



Cost-effective IoT-based intelligent irrigation system

C. S. Anagha¹ · Pranav M. Pawar¹ · P. S. Tamizharasan¹

Received: 3 August 2021 / Revised: 12 October 2021 / Accepted: 8 November 2021 / Published online: 28 January 2023

© The Author(s) under exclusive licence to The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2023

Abstract Agriculture contributes to the growth of human civilization. An adequate amount of water (irrigation) is needed for healthy crops and to increase productivity. Water scarcity is a major problem the world faces, where agriculture consumes a significant portion of freshwater. Many researchers concentrate on imparting intelligence in irrigation systems using machine learning (ML) in recent days. With the emergence of Internet of Things (IoT) technology, devices can communicate with each other. It makes systems like IoT and ML a successful solution for precision agriculture to reduce human intervention in plant irrigation. The paper presented a detailed comparative review of state-of-the-art work in the intelligent automated irrigation system, and contributed the IoT based cost-effective intelligent irrigation system. The developed system uses temperature, soil moisture, humidity, and weather forecast data to take intelligent decisions to automate irrigation using an ML algorithm. The proposed system shows 99.6% accuracy for the accurate prediction of soil moisture as compared with state-of-the-art. The proposed system is also cost-efficient in terms of time (by reducing the time require for training a model), and money (by saving power and human labour requirement).

Keywords Agriculture · Irrigation system · IoT · Machine learning

1 Introduction

Agriculture and human civilization have a very tight bond. Cultivation of crops is mandatory for food production, the growth of the country's economy, to enrich the beauty of the land, etc. Plants contribute to natural air purification. Most countries strive to protect and improve their green spaces. However, the issue stems from improper watering and adaptation to changing weather conditions. If plants are not appropriately irrigated, they cannot survive. While water scarcity is a problem that the world faces, a significant portion of freshwater is consumed by agriculture (Mancosu et al. 2015; Vanani et al. 2017). If water wastage could be controlled in agriculture, it is possible to save a substantial percentage of freshwater. Maintenance of soil moisture is essential for the uninterrupted growth of the plant. There are different methods used for irrigation, like sprinklers, drip irrigation, etc. The water should be given so that the plant gets an adequate amount of water for the uninterrupted growth and productivity of the crop. While maintaining soil moisture, moisture levels also are considered. It should not be overwatered or should not be underwater. Soil moisture also depends on the type of soil, humidity, rain, temperature, etc. (Mintz and Walker 1993; Ostad-Ali-Askari et al. 2017). In traditional irrigation systems, humans assess whether the moisture is enough and decide to irrigate the plant. But it will be advantageous to reduce human intervention in watering and monitoring with an automated system. Human labour can be reduced by establishing an automated intelligent irrigation system. The amount of water adequate for the plant can be determined by using an intelligent irrigation system and thus can reduce the wastage of water as well. For this, farmers can seek the help of modern technologies like sensors, the IoT, ML algorithms. These technologies

✉ Pranav M. Pawar
pranav@dubai.bits-pilani.ac.in

¹ Department of Computer Science, Birla Institute of Technology and Science Pilani, Dubai Campus, Dubai, UAE

can contribute to the automation and monitoring of crops without the use of extensive human resources.

With the high usage of computers and smart devices, there exist systems for effective monitoring of crops. There have been studies on crop disease detection and management. When humans started to use modern machinery in agriculture and farming, their labor was reduced by a lot. Even though watering and care at the accurate time is a must in the growth of plants. Watering plants is a time-consuming process. If it could be automated with intelligence, then farmers could utilize their time for other farming activities. A traditional irrigation system does not have intelligence, which will cause wastage of water, or sometimes the plant may not get an adequate amount of water. An intelligent irrigation system can make decisions on its own about when to water and the amount of water required to water the crop (Nawandar and Satpute 2019; Ostad-Ali-Askari and Shayannejad 2020). This system can reduce overwatering and underwatering issues, resulting in fewer plant deaths, root rot issues, and plant drying issues.

Drip irrigation is a water-saving system that slowly sprays water into the soil or plant roots. Plant diseases can be reduced by using this system. Drip irrigation is suitable for all types of soil. The problem with the system is the maintenance of the pipe from clogging. But that is not required at short intervals. With the developments in sensors, IoT and ML, it is possible to innovate a system that can automate and impart intelligence to irrigation in agriculture. The sensor can be used to check water moisture in the soil, temperature, humidity, etc. Climate, humidity, and soil temperature are all factors that influence soil moisture. This data can be used to prevent water wastage by calculating the optimum requirement for water. M2M allows devices to communicate with each other without human intervention. IoT consists of intelligent devices connected to the web. IoT uses embedded systems, such as processors, sensors, and communication hardware, to gather, transmit and act on data they obtain from their surroundings. These devices can send data from a sensor to another edge device or a gateway (Al-Fuqaha et al. 2015). ML programs are concerned with how to improve a model with experience automatically. ML algorithms can learn from data and make predictions based on that data.

The first contribution of the paper is a detailed comparative review of state-of-the-art techniques in the last five years (2017–2021) for the automated intelligent irrigation system. The study is performed concerning mechanism used, advantages, disadvantages, ML algorithms used, dataset/data collection methodology, performance measures, and accuracy percentage. It provides a brief idea to a researcher about state-of-the-art techniques and research horizons in the intelligent irrigation system. The second significant contribution of the paper is the IoT-based intelligent irrigation system. In this paper, the research collected data such as temperature,

humidity, and soil moisture of the plant and online weather data and trained the models and used them to predict the soil moisture after a particular amount of time. This system is an automatic watering system that waters based on the plant's water requirements, reducing human intervention and eliminating water waste and power consumption while ensuring continuous growth. The proposed research predicts the water requirements with regression models, K-nearest neighbor (KNN), Support Vector Machine (SVM) and Random Forest (RF). The experimentation shows that KNN and RF show the accuracy of prediction up to 0.996 (99.6%), which offers a good contribution to the agriculture field using Machine Learning algorithms. The proposed system is considered to be cost efficient as it won't waste water as it will only water the plants when required. The system will also use less power since it is using drip irrigation, which is also controlled wastage of water. The motor won't work to irrigate extra water and thus will save power. The sensors will also work in sleep and listen mode, which also helps in saving power. Thus, overall system will be cost efficient by reducing power requirement, human labour, and money in terms reducing water and power bills.

