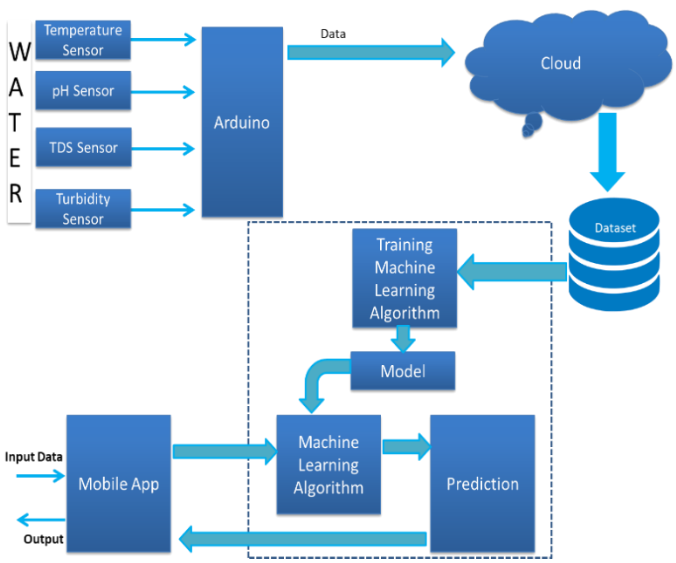

Fig. 2. Framework for Water Quality

5.1 Main Components of Proposed Framework

5.1.1 Arduino UNO WiFi REV2

Arduino UNO WiFi Rev.2 is an easy way to begin with basic IoT with its typical UNO small-factor design. If we want to create a sensor network associated with our home or office router or Bluetooth low-energy device that sends data to cell phones, Arduino UNO WiFi Rev.2 is the one-stop solution for several simple IoT solicitation situations. By including this board in the device, we can use its protected ECC608 crypto chip accelerator to connect it to the WiFi system. Arduino Uno WiFi has WiFi, Bluetooth, and other improvements while otherwise functioning identically to Arduino Uno Rev3. It contains an integrated IMU (Inertial Measurement Unit) LSM6DS3TR and uses Microchip's brand-new ATmega4809 8-bit microcontroller. A Wi-Fi Module is an integrated SoC that can function as an access point or provide access to a Wi-Fi network. It has a built-in TCP/IP protocol stack. It is NINA-W102 from U-Blox. Arduino UNO WiFi Rev.2 contains a USB port, a power jack, an ICSP header, six analog inputs, 14 digital input/output pins, five of which can be used as PWM outputs, and a reset button. All things required to support the microcontroller are incorporated. To begin, just plug in a USB cord to the PC, or power it with a battery or AC adapter [32].

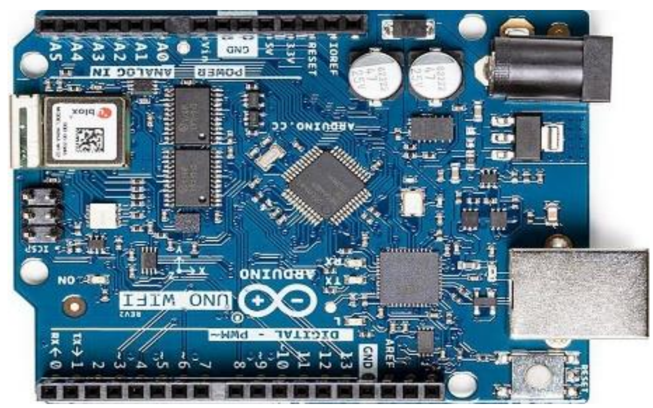

Fig. 3. Arduino UNO WiFi REV2

5.1.2 Temperature sensor DS18B20

The water's warmth or coldness is gauged by its temperature. Since temperature directly distresses the quantity of dissolved oxygen (DO) that aquatic organisms can access, it is a critical water quality metric. The types of aquatic organisms that can persist in water can also be determined by temperature measurements [33].

Fig. 4. Temperature sensor

5.1.3 Turbidity sensor SKU: B306

Water clarity is determined by turbidity. This measurement decides how many floating particles, like clay, silt, sand, and plant debris are present in water, which has an impact on how much sunlight reaches aquatic plants. When spawning grounds and eggs are covered by soil, excessive turbidity can lower aquatic life's reproductive rates. Nephelometric Turbidity (NTU) Units are used to measure turbidity [33].

