ISSN [Online]: 2583-2654

Prediction of Water Requirement for Mixed Crops Using Machine Learning Algorithm

Dr. Geetha C. Megharaj, Prof. Thara Devi M, Chandralekha V, Indhu S, Janani J, Naveena B
Department of CSE, Dr. T. Thimmaiah Institute of Technology, KGF, Karnataka, India

Abstract: The existing plant watering systems are often unreliable and not accurate. There is a possibility of using incorrect data when stress levels of plants are considered for watering and capturing the images of the field is expensive. To overcome this, we have proposed a system called Automatic Plant Watering System (APWS). System aims to automatically water varieties of plants without human interruption with machine learning techniques. The workflow of plant watering system automatically includes four stages likely data collection, pre-processing, training and classification. In the first stage, the data is generated on the threshold value of the variety of plant and used in training the K-means technique and the hardware components are assembled i.e., Arduino sensors, relays and pumps are connected. In the second stage, the data read from the sensors are programmed using microcontroller that controls the pumps. Following the previous stage, third stage is about analyzing the data collected from the microcontroller and train K-means model to classify water requirements. The final stage is about testing the model, fine-tunning and deploying our system for plant care.

Keywords: Automatic Plant Watering, K-Means, Arduino, Sensors, IoT, Machine Learning

I. INTRODUCTION

An automatic irrigation scheduling based on thermal stress has been investigated to characterize plant water stress relationship and to estimate crop. Varying tolerances to temperature and humidity levels have been recorded by the plant species. A single thermal stress content may not account for these variations, leading to inaccurate assessments of stress for certain plant species. The threshold values used in thermal stress indices are often determined empirically and may vary based on the source or the specific plant species. Using incorrect threshold values can lead to inaccurate assessments of stress levels. Irrigation System based on smart Phone applications have connectivity issues: smart irrigation systems rely on stable internet connections. If there are connectivity issues or outages, it may disrupt the system's functionality. Users may need technical knowledge to troubleshoot and address issues that arise. Ensuring compatibility between different components and the mobile application can be, especially it using products from different manufactures. Setting up an image processing system for agriculture can be expensive. Costs may include acquiring high-resolution cameras, specialized software, and computing equipment. Image sensors requires calibration accuracy. Failure to calibrate the sensors or neglecting maintenance can result in inaccurate data and flawed irrigation decisions .With the help of moisture level of soil and temperature sensor, we aim to water plants only when it is needed, avoiding overwatering. This precision helps conserve water resource. Automatic watering systems can be customized to meet the specific moisture amount of water for its optimal growth and health. By combining moisture of soil and data of temperature with automated watering, these systems offer a holistic approach to plant care, promoting water efficiency, environmental sustainability and healthier plant growth.

II. RELATED WORKS

In [1] the authors Kasara Sai Pratyush Reddy et al., have implemented model , which predicts the water

ISSN [Online]: 2583-2654

requirement for a crop. It uses decision-tree algorithm on the data sensed from the field to predict results. The result obtained through decision-tree algorithm is sent through a mail alert to the farmers which help in performing action with respect to supply of water, farmer has to manually on or off the water pump for watering to crops i.e., human interruption is required.

The authors M Senthil Vadivu et al., in [2] have provided a system which is designed to reduce water usage and provide optimal irrigation to plants. The system also includes a mobile application that allows farmers to monitor the system remotely and control the motor pump from their smart phones.

Narmadha S et al., in [3] have analyzed the current state of post-harvest loses, the hard benefits of IoT. The tool was developed for soil data analysis in agriculture. The used method has vital implications for the agricultural industry and can help stakeholders in the sector to make actions by inheriting IoT based solutions to reduce post-harvest losses.

In [4] the authors Muhammad Ibrar et al, have implemented system that has the ability to inform farmers about the water levels in the irrigation channel in real time. They have suggested a smart irrigation system based on the IoT using the information of the blockage of the irrigation channel. This is a smart, real time, easy to use, and inexpensive method of water management which opens up opportunity for the government and business to invest in their research work.

The authors Sameer Quazi et al., in [5] have provided a system called IoT-Equipped and AI-Enabled next generation smart agriculture, gives detailed tutorial on available smart agriculture systems through IoT and AI techniques, a critical review of these two technologies and discussion about future trends. They have given the importance of smart agriculture practices with growing gaps in global food demand verses current food generation.

Dhanashri Jadhav et al., in [6] have done a work , that explains the how the irrigation system can be used to reduce water consumption. Their work supports more water controlling for agricultural land and reduces power consumption.