The paper provides a detailed literature review of state-of-the-art working automated intelligent irrigation systems in the following section of this paper. Section 3 consists of the implementation and architecture of our system. Section 4 includes the proposed algorithms and ML algorithm, which is used in the current work. Section 5 discusses the obtained results, and Sect. 6 contains the conclusion and future work.

2 Comparative review of state-of-the-art

Many recent studies have been conducted on automated intelligent irrigation systems using different methodologies used. The section discusses major state-of-the-artwork from (2017–2021) in the automated intelligent irrigation system with comparative analysis. The comparative review of state-of-the-artwork is mentioned in Table 1.

In the paper (Goldstein et al. 2018), they collected soil moisture from the plot and temperature data to forecast the irrigation plan for the upcoming week and construct an irrigation strategy weekly for plots, using ML algorithms. They used regression and classification models. Compared to the agronomist's prediction, a linear regression model did not have a high success rate, but the regression tree model, known as gradient boosted regression trees (GBRT), resulted in 93% accuracy and boosted tree classifiers (BTC) resulted in 95% accuracy. GBRT is regarded as one of the most efficient and admired prediction models, with exceptionally high accuracy. But GBRT-based regression trees are challenging to interpret since they include fifty decision trees.

Table 1 Comparative review of state-of-the-art work in automated intelligent irrigation system

Ref. no.	Dataset/data collection methodology	Algorithm	Performance measures	Accuracy (%)
Goldstein et al. (2018)	Soil moisture data collected from uMANAG2. Soil moisture collected from different plots using sensors. Meteorological data collected for temperature, humidity, and solar radiation	GBRT, BTC	Accuracy	GBRT: 93 BTC: 95
Mota et al. (2018)	Sensor data of soil moisture of both irrigated and non-irrigated trees. Climate conditions. Three leaf and stem moisture data of both irrigated and non-irrigated trees	Linear(L) and polynomial(P) regression	R ²	L: 38 P: 43
Goap et al. (2018)	Sensor data (soil moisture, air temperature, air relative humidity, soil temperature, and radiation) of the field. Weather forecast data collected from internet	SVR followed by k-means clustering	Accuracy	96
Ebin et al. (2019)	Data from Temperature-, soil-moisture-, and light-intensity-sensors. Past and current weather data	ID3	–	–
Kondaveti et al. (2019)	Soil moisture sensor data, humidity sensor data	Romyan's algorithm	–	–
Blasi et al. (2021)	Sensor data of soil moisture, temperature. Data from Directorate of Agriculture of the Southern Jordan Valley. Data about the weather from Taqs Al-Arab website and the Department of Meteorology	DT	Accuracy	97.86
Velmurugan (2020)	Soil moisture and temperature data from the sensors	Tree-based ML algorithm	Accuracy	98
Shekhar et al. (2017)	Soil moisture and temperature data collected using sensors	K-NN	–	–
AlZu'bi et al. (2019)	Soil moisture, humidity and temperature data. IP camera data of images of leaves and soil cracks	SVM, NN, RF	Accuracy	SVM: 95 NN: 96 RF: 83
Mahajan et al. (2018)	Sensor data of soil moisture, Field parameters. Theta dataset	Clustering	–	–
Nawandar and Satpute (2019)	Soil moisture, humidity and temperature data collected from deployed sensors	NN, Bayesian	MSE	37.7
Alipio et al. (2019)	Sensor data of light intensity, pH, electrical conductivity, water temperature, and relative humidity	Bayesian	Accuracy	84.53
Rawal (2017)	Soil moisture data collected using Sensor	No ML algorithm	–	–
Sidhartha et al. (2021)	Sensor data of soil moisture, temperature, and humidity	Customized algorithm	–	–
Ramya et al. (2020)	Soil moisture, humidity, temperature, and UV radiation data from sensors. Weather data	SVR and bagging algorithm	R ²	98

The work in Mota et al. (2018) discusses the problems faced by chestnut trees in the summer. These trees need

optimum irrigation for fruit productivity. They found the relation between soil, tree, and water. They found that the

photosynthetic productivity of chestnut trees is dependent on soil moisture. The data set used for the study are sensor data of soil moisture of both irrigated and non-irrigated trees, climate conditions, tree leaf and stem moisture data of both watered and not-watered trees. The regression model uses for the photosynthetic rate and mid-day stem water relationship.

The paper (Goap et al. 2018) discusses soil moisture, soil temperature, air temperature, ultraviolet (UV) light radiation, and relative humidity of the crop field. The algorithm considers data from sensors and the weather forecast data for the near future to impart intelligence to the system. The paper points out the importance of weather forecasting in irrigation systems. Prediction of future Soil Moisture Differences (SMD) is made using a trained Support Vector Regression (SVR) model. The output value of SMD is provided as input to the k-means clustering to improve the prediction result of the difference in soil moisture. The soil moisture difference is taken as the centroid value of the k-means. The result of the k-means clustering is used to make the irrigation decision. A real-time monitor was implemented for irrigation management. The system will start and stop irrigation automatically by considering threshold values for starting and stopping. The algorithm has a 96% accuracy rate.

This research (Ebin et al. 2019) developed an automated irrigation system that monitors past weather conditions, soil moisture, and current weather conditions. The system can reduce human intervention in agriculture by imparting proper irrigation methods using ML algorithms. The ID3 algorithm learns according to data from temperature sensors, soil moisture sensors, and light intensity sensors, in addition to weather data. The ID3 algorithm is resistant to outliers and missing values. Also, ID3 requires less cleaning of data than other algorithms. However, the performance of the external data is inadequate.