Fig. 5. Turbidity sensor

5.1.4 TDS sensor SKU: SEN-0244

TDS is a word used to define the quantity of mineral and salt contaminants in water (TDS). PPM units are used to measure TDS. TDS indicates the number of contaminants per million units of water. For instance, water for drinking should have a ppm of less than 500, and water for agriculture should have a ppm of fewer than 1200 [33].

Fig. 6. TDS sensor

5.1.5 SKU-B305 pH Sensor

The concentration of hydrogen ions in water is gauged using the pH scale. A pH sensor measures the water's acidity or basicity, which has a direct impact on aquatic species' ability to survive. With 7 being neutral, the pH scale ranges from 0 (extremely acidic) to 14 (very basic). Utmost water falls within the 5.5 to 8.5 pH range. Chemicals can dissolve differently in water depending on pH changes. Fish and other aquatic species can die if the pH is too high (below 4) or too low [33].

Fig. 7. pH sensor

5.1.6 Cloud storage

Data about water quality is centralized in the cloud. ML techniques such as SVM, RF, DT, Logistic regression, and Nave Bayes are available on the Firebase cloud. The right amount of data needed for a specific crop is obtained using ML algorithms applied to water quality data, and registered users can access this information using mobile applications. Those who have registered can examine ML predictions. Those who have registered can examine the temperature, TDS, pH, and turbidity of their crop water.

5.2 Machine Learning Algorithms

5.2.1 Support Vector Machine Classifier

SVMs refer to a group of learning algorithms that utilize regression and classification techniques to categorize data configurations. Their objective is to classify new samples by determining which side of a gap they belong to. SVM models are employed to classify the quality of water used in smooth irrigation into different classes as accurately as possible [34].

5.2.2 Random Forests

An effective ensemble learning method used frequently in classification applications is RF. It categorizes using conclusions from uncountable DT it generates during training, where an output of forest is a mode of targeted outputs from each DT. Using random samples of training data, RF creates DT, lowering variance in the final model, improving performance, and preventing overfitting [35].

5.2.3 Logistic Regression

The method used to connect the dependent variable to one or more autonomous variables is known as logistic regression. In some instances, the terms predictor and predictor, respectively, are used to describe both the dependent and independent variables. Variables unrelated to the prediction of plant type include temperature and humidity differences, soil moisture, and pH levels. The formula below has been created [35].

$$Y = B_0 + B_1X_1 + B_2X_2 + \cdots + B_kX_k + \epsilon \quad (1)$$

Where,

Y = Response variable (Predicted variable)

Xi = Independent variable

β0 = Y-intercept (constant)

βi = Slope for each independent variable

ε = error (Residual)

5.2.4 Naive Bayes

The Naive Bayes techniques are a set of supervised learning algorithms that use Bayes' theorem for probabilistic classification. These algorithms assume that the features are independent of each other. Each feature is expected to increase the probability that the trial belongs to a particular class. Although the Naive Bayes models are considered to be among the simplest Bayesian network models, they can achieve high levels of accuracy when used in combination with kernel density approximation. Naive Bayes is a classification algorithm that performs well for both binary and multiclass classification. It performs better in cases where there are definite input variables, especially when dealing with numerical input variables. Naive Bayes is useful for predicting data and making predictions based on previous performance [35].

5.2.5 Decision Tree

DT is a non-parametric supervised learning approach that is utilized for regression and classification applications. It is structured hierarchically and comprises a root node, branches, internal nodes, and leaf nodes. We can assess our possibilities with the use of DT. DTs are great tools for supporting in choosing one course of action over others. They offer a very valuable framework within which we can present options and research the potential results of those options. Using the tree-like model of decisions and their possible results, like outcomes, events, chances, utility, and resource costs, the DT is a decision support tool. A single technique for showing an algorithm that especially uses provisional control declarations is to use this one [36].

6. Simulation and Results

The data set of 9000 water samples was developed for experimental exploration, and real-time data on water quality was used for water quality prediction. Details about the water's temperature, pH, TDS, and turbidity for a particular crop in a particular area are included. The dataset is made up of training data in the amount of 30% and testing data in the amount of 70%. In this study, we only build decision rules using 30% of the training data. Five ML algorithms were utilized in an experimental study: SVM, RF, Linear Regression, Naive Bayes, and DT. The following formulas were employed to calculate accuracy:

6.1 Accuracy

The proportion of redress expectations of indications of the malady to the whole number of inputs [37].