In [7] the authors Keyurbhai A.Jani et al., have represented frame work is to monitor various forms of sensors for IoT and make decisions based on analysis of sensors data. Their suggested frame work takes care of most of the requirements of crops during the development phase. They suggested a model to monitor IoT sensor devices that are low cost, which collect data from soil, air, water, insects, and make appropriate results based on analysis of sensors data.

The authors Susan A. O'Shaughnessy et al., in [8] have provided a system that uses CWSIT-TT method for characterizing plant water stress and guiding irrigation schedules.

Amarendra Goap et al., in [9] , have presented a smart system based on an open-source technology to predict irrigation needs of the field. The required output was achieved by considering various parameters.

In [10] the authors Anat Goldstein et al., have represented a system were, the sensors will collect the data for monitoring plant needs in real-time. Based on these data, a weekly irrigation plan is defined by the company's agronomist. The dataset was constructed by integrating data collected from 22 soil sensors spread in four major plots.

III. PROPOSED METHOD

A Proposed System Architecture

The proposed APWS system aims to automatically water varieties of plants without human interruption by the form of dynamically identifying number of clusters based on K means clustering technique. It consists of several stages collection of data, preprocessing of data, training and classification. The stages involved in the proposed system are illustrated in Fig 1.

ISSN [Online]: 2583-2654

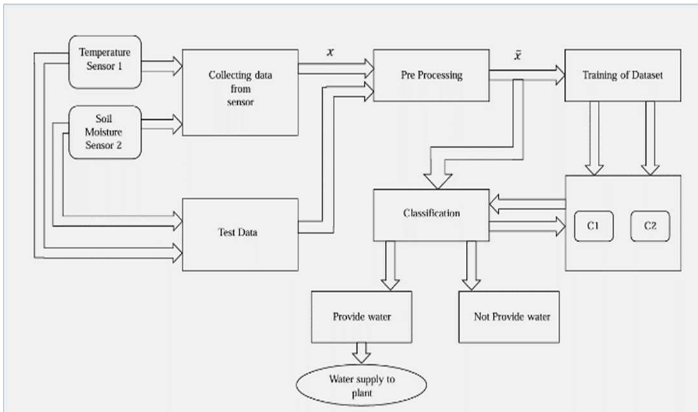

Fig 1: Proposed System Architecture

B Data Flow Diagram

The system is designed to detect prediction of water requirement for various plants which is shown in Fig 2 with specific data flow .First, the automated system acquires the input data set from the soil moisture and temperature sensors. These data are then pre-processed which includes normalization and data cleaning. The pre-processed data set is used for training the machine learning algorithm.

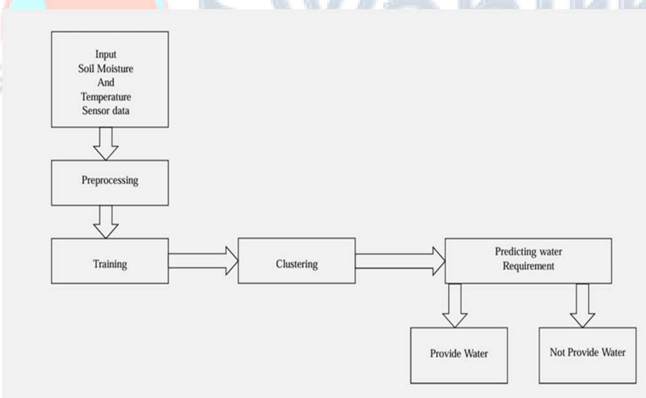

Fig 2: Dataflow Diagram for Proposed System

C Data Generation

The hardware components such as Arduino Mega, sensor to detect moisture of soil, sensor to know the temperature, Relay Modules and water pumps are connected. The Arduino is connected with all the sensors which in turn is connected to Arduino IDE and relay control is set up for the water pumps. The board is programmed to read sensor data and establish serial communication to send sensor data to a computer which is used during classification. It also implements control logic to turn water pumps set or reset based on incoming commands. The data is generated for the moisture of soil and temperature sensors based on threshold of variety of plants i.e Ragi, Rose and Pothos. This generated data is provided the next stage.

D Pre-Processing

The Data preprocessing refers to the steps and techniques applied to the raw data before it is fed into system getting to know algorithms for training and testing. The aim of pre-processing is to collect sensors data that may have missing values, outliers or repeated data. Hence, we will first pre-process the collected data. The dataset is the input data X = {x0,x1,……,xn}. Let xi be a value in input dataset X such that X = xi for xi = xij such that xij is the jth feature of xi. where, X is the input dataset with n objects, n denotes the number of objects in X, i = 1,2, 3,…….,n, j = 1,2, 3,…….,d, xi is the ithrecord of object in X, d is the dimension of the object. The steps to be taken in preprocessing are: 1) Normalization 2) Data Cleaning.