The paper (Kondaveti et al. 2019) describes the contribution to protecting agriculture in urban areas by introducing an automated irrigation system. This algorithm uses a rainfall prediction algorithm to predict crops suitable for a specific location. They introduced Romyan's method to determine when to switch on irrigation, thus reducing electricity and water usage. The paper gives a brief description of different types of soils and how they differ in retaining moisture. They use a linear regression model to predict rainfall using parameters mean temperature, max temperature, min temperature, mean humidity, sea level pressure, and wind speed. The rainfall prediction gives almost high accurate results but more complex for real-time implementations.

The problems faced in the agriculture sector in the Jordan valley is discussed in Blasi et al. (2021). The research examines the artificial intelligence-based solution for a water irrigation system to solve the problem of the agriculture sector. Sensor data is used to construct the intelligent irrigation

system using a traditional drip irrigation system with a pipe control mechanism and an ML algorithm for automation. Selected information (Soil Humidity, Soil Type, Soil Salinity, and Temperature) from the data set collected from the Directorate of Agriculture of the Southern Jordan Valley was analyzed using python programming. Data is pre-processed from numerical to categorical data to apply the decision tree (DT) algorithm. After training and testing, the data from sensors is used to determine the irrigation needs. Humidity levels and temperature also decide whether to stop or start irrigation. In this work, a high accuracy (97.86%) was achieved using the simple DT algorithm.

The work done in Velmurugan (2020) is used to solve the water wastage in Indian agriculture and avoid unplanned water usage causes a decrease in ground-level water. This system uses data from the sensors and weather forecast data from the internet. The gateway collects sensor data and sends it to a base station. This data can be visualized and analyzed using server-side software.

The paper (Shekhar et al. 2017) discusses traditional irrigation systems, ICT-based agricultural monitoring, the scope of machine learning in agriculture, and IoT based irrigation systems. Using an intelligent method, they are trying to irrigate by taking data like soil moisture and temperature to predict the soil type and turn on the water accordingly. The details of soil conditions and the amount of water irrigated are saved in the cloud and assessed by the farmer using his cell phone. The system uses an Arduino as a microcontroller and a Raspberry Pi3 as a processing unit. The KNN algorithm used for this work proposed to give higher accuracy.

This study (AlZu'bi et al. 2019) uses the Internet of Multimedia Things (IoMT) and ML algorithms to construct an intelligent irrigation system. This system can reduce water wastage with intelligence applied to the system. The system consists of a set of hardware units along with software programs. The Arduino MEGA 2560 is a microcontroller board to link sensor data. The ESP8266 chip is used to collect information from the microcontroller to the server. Soil moisture sensors, humidity, and temperature sensors are used to collect environmental data. To avoid water wastage during the rainy season, a raindrop sensor was added to the system. A relay is used to switch on and off the water pump. The Waikato Environment for Knowledge Analysis (WEKA) is used for classification. IP cameras have been used to capture soil cracks and the yellowing of leaves to add to the decision about watering, along with other sensor data. The classification algorithms used are SVM, Convolutional Neural Networks (CNN), and RF. Precision in the NN is 0.96. SVM yields a score of 0.95. Recall for Neural Network was 0.80. When it comes to high accuracy, the Neural Network is the way to go. However, when it comes to training time, SVM is the best option.

The paper (Mahajan et al. 2018) describes a prototype for an automated, IoT and ML-based irrigation system that uses sensor data. The irrigation is done by collecting sensor data such as moisture, temperature, age of the plant, type of the plant, soil, etc., of the plant. A gateway captures the sensor values and sends them to a Raspberry Pi control unit with KNN algorithm. Instead of scheduling the irrigation periodically, the system will predict the condition of the soil for watering based on sensor data and qualified data and apply the ML algorithm. They selected water scheduling parameters like temperature, soil CO₂, and humidity sensor values. They used this dataset to implement the Multifractal Downscaling Model algorithm. This gives a novel approach by resolving heterogeneity in soil moisture.

The contribution discussed in Nawandar and Satpute (2019) uses IoT and neural networks (NN) to develop a low-cost intelligent system for smart irrigation farming. To make decisions about future irrigation schedules using learning algorithms, ground surface evaporation, transpiration from leaves, stems, flowers, etc., is used. The experiment took user data as input for the estimation of the required amount of water. The planting area can be divided into zones according to the type of crop or moisture. The moisture sensor is placed in areas where the average humidity is present. If various modes of irrigation can be imparted, it will help when unpredictable weather changes occur. This drip irrigation system can save up to 67% of water by using the NN method.

The work (Alipio et al. 2019) talks about site-specific farming. This method continuously monitors farm-specific soil or plants to enhance the agriculture process. The work integrates IoT, sensors, Data Analytics, and a Web interface. Sensors are integrated into hydroponic farms to monitor the parameters used for the hardware. The software includes a cloud server, analysis of data, and ML predictions. The application developed to control and monitor the farm parameters using sensors and actuators. The sensor values were collected and built into a Bayesian Network to classify and predict the ideal value of each actuator and autonomously control the hydroponic farm. The model achieves an accuracy of 84.40%. “Think to Speak” IoT platform used for the dashboard. The problem with this kind of system is the high cost.

Another study in Rawal (2017) uses a microcontroller ATMEGA328P on the Arduino platform as a control unit. They put a system that uses a GSM-GPRS SIM900A modem, and a web page was created to check the farmer’s sprinkler status. The GSM modem is used to transfer data obtained from sensors to the internet. They aim to avoid over-and under-watering the plant, using a little human intervention. The author used THINGSPEAK open data platform and API to analyze the data and automatically respond to sensor results. Thus, the sprinkler can be on and off until the

desired moisture is obtained. Farmers can check the status and control the water by switching it on or off using an IoT-based system. In this work, no ML algorithm is applied to incorporate intelligence into the system.