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})} \quad (2)$$

6.2 Precision

The ratio of correctly predicted parameters to the total number of parameters to be classified is called precision [37].

$$\text{Precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})} \quad (3)$$

6.3 Recall

The percentage of all positives to the total number of parameters. It is also known as the true positive rate or affectability. Wherever true positive, TN = true negative, FP = false positive, and FN = false negative [37].

$$\text{Recall} = \frac{\text{TP}}{(\text{TP}+\text{FN})} \quad (4)$$

6.4 F1-Score

An ML evaluation metric known as the F1 score measures the accuracy of models. It is a combination of the model's recall and precision. How many times the model is appropriately predicted during an occupied dataset is determined by accuracy statistics. In maximum cases, an F1 score is more useful than accuracy, mostly if your class is distributed irregularly. When FP and FN costs about similar, accuracy performs best. It is better to embrace both precision and recall if the costs of FP and FN differ considerably [37].

$$\text{F1 Score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (5)$$

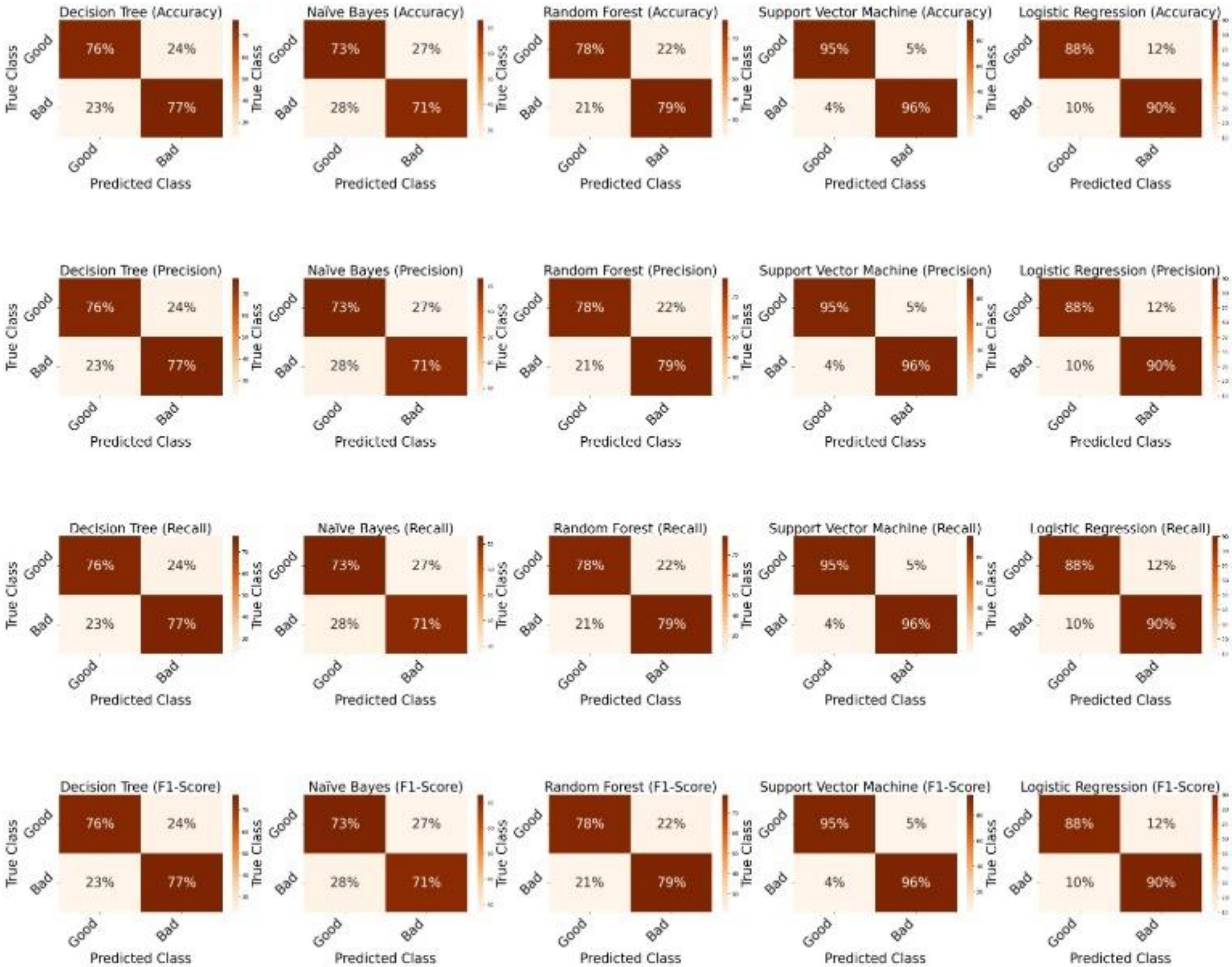

Fig. 8. Confusion Matrix

Accuracy results of ML methods are displayed in Table 1 and the following figures 8, 9, 10, 11, and 12. In this graph, SVM outperforms RF, Logistic Regression, DT, and Nave Bayes in terms of accuracy results for ML algorithms. SVM's result is greater than 90%, while logistic regression's is 88%, and the accuracy results of RF, DT, and naive Bayes are all less than 80%. Table 2 shows the water quality index and how much percent water is healthy for agriculture.

6.5 Decision Tree (Accuracy, Precision, Recall, F1-Score):

The Decision Tree model exhibited an Accuracy of 76%, which represents the proportion of correctly classified instances. Its Precision was 76%, indicating the percentage of true positive predictions out of all positive predictions. The Recall or Sensitivity was 77%, signifying the percentage of true positives correctly identified out of all actual positive cases. The F1-Score, a balanced measure of Precision and Recall, was 76%, showcasing a fair trade-off between precision and recall. The Decision Tree's performance is acceptable but not exceptional, and it offers interpretability.

6.6 Naïve Bayes (Accuracy, Precision, Recall, F1-Score):

The Naïve Bayes model achieved an Accuracy of 71%, suggesting a reasonable rate of correct classifications. Its Precision was 71%, indicating that 71% of the positive predictions were accurate. The Recall, at 71%, shows the model's ability to identify true positives out of all actual positives. The F1-Score, which combines Precision and Recall, was 71%, reflecting a balanced performance between precision and recall. Naïve Bayes is known for its simplicity and efficiency in text and categorical data classification.

6.7 Random Forest (Accuracy, Precision, Recall, F1-Score):

The Random Forest model demonstrated an Accuracy of 79%, indicating a high proportion of correct predictions. The Precision was 78%, denoting that 78% of its positive predictions were accurate. The Recall, at 79%, shows a good ability to correctly identify actual positives. The F1-Score was 79%, signifying a well-balanced performance between precision and recall. Random Forests excel in handling complex datasets and offer robust performance.

6.8 Support Vector Machine (Accuracy, Precision, Recall, F1-Score):

The Support Vector Machine model displayed an impressive Accuracy of 96%, indicating an exceptional rate of correct predictions. Its Precision was 96%, showcasing a high percentage of true positive predictions. The Recall, at 96%, indicates the model's strong ability to identify true positives. The F1-Score, also at 96%, highlights the high balance between precision and recall. Support Vector Machines are powerful but computationally intensive models known for their accuracy.

6.9 Logistic Regression (Accuracy, Precision, Recall, F1-Score):

The Logistic Regression model achieved an Accuracy of 90%, signifying a substantial proportion of correct predictions. Its Precision was 90%, indicating that 90% of its positive predictions were accurate. The Recall, at 90%, reflects the model's ability to correctly identify actual positives. The F1-Score, at 90%, demonstrates a strong balance between precision and recall. Logistic Regression offers a good trade-off between performance and interpretability.

Table 1

Accuracy of ML algorithms

Algorithm	Accuracy	Precision	Recall	F1-Score
Decision Tree	76	75	79	87.34
Naïve Bayes	72	75	70	73
Random Forest	79	79.11	69	80
SVM	94.18	91	90	92
Logistic Regression	88.22	84	85	87