Due to water scarcity, farmers turn to groundwater resources, resulting in a reduction in ground-level water (Sidhartha et al. 2021). The lack of human labor for agriculture in cities highlights the need for an automated system. They made a computerized system using sensors that collected temperature, moisture, and humidity data. The system uses a DC-operated fan to manage the temperature. The Arduino gets the sensor data and decides to switch on or switch off the motor. To get the time RTC timer circuit connected with Arduino. The data is uploaded to a platform via XBee for real-time monitoring. This system is low-cost irrigation management. The drawback of this system is the lack of use of the ML algorithm to impart intelligence.

This work (Ramya et al. 2020) consists of five stages: data acquisition, online weather data collection, soil prediction, real-time monitoring utilizing a web interface, and an IoT based motor controller. Soil moisture, humidity, temperature, and UV radiation are all published by the ESP32 node. Data collected using specific sensors, and a Raspberry Pi node collects data using the Message Queuing Telemetry Transport (MQTT) cloud service. Sensor data, along with weather data from the farm location, is used to calculate the evapotranspiration rate using the Penman-Monteith model, contributing to the water consumption rate. It is kept on a server. Soil moisture is predicted using the Bagging ensemble learning model. They are using sensor data, weather data, and evapotranspiration rate. The idea of the Bagging ensemble model is different samples (with replacement) fit on several models of the dataset. Then, these models are added by applying the average, weighted average, or a voting system. A web interface is used to set the user’s required water threshold value and monitor real-time data. The Python code is used to start and stop the relay switch of the water pump automatically. Initially, the bagging algorithm was applied to differentiate between real and the predicted soil moisture, compared to the expected value using SVR and bagging methods. The results show the R² value of the Bagging model is highest and the mean square error (MSE) of the Bagging model is more minor, while the R² value of the SVR is less compared with the bagging model and the MSE of the SVR model is high.

The detailed comparative review shows that automated intelligent irrigation system improve the performance of precision agriculture, but it is also necessary to improve the system in terms of cost (time and money), in order to reach the technology at grass-root level. There is also need to improve the mechanism in terms complexity of system, accuracy, and reducing training time of ML algorithms.

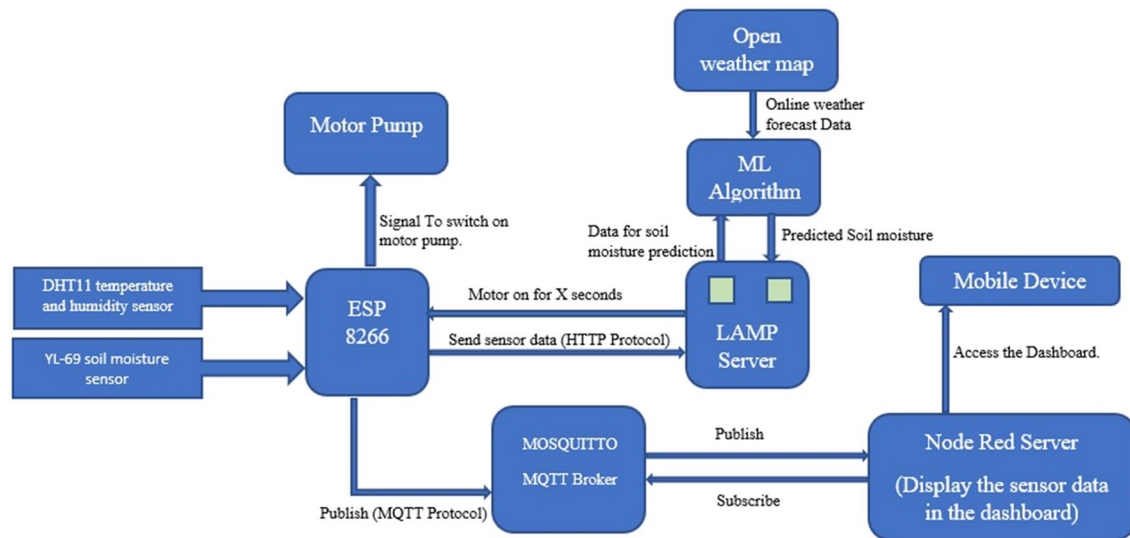


Fig. 1 Proposed system architecture and data flow

The proposed work make a trial to develop cost-efficient automated intelligent irrigation system by considering water scarcity problem in UAE.

3 Proposed system methodology

The proposed system makes use of IoT and ML algorithms. This system will help reduce water wastage while providing irrigation for the flawless growth of the crop. The system will decide when to irrigate and switch on the motor pump automatically without human intervention by reducing human labor by watering.

3.1 System architecture and workflow

To implement an intelligent irrigation system, first collected the temperature, soil moisture, and humidity data from the field using sensors. The proposed work exploits the DH11 sensor to collect temperature and humidity and the YL-69 sensor to collect real-time soil moisture from the field. Figure 1 shows overall architecture of system and the data flow, which is explained above.

The ESP8266 microcontroller reads the data collected by the sensor. The collected data is sent to the Linux, Apache, MySQL, and PHP (LAMP) server using the Hypertext Transmission Protocol (HTTP) and the data is stored in a MariaDB database. At the same time, the microcontroller publishes the sensor data to the Mosquitto MQTT Broker. The Node-Red server is used to collect the real-time sensor data from the MQTT broker, and it was displayed on a dashboard on the server. This dashboard can be accessed via any

smart device so that it is possible to monitor the real-time sensor data using a mobile phone or laptop. The collected real-time data is used to estimate the future soil moisture. This prediction utilizes online weather forecast data to get the maximum temperature, minimum temperature, and average temperature. The proposed work uses OpenWeatherMap API for collecting the forecast data.

The trained ML model is used to predict the future soil moisture using the dataset. This predicted soil moisture value is used to determine whether the motor pump should be turned on. The water needed is calculated and the time to keep the motor on to pump the desired amount of water is determined. This information is sent to the ESP8266 microcontroller, and the microcontroller will send the signal to switch on the motor accordingly. Figure 2 shows the experimentation setup for the experiment. The experimentation is performed in controlled environment during month of March. The plant used during experimentation is Fenu-greek plan, which is a fast growing plant. The plat is planted balcony of lab. The balcony roof was covered, so plant can get a controlled sunlight and the change in temperature and humidity outside will be same for the plant also. The collected information (temperature, humidity, and soil-moisture) on Node Red Server is visualized using simple user interface on mobile device. Figure 3 shows the dashboard view from mobile device before and after watering.

3.2 Data collection and processing

The data is collected via sensors. A temperature-, humidity—and soil-moisture-sensors (DH11 sensor for temperature and humidity, YL69 for soil-moisture) are used to collect the data of the plant's context. These collected data

Fig. 2 Experimentation setup: **a** Circuit setup. **b** Over all experimentation

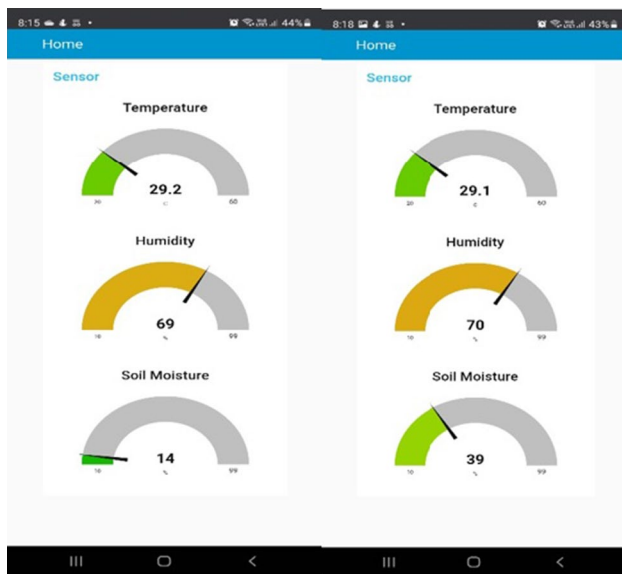
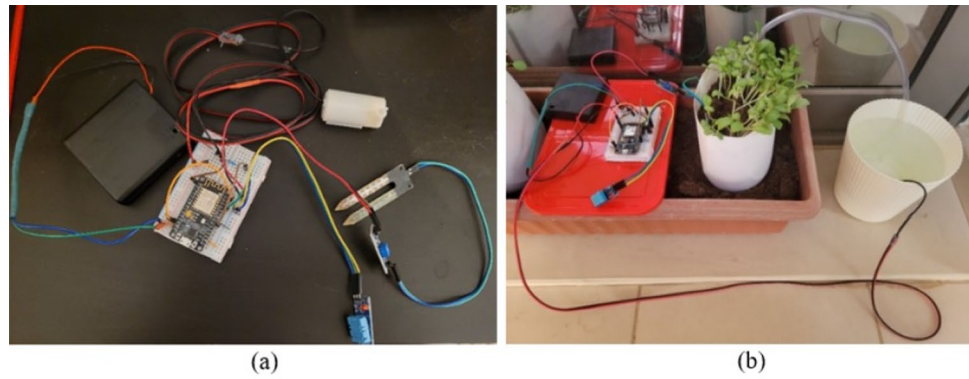


Fig. 3 Dashboard view: **a** Before watering. **b** After watering

are sent to the server. The collected data is combined with the online weather forecast data of maximum temperature, minimum temperature, and average temperature to make a one record. Each field of the collected records is verified to ensure that all required values are collected and then inserted into the database. This data is collected every 5 min, and the soil moisture after 20 min from the point of data collection recorded. The data is collected for thirty days. The data collected via sensors and the online weather data is combined to create data-set for training ML model. The dataset includes:

- Current soil moisture data.
- Humidity.
- Current temperature.
- The maximum temperature of the day
- Minimum temperature of the day.
- An average temperature of the day.

- Soil moisture (moisture after 20 min at the current time).

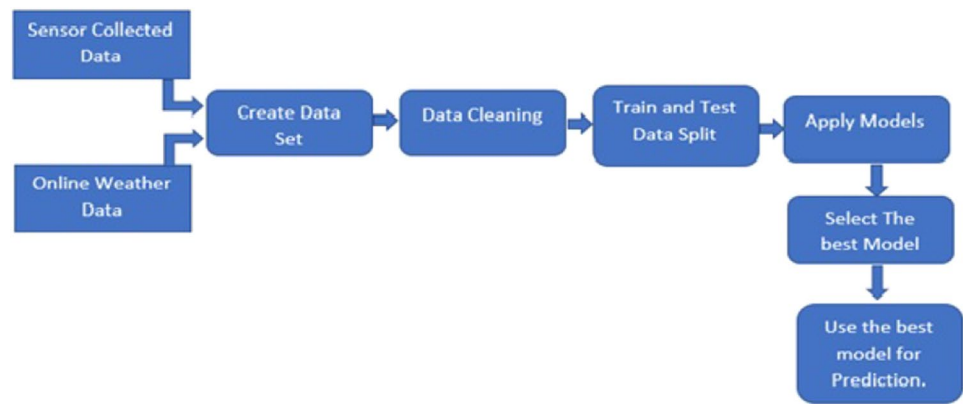
The dataset is cleaned by eliminating rows that contains unknown values and only preserving valued datasets. The dataset consist of total 9000 records.

3.3 Model selection architecture for soil moisture predictions

The different ML models were used to select the best model for soil moisture prediction. The work used KNN for prediction in this project, which gives the best R^2 -value, MSE, and MAE (Mean Absolute Error) values compared to other experimented models. The current soil moisture data, humidity, current temperature, maximum temperature of the day, minimum temperature of the day, and an average temperature of the day are used as features to predict the Next soil moisture. The predicted soil moisture is used to switch off and switch on the motor pump and calculate the amount of water to be irrigated. The pump will be on till the calculated amount of water is pumped, and then it will go off. The user can monitor the plant's current temperature, humidity, and soil moisture using his handset. This system can detect the remaining water in the tank using a water level detector and message the user about the water requirements. The user will not need to bother about the plant's irrigation as the weather changes since the system automatically detects the water needs and serve the purpose accordingly.