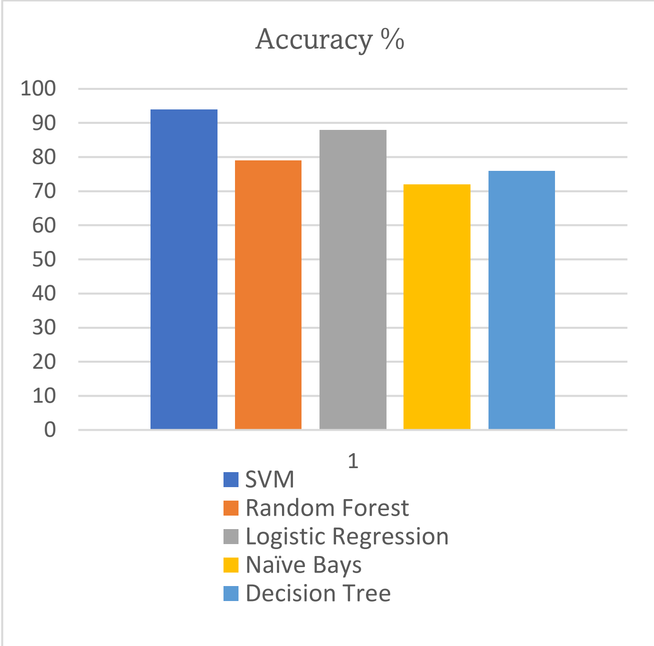

Fig. 9. Accuracy of classifiers

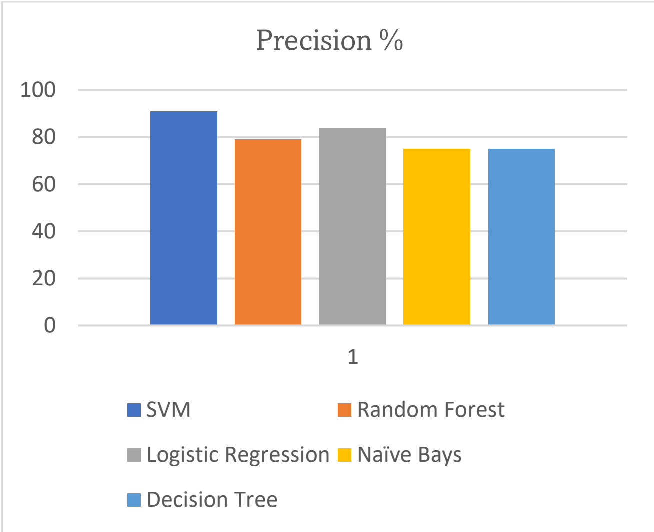

Fig. 10. Precision of classifiers for classification of agriculture fields

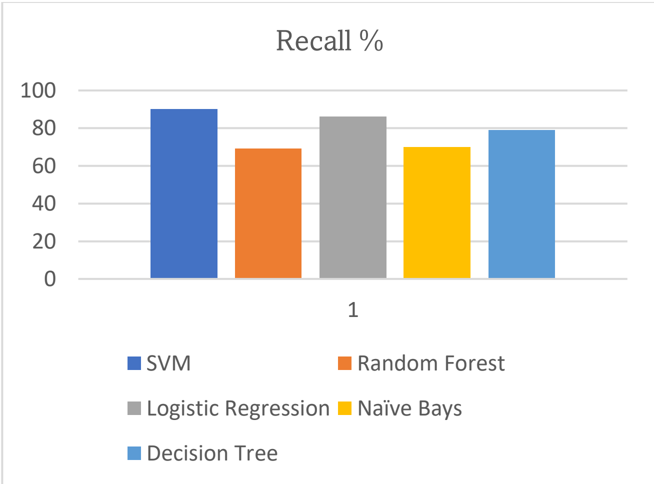

Fig. 11. Recall of classifiers for classification of agriculture fields

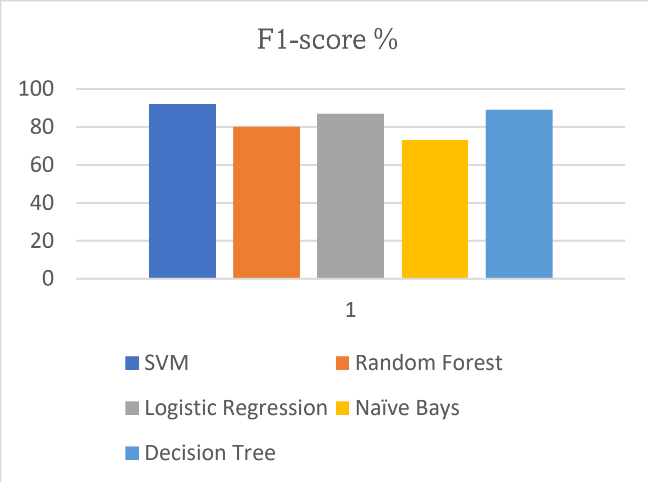

Fig. 12. F1-score of classifiers for classification of agriculture fields

6.10 Water Quality Index (WQI) Calculation

The water quality index can express the overall water quality status in a single term.

Table 2

Water Quality Index Range

Classification	WQI Range	Description
Good	80-100	Suitable for use
Fair	50-79	Minor impairment
Marginal	30-49	At Risk
Poor	0-29	unsuitable

WQI was calculated using Eq. (6) for the WQI assessment method, while the water quality status was assigned using the classification system developed by Brown et al. (1970) [41] [42] in Table 2

$$WQI = \sum_{i=1}^{n} QiWi \quad (6)$$

Here Qi is the sub-index i-th water quality parameter, Wi is the weight of the i-th water quality parameter, and n is the number of water quality parameters.