The automatic irrigation system needs to implement automatic prediction of soil moisture to calculate the crop's water needs. This can be done with the help of ML algorithms. Since the value of soil moisture to be predicted is a continuous value, regression models are chosen to predict the soil moisture value. The different steps included in the prediction of soil moisture are depicted in Fig. 4. The created dataset of 9000 records is used for training ML models. The collected data is split for testing and training using the `train_test_split` function with random state 42 of the model selection package from the sklearn library. The dataset is

Fig. 4 Model selection architecture for soil moisture prediction



split into 80% for training-data and 20% for testing-data. Using 80% of the training set, the model is trained with different ML models. The trained models are tested with the remaining test data set. The models were cross-validated for 10-folds to get a better validation score. In this paper, KNN is used for prediction and thus for decision making on automated watering, as it shows maximum accuracy after testing.

4 Algorithm

The Algorithm 1 for data collection is shown below, it is used to collect required data from sensors and weather

data from weather forecast portal, and form the dataset for applying ML models. First, the data collected for current soil moisture (CSM), temperature (ST), and humidity (SH) from sensors are stored in structure record sensor data (SD) with Time. Similarly, MaxTemp (Maximum temperature), MinTemp (Minimum temperature) and AvgTemp (Average temperature) are collected from online weather forecast portal and stored in structure record weather forecast data (WD). The data from sensors and weather forecast portal is combined and stored in structure record collected data (CD) if none of the value from SD and WD is NULL. The data is collected after every 5 min and written to the database CD.

Algorithm 1 DATA COLLECTION

- 1: **procedure** DATACOLLECTION()
 - 2: Initialize sensor data (SD) = {ST, SH, CSM, Time}
 - 3: Initialize online weather forecast data (WD) = {MaxTemp, AvgTemp, MinTemp} from weather forecast web portal.
 - 4: **if** SD and WD does not contain NULL value **then**
 - 5: Combine SD and WD as CD.
 - 6: Store collected data in database, CD = {ST, SH, CSM, MaxTemp, AvgTemp, MinTemp}
 - 7: **end if**
 - 8: **end procedure**
-

A new column, “Next Soil Moisture” (NSM), is created from soil moisture value after 20 min. After data collection, the next step is to choose an ML algorithm to train and predict the soil moisture with the collected dataset. The ML algorithms considered are Regression models of KNN, Linear Regression (LR), SVM, and RF using Python.

KNN is a method for classifying cases depending on their similarity to other cases. Cases that are close to each other are referred to as neighbours. KNN calculates the distance of unknown data for all cases, selects the k observations in the

training data nearest to the unknown data point, and predicts the response using the most popular response value from their K nearest neighbours. The value of K is determined by a trial-and-error method. The Makowski distance is given by equation 1. Here, x_i 's are the input vectors and y_i 's are the output vectors and c is the content.

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^c \right)^{\frac{1}{c}}$$

Next, another algorithm considered is the Multiple Linear Regression Algorithm to train the model. Multiple Linear Regression is an extension of Linear Regression. MLR is an arithmetic method that uses several independent variables to predict the consequence of a dependent variable. The calculation formula is given in equation 2.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon$$

where for $i = n$ observations, y_i = dependent variable, β_0 = y-intercept (a content term), β_i = slope intercept for independent variables, x_i = independent variables, ϵ = residual term or error term for the model.

Support Vector Machine is an important supervised learning algorithm. SVM works by choosing a hyperplane by increasing the distance from the closest instance of both classes to the hyperplane. The support vectors are a small subset of training data that are used to define the optimal hyperplane.

$$\arg \min_{w,b} \frac{1}{2} \|w\|^2 A = \pi r^2$$

subject to

$$y_i(w \cdot x_i + b) \geq 1, i \in [1, n]$$

The Random Forest Classifier is an ensemble classifier. The RF consists of many decision trees and outputs a class that is the combined mode of the outputs of individual trees. The process by which RF randomly chooses each individual tree

from the database with replacement is called bagging. Overfitting may happen if more iterations are performed with the Random Forest. RF will choose random vectors to create multiple decision trees.

All the above algorithms are modelled using 10-fold cross-validation (as stated in previous section) and found the average score of R^2 value, MSE value, and MAE value. The experimentation performed found that the KNN could more accurately predict soil moisture than other models by considering the metrics. The other models also work well with our dataset. The results are discussed in the result section.

Then the next step was to use the KNN model to implement an automated intelligent irrigation system. The advantage of the KNN algorithm is that it requires no training before making predictions, and new data can be added effortlessly without affecting the algorithm's accuracy. Moreover, KNN is extremely easy to implement. But they are prone to outliers or missing values. So, to make sure the data does not contain any missing values, the data written on the server is checked for correctness. After deciding to use the KNN model for prediction, the model is made to predict NSM and used this prediction to make decisions for automated watering.

The Algorithm 2 for making decision regarding watering a plant is given and explain below. The algorithm 1 works on the data collected using DataCollection() procedure mention in Algorithm 1. It applies KNN model on collected data to predict the NSM column. If predicted NSM value is less than minimum soil moisture, then system calculate the water requirement and pump run-time, and turn on the motor accordingly for watering a plants.

Algorithm 2 AUTOMATED WATERING

- 1: **Begin**
 - 2: Set min and max soil moisture
 - 3: Call procedure DataCollection() for collecting data
 - 4: Apply KNN model to predict NSM.
 - 5: **if** NSM \leq min soil moisture **then**
 - 6: Calculate Water Requirement (WR)
 - 7: Calculate pump run-time in seconds (X)
 - 8: Turn on the motor for X seconds
 - 9: **end if**
 - 10: **end**
-

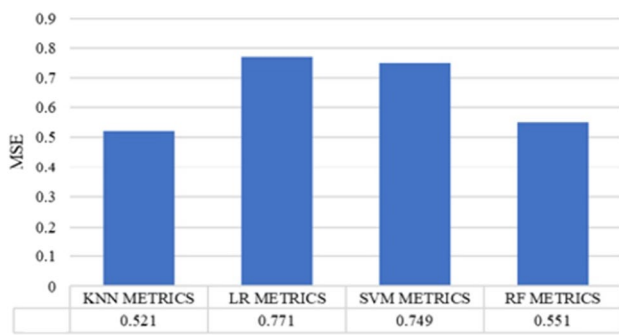


Fig. 5 Comparison of MSE values for different algorithms

5 Results and discussion

The data set used in this work consists of the features ‘curr_temp’, ‘curr_humidity’, ‘curr_soilmoisture’, ‘max-temp’, ‘avgtemp’, and ‘mintemp’ as independent variables and ‘Next_SoilMoisture’ as a dependent variable. The ML Algorithms used for the experiment were Regression models of KNN with four as the optimum K value, Linear Regression (LR), SVM, and RF using Python. As mentioned in previous sections, the dataset is divided into an 80% train set and a 20% test set and all the models were cross-validated for ten folds. The R^2 value, MSE and MAE were obtained to compare the models. All results are shown in Figs. 5, 6 and 7.

The comparison of MSE values is given in Fig. 5. The mean square error is the average of the square of the variation between the actual and predicted values. The experimentation obtained MSE of 0.521 for KNN after 10 -fold cross-validation. Similarly, MSE for LR, SVM, and RF are 0.771, 0.749 and 0.551, respectively. The MSE is close to zero, which means the model is a perfect fit. So, a model with less MSE is chosen. The work chosen the KNN model based on these metrics.

Next, the work also found the MAE values for all the models. The comparison of MAE values is shown in Fig. 6. The Mean Absolute Error is an absolute value

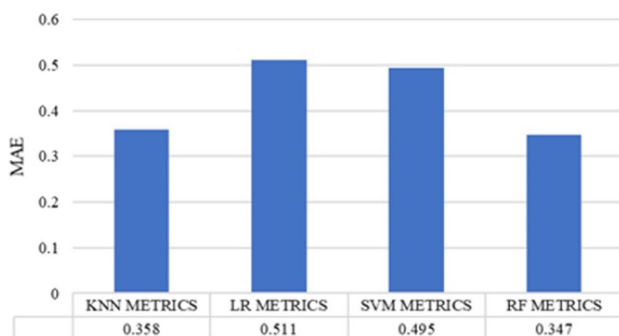


Fig. 6 Comparison of MAE values for different algorithms

that is used to compute the median. This metric is resistant to outliers. As the MAE value is lower, the model is better. Figure 6 shows the MAE value obtained for different algorithms using our dataset. RF gives the least value of MAE at 0.347. Next, KNN gives 0.358. The LR and SVM MAE values are a little higher. But both are in range 0.5. This indicates the algorithms are fit for our dataset.

The other metric used is the R^2 value. It indicates how the model will forecast perfectly future samples. The maximum possible score is 1. The R^2 value must be close to 1 to select a model for our prediction. Figure 7 shows the comparison of R^2 values for different models. All the models gave us an exceptionally good score even though KNN and RF received the highest score of 0.996.

The KNN is chosen to predict soil moisture after comparing the above three metrics to all of the algorithms. Table 2 shows the predicted values using KNN and other algorithms on some of the test datasets. There is negligible difference between the predicted soil moisture values and actual soil moisture values. These values indicates that, this prediction model can be used without failure.

After selecting the KNN model, this model is used to predict soil moisture, and the predicted value is used to determine whether the motor pump should be turned on or off. The signal to switch on and the time duration to keep the motor on till the desired amount of water is pumped is sent to Arduino. Then, Arduino will start the pump accordingly. From the above discussion, it is evident that the ML algorithms could contribute more to the enrichment of agriculture in future.

The Table 3 also shows the comparison of proposed work with state-of-the-art work in Blasi et al. (2021). The experimentation performed in Blasi et al. (2021) is in middle-east region, which is considered as important reason to select the work for comparison. The comparison shows that the proposed approach shows better accuracy than (Blasi et al. 2021). The reason of improved accuracy is size of dataset used, where proposed model in paper is trained on 9000 records in comparison of model training on 1498 records in

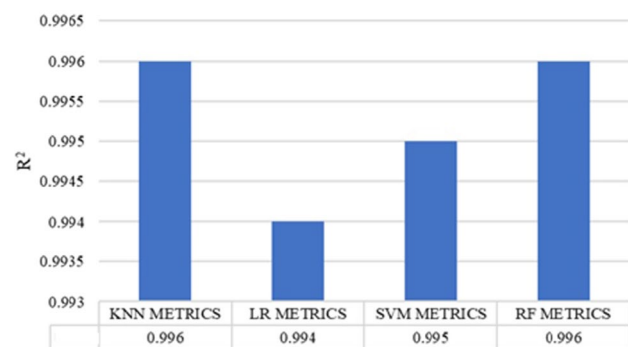


Fig. 7 Comparison of R^2 values for different algorithms

Table 2 Actual and Predicted soil moisture values

Actual soil moisture values	Predicted soil moisture values			
	KNN	LR	SVM	RF
52	52	51.807	52.096	52
59	59.35	58.961	59.435	59.31
52	52.25	51.973	52.081	52.12
20	20	19.759	21.348	20
46	46	45.577	45.592	45.21
62	62	61.662	61.874	61.98
62	61.5	61.938	61.434	61.3
47	47.25	47.591	47.774	47.09
42	43.25	42.602	42.464	42.98
46	46.25	46.956	46.936	46.08
20	20.25	20.743	21.829	20
34	34	33.693	33.17	34.04
49	49	48.911	48.996	49

Table 3 Comparison with state-of-the-art

Model	Prediction accuracy
Proposed model	99.6
AlZu'bi et al. (2019)	97.86

Blasi et al. (2021). The proposed algorithm (using KNN) is also cost-effective in terms time, as required training time is less and updating a existing model if new data-points is added in dataset is also minimum. The decision-tree based algorithm (in Blasi et al. 2021) require more training time and adding new data-points is time-intensive as whole tree needs to be rebuild again with respect to new data point.