Table 3

Good and bad water quality on this dataset in terms of WQI

Good Quality	Bad Quality
88.12 (%)	11.88 (%)

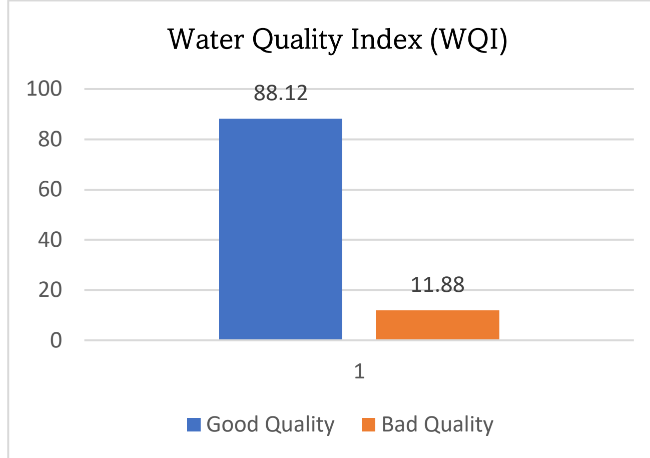

Fig. 13. Good and bad water quality on this dataset in terms of WQI

7. Conclusion

The smart irrigation system described in this article makes use of cloud computing, IoT, and ML frameworks. This study consists of four sensors i-e Temperature, TDS, pH, and Turbidity. ML techniques were employed in this framework to forecast the quantity of usable grade water required for agricultural production. As a result, a quantifiable amount of high-quality water is used for irrigation. Therefore the agricultural industry will alter as a result of intelligent irrigation.

Among the classification algorithms, the SVM demonstrates the highest performance with an accuracy of 95%, precision of 91, recall of 90, and F1-score of 92. The Logistic Regression also performed well with an accuracy of 88%, while other methods, such as RF, Naïve Bayes, and Decision Trees are less accurate.

These findings have significant implications for water management since they allow for proactive decision-making and prompt risk-reduction measures by providing an accurate estimate of the parameters governing water quality. Ensuring that water quality is accurately classified enables efficient monitoring and the timely identification of critical situations that demand prompt care.

8. Future Work

The use of IoT and ML for quality water monitoring has the potential to completely change how we manage and protect our water resources. Here are some potential future routes for this technology, which can gather real-time data from many sources and is one of the primary benefits of IoT-based water quality monitoring. This information can be used to spot problems with water quality as they arise, enabling quick response and correction. Using ML methods, predictive modeling can be used to create water quality models based on historical data.

Autonomous sensor-based quality water monitoring: In remote areas, water quality parameters can be continuously monitored. When it comes to monitoring quality water in hard-to-reach places like deep lakes or secluded rivers, technology can be especially helpful. *Integration with other data sources:* To acquire a more thorough understanding of water quality problems and their causes, water quality monitoring data can be merged with other data sources such as weather data, land use data, or hydrological data.

Creation of decision-support systems for the control of water quality that can be aided by ML algorithms. Water resource managers can use these tools to make well-informed choices regarding strategies for managing and monitoring water quality. Some necessary minor concerns regarding the dependability and security of the various data processing systems and processes can also be addressed.

9. References

[1] M. Atanu, and D. Biswas, "Assessment of drinking water quality using water quality index: a review", Water Conservation Science and Engineering, 2023, 8, no. 1, 6,.

[2] Xu. Dinghong, K. R. Abbasi, K. Hussain, A. Albaker, A. I. Almulhim, and R. Alvarado, "Analyzing the factors contribute to achieving sustainable development goals in Pakistan: A novel policy framework", Energy Strategy Reviews 45, 2023, 101050.

[3] B. Das, S. M. Ali, M. Z. Shaikh, A. F. Chandio, M. A. Rahu, J. K. Pabani, and M. R. Khalil. "Linear regression based crop suggestive system for local pakistani farmers", Global Conference on Wireless and Optical Technologies, 2023, pp. 1-6. IEEE.

[4] G. U. Md, S. Nash, A. Rahman, and A. I. Olbert. "A sophisticated model for rating water quality", Science of the Total Environment, 2023, 161614.

[5] G. S. Lakshmi, P. N. Prasanna, G. Asha, G. Sandhya, S. V. Sharma, S. Shilpashree, and S. G. Subramanya. "An intelligent IOT sensor coupled precision irrigation model for agriculture", Measurement: Sensors 25, 2023, 100608.

[6] R. Hajar, F. Nadi, S. S. Lam, M. I. Abdullah, N. A. Kasan, R. F. Rahmat, and W. A. W. Mahari. "Internet of Things in aquaculture: A review of the challenges and potential solutions based on current and future trends", Smart Agricultural Technology, 2023, 100187.