6 Conclusion and future work

The research proposed a cost effective IoT-based intelligent irrigation system. With this system, the amount of water wasted for irrigation purposes can be reduced, since this system will not overwater the crop. This system will also help to maintain the needed soil moisture for the crop, thus making for the flawless growth of the crop. Since there is no need for human intervention, this model can help to reduce farmers' work to a great extent. The system checks soil moisture every 20 min, so any changes in the soil moisture can be detected as soon as possible and the decision to water. The developed system is said to be intelligence as it predict

the soil moisture using ML algorithm. It will help not to waste water unnecessarily, and also shows efficiency compared to the earlier automatic timer-based irrigation system. The developed system is also cost effective in terms of time and money, and it also shows 99.6% prediction accuracy. The developed system can be extended to predict the other requirements such as the need for minerals, light etc. In the future, it is possible to create smart farms using different ML techniques and produce more fruitful farming with less wastage of water. It is also possible to make more accurate decisions by using image processing and deep learning techniques.

Funding Not applicable.

Declarations

Conflict of interest No any conflict of interest for the article.

Research involving Human Participants and/or Animals Not applicable.

Informed consent No any informed consent for the article.

References

- Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of things: a survey on enabling technologies, protocols, and applications. *IEEE Commun Surv Tutor* 17(4):2347–2376
- Alipio MI, Cruz AEMD, Doria JDA, Fruto RMS (2019) On the design of Nutrient Film Technique hydroponics farm for smart agriculture. *Eng Agric Environ Food* 12(3):315–324
- AlZu'bi S, Hawashin B, Mujahed M, Jararweh Y, Gupta BB (2019) An efficient employment of internet of multimedia things in smart and future agriculture. *Multimed Tools Appl* 78(20):29581–29605
- Blasi AH, Abbadi MA, Al-Huweimel R (2021) Machine learning approach for an automatic irrigation system in southern Jordan valley. *Eng Technol Appl Sci Res* 11(1):6609–6613
- Ebin PM, Kavitha NR, Mathew JK (2019) Automated irrigation system based on machine learning concept volume 8, no. 2, March–April 2019 international journal of information systems and computer sciences. <http://warse.org/IJISCS/static/pdf/file/ijiscs02822019.pdf>
- Goap A, Sharma D, Shukla AK, Krishna CR (2018) An IoT based smart irrigation management system using Machine learning and open source technologies. *Comput Electron Agric* 155(41–49):2018. <https://doi.org/10.1016/j.compag.2018.09.040>
- Goldstein A, Fink L, Meitin A, Bohadana S, Lutenberg O, Ravid G (2018) Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge. *Precis Agric* 19(3):421–444
- Kondaveti R, Reddy A, Palabtlala S (2019) Smart irrigation system using machine learning and IOT. In: 2019 international conference on vision towards emerging trends in communication and networking (ViTECoN), pp. 1–11. IEEE

- Mahajan Y, Pachpande M, Sonje R, Yelapure S, Navale M (2018) Effective use of water in irrigation system using machine learning. *Int J Curr Res* 10(01):63873–63875
- Mancosu N, Snyder RL, Kyriakakis G, Spano D (2015) Water scarcity and future challenges for food production. *Water* 7(3):975–992
- Mintz Y, Walker GK (1993) Global fields of soil moisture and land surface evapotranspiration derived from observed precipitation and surface air temperature. *J Appl Meteorol Climatol* 32(8):1305–1334
- Mota M, Marques T, Pinto T, Raimundo F, Borges A, Caço J, Gomes-Laranjo J (2018) Relating plant and soil water content to encourage smart watering in chestnut trees. *Agric Water Manag* 203:30–36. <https://doi.org/10.1016/j.agwat.2018.02.002>
- Nawandar NK, Satpute VR (2019) IoT based low cost and intelligent module for smart irrigation system. *Comput Electron Agric* 162:979–990
- Ostad-Ali-Askari K, Shayannejad M (2020) Impermanent changes investigation of shape factors of the volumetric balance model for water development in surface irrigation. *Model Earth Syst Environ* 6(3):1573–1580. <https://doi.org/10.1007/s40808-020-00771-4>
- Ostad-Ali-Askari K, Shayannejad M, Eslamian S (2017) Deficit irrigation: optimization models. *Management of drought and water scarcity, handbook of drought and water scarcity*
- Ramya S, Swetha AM, Doraipandian M (2020) IoT framework for smart irrigation using machine learning technique. *J Comput Sci* 16(3):355–363
- Rawal S (2017) IOT based smart irrigation system. *Int J Comput Appl* 159(8):7–11
- Shekhar Y, Dagur E, Mishra S, Sankaranarayanan S (2017) Intelligent IoT based automated irrigation system. *Int J Appl Eng Res* 12(18):7306–7320
- Sidhartha G, Dwarakanath ML, Vineeth KC, Sukruth R (2021). IoT based Greenhouse Irrigation System. *Int J Res Appl Sci Eng Technol (IJRASET)*, 9(III)
- Vanani HR, Shayannejad M, Tudeshki ARS, Ostad-Ali-Askari K, Eslamian S, Mohri-Esfahani E, Jabbari H (2017) Development of a new method for determination of infiltration coefficients in furrow irrigation with natural non-uniformity of slope. *Sustain Water Resour Manag* 3(2):163–169
- Velmurugan S (2020) An IOT based smart irrigation system using soil moisture and weather prediction. *Int J Eng Res Technol (IJERT)*. ISSN:2278-0181

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.