[7] U. O. Chinanu, and J. C. Egbueri. "Drinking water quality assessment based on statistical analysis and three water quality indices (MWQI, IWQI, and EWQI): a case study", Environment, Development and Sustainability, 2023, vol 25, no. 1, pp: 686-707.

[8] C. Qi, G. Yu, and Z. Qiao. "Application and recent progress of inland water monitoring using remote sensing techniques", Environmental Monitoring and Assessment 195, no. 1, 2023: pp: 1-16.

[9] R. Veerachamy, R. Ramalakshmi, B. P. Kavin, I. Hussain, A. R. H. Almaliki, A. A. Almaliki, A. Y. Elnaggar, and E. E. Hussein. "Exploiting IoT and its enabled technologies for irrigation needs in agriculture", Water vol. 14, no. 5, 2022.

[10] A. A. Mirani, M. S. Memon, M. A. Rahu, M. N. Bhatti, and U. R. Shaikh. "A review of agro-industry in IoT: applications and challenges", Quaid-E-Awam University Research Journal of Engineering, Science & Technology, Nawabshah. 17, no. 01, 2019, 28-33.

[11] U. Md. Galal, S. Nash, A. Rahman, and A. I. Olbert. "Performance analysis of the water quality index model for predicting water state using machine learning techniques", Process Safety and Environmental Protection 169, 2023, pp: 808-828.

[12] E. M. Hamdy, M. Elbagory, A. A. Tamma, M. Gad, S. Elsayed, H. Hussein, F. S. Moghanm, A. El-Dein Omara, A. Kovács, and S. Péter. "Evaluation of groundwater quality for irrigation in deep aquifers using multiple graphical and indexing approaches supported with machine learning models and GIS techniques, Souf Valley, Algeria", Water 15, no. 1, 2023, 182.

[13] E. Naser, and M. Hooshmandzadeh. "Comparison of river water quality assessment methods using the tree model and the nearest neighbor method (A case study: Ahvaz Hydrometric Station)", 2023.

[14] T. Hirofumi. "Water quality monitoring." Analytical Sciences vol 39, no. 1, 2023, pp: 1-3.

[15] G. Fatemeh, and D. Kang. "Application of machine learning in water resources management: a systematic literature review", Water 15, no. 4, 2023.

[16] A. Mohammad, and M. Kheimi. "Stacking Artificial Intelligence Models for Predicting Water Quality Parameters in Rivers", Journal of Ecological Engineering vol 24, no. 2, 2023, pp: 152-164.

[17] A. Chiara, and F. Castelli. "Prediction of the ecological status of surface water bodies with supervised machine learning classifiers", Science of the Total Environment, 2023.

[18] M. J. Kumar. "Application of machine learning algorithms in the investigation of groundwater quality parameters over YSR district, India", Turkish Journal of Engineering, vol 7, no. 1, 2023, pp: 64-72.

[19] M. A. Rahu, S. Karim, R. Shams, A. A. Soomro, and A. F. Chandio. "Wireless sensor networks-based smart agriculture: sensing technologies, application, and future directions", Sukkur IBA Journal of Emerging Technologies Vol 5, no. 2, 2022, pp: 13-32.

[20] F. K. Shaikh, S. Karim, S. Zeadally, and J. Nebhen. "Recent trends in internet of things enabled sensor technologies for smart agriculture", IEEE Internet of Things Journal, 2022.

[21] M. A. Rahu, P. Kumar, S. Karim, and A. A. Mirani. "Agricultural environmental monitoring: a WSN perspective", University of Sindh Journal of Information and Communication Technology 2, no. 1, 2018, pp: 17-24.

[22] M. A. Rahu, "Energy Harvesting for water quality monitoring using floating sensor networks: a generic framework", Sukkur IBA Journal of Emerging Technologies 1, no. 2, 2018, pp: 19-32.

[23] X. Dong, M. C. Vuran, and S. Irmak, “Autonomous precision agriculture through integration of wireless underground sensor networks with center pivot irrigation systems”, Ad Hoc Networks, vol. 11, no. 7, 2013. pp. 1975–1987.

[24] G. Vellidis, V. Liakos, J. H. Andrews, et al., “Development and assessment of a smartphone application for irrigation scheduling in cotton”, Computers and Electronics in Agriculture, vol. 127, 2016, pp.249–259.

[25] C. Goumopoulos, B. O’ Flynn, and A. Kameas, “Automated zone-specific irrigation with wireless sensor/actuator network and adaptable decision support”, Computers and Electronics in Agriculture, vol. 105, 2014, pp. 20–33.

[26] B. M. A. Abubaker, Y. Shuang-En, S. Guang-Cheng, and M. Alhadi, “Impact of different water harvesting techniques on soil moisture content and yield components of sorghum”, Pakistan Journal of Agricultural Sciences, vol. 51, 2014.

[27] J. Gutierrez, J. F. Villa-Medina, A. Nieto-Garibay, and M. A. Porta-Gandara, “Automated irrigation system using a wireless sensor network and GPRS module”, IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 1, 2014, pp.166–176.

[28] G. Vellidis, V. Liakos, J. H. Andrews, et al., “Development and assessment of a smartphone application for irrigation scheduling in cotton”, Computers and Electronics in Agriculture, vol. 127, 2016, pp. 249–259.

[29] M. R. M. Kassim, I. Mat, and A. N. Harun, “Wireless sensor network in precision agriculture application”, Proceedings of the International Conference on Computer, Information and Telecommunication Systems, Jeju, Republic of Korea, 2014.

[30] A. A. Mirani, M. S. Memon, R. Chohan, I. N. Sodhar, and M. A. Rahu. "Irrigation scheduling, water pollution monitoring in IoT: A Review", 2021.

[31] C. S. Gyan, H. R. Mian, S. Mohammadi, M. Rodriguez, K. Hewage, and R. Sadiq. "Digital water: artificial intelligence and soft computing applications for drinking water quality assessment", Clean Technologies and Environmental Policy, 2023, pp.1-30.

[32] Y. Yan, X. Huang, Xi-Qiao Wu, Chao-rong Liu, Shi-yong Zhao, and Xiao-hua Zhu. "The spatiotemporal variations characteristic and pollution evaluation of water quality in Qujiang River, China", Applied Water Science 13, no. 2, 2023.

[33] S. Janjua, I. Hassan, S. Muhammad, S. Ahmed, and A. Ahmed. "Water management in Pakistan's Indus Basin: challenges and opportunities", Water Policy, no. 6, 2021, pp.1329-1343.

[34] N. Sohu, N. A. Zardari, M. A. Rahu, A. A. Mirani, N. H. Phulpoto, "Spectrum sensing in ISM Band Using Cognitive Radio", QUEST Research Journal of Science and Technology, Vol 17 no. 01, 2019, pp.1329-1343.

[35] G. M. Faezeh, T. Rajaee, and M. Z. Kermani. "Decision tree models in predicting water quality parameters of dissolved oxygen and phosphorus in lake water", Sustainable Water Resources Management, vol 9, no. 1, 2023.

[36] D. Pakorn, S. Pinthong, N. Salaeh, J. Weekaew, T. T. Tran, and Q. B. Pham. "Comparative study of machine learning methods and GR2M model for monthly runoff prediction", Ain Shams Engineering Journal, vol 14, no. 4, 2023.

[37] Y. Naima, M. Neji, M. Kallel, A. Wali, and S. Hajji. "A comparative approach of ML algorithms to rank irrigation water quality: the case of Oriental-Coast shallow aquifer in Cap-Bon, northeaster of Tunisia", Modeling Earth Systems and Environment, 2023, pp.1-14.

[38] M. Alexandros K., and B. E. Jimenez-Cisneros. "UNESCO’s contribution to face global water challenges", Water vol. 11, no. 2, 2019.

[39] M. H. Lashari, S. Karim, M. Alhussein, A. A. Hoshu, K. Aurangzeb, and M. S. Anwar. "Internet of things-based sustainable environment management for large indoor facilities", PeerJ Computer Science, 2023.

[40] M. A. Rahu, A. F. Chandio, K. Aurangzeb, S. Karim, M. Alhussein, and M. S. Anwar. "Towards design of Internet of Things and machine learning-enabled frameworks for analysis and prediction of water quality", IEEE Access, 2023.

[41] G. Uddin Md, S. Nash, A. Rahman, and A. I. Olbert. "Performance analysis of the water quality index model for predicting water state using machine learning techniques", Process Safety and Environmental Protection, 2023. pp: 808-828.

[42] M. Marselina, F. Wibowo, A. Mushfiroh, “Water quality index assessment methods for surface water: A case study of the Citarum River in Indonesia”, Heliyon. 2022, Vol 3; 8 (7).

[43] M. H. Lashari, A. A. Memon, S. A. Shah, K. Nenwani, F. Shafqat, “IoT-based poultry environment monitoring system”, IEEE International Conference on Internet of Things and Intelligence System, 2018, pp. 1-5). IEEE.