1. Normalization

Normalization is a data preprocessing technique used to scale numerical values within a specific range, usually between 0 and 1. Normalization is essential in various data analysis and machine learning tasks to ensure the different features. Normalization is a crucial step in data preprocessing, ensuring that numerical data falls within a consistent range. Normalization helps in preventing definite features, making the data more suitable for machine learning algorithms. The below given formula rescales each value of X.

$$X_{normalized}=X-X_{min}/X_{max}-X_{min} \quad \ldots\ldots\ldots\ldots(1)$$

Where, X is the original value, Xmin is the minimum value of the feature, Xmax is the maximum value of the feature.

2. Data cleaning

Data cleaning refers to the process of identifying and correcting errors, inconsistencies and inaccuracies within a dataset to improve its quality and reliability for

ISSN [Online]: 2583-2654

analysis such as holding missing values that is dropping rows or columns with missing values, eliminate duplicates which means identifying and eliminate duplicate entries in dataset, consistency checks that is verifying data for consistency and rectifying any discrepancies.

Algorithm:

Input: X with n objects

Output: $\bar{X}$ with m objects | {m ≠ n, m < n}

Begin

1. Check for null values in the dataset, if xi is equal to Null then delete it.
2. Check whether the values lie in the defined range.
3. If xi is less than minimum threshold, then remove it
4. If xi is greater than maximum threshold, then remove it End

E Training

The training phase in machine learning involves teaching a model to make predictions or perform a task by exposing it to a labelled dataset. The intention is to educate a model that could make correct predictions on new, unseen facts via way of means of gaining knowledge of styles and relationships from the training set. In this stage, the proposed APWS system iteratively splits the input plant monitoring dataset ($\bar{X}$) into K dissimilar clusters based on K-Means technique and it represents each individual cluster into separate plant monitoring data for analysis process. The clustering stage consists of four steps, in the first step, it select the k centroid vectors $\bar{V}=\bar{vr}$, for r = 0,1,…..,k, $\bar{V} \subseteq \bar{X}, \forall \bar{vr} \in \bar{X}$ over the actual plant monitoring dataset $\bar{X}$ = xi for xi = xij , for i=0,1,…..,n, j=0,1,2,…..,d based on predetermined knowledge, where k denotes the number of clusters or centroid values, $\bar{V}$ is the centroid vector set with k vectors, vr denotes the rth centroid vector with d plant monitoring data which belongs into the centroid vector set $\bar{V}$, $\bar{X}$ is the plant monitoring data vector set with n vectors and d represents the size of the vector.

In the second step, it measures the dissimilarity $D(\bar{X},\bar{V})$ between the plant monitoring data set vector set $\bar{X}$ = xi ,for i=0,1,…..,n and centroid vector set $\bar{V} = \bar{vr}$ for r = 0,1,…….,k based on Euclidean distance and is defined in the Equation (2).

$$D(\bar{X}, \bar{V}) = \{\{d(\bar{X}ij, \bar{V}rj)_{r=0,i=0,j=0}^{k-1,n-1,d-1}\}\forall \bar{vr} \in \bar{V}, \bar{V} \subseteq \bar{X}, \forall xi \in \bar{X}, \forall xij \in xi\} \ldots\ldots\ldots\ldots(2)$$

Where, $d(\bar{X}ij, \bar{V}rj)$ denotes the Euclidean distance between jth plant monitoring data of the rth centroid vector and ith input plant monitoring dataset and is defined in the equation (3).

$$d(xij, vrj) = \{[\sum_{j=0}^{d-0} (xij - vrj)^2]^{1/2}\} \ldots\ldots\ldots(3)$$

Next step, it place the each individual ith vector xi in the $\bar{X}$ into their closed centroid vector $\bar{vr}$ of its respective cluster in the cr cluster set C= cr for r = 0,1,….,k based on the highest similarity and is defined in the equation (4)

$$cr = Min\ \{\{d(xi, \bar{vr})_{r-0,i=0}^{k-1,n-1}\}, \forall cr \subseteq \bar{X}\} \ldots..(4)$$

Where, C denotes the predetermined cluster with k clusters, cr represents the rth cluster in the cluster set C and in the last step, it updates the rth centroid vector value $\bar{V}=\bar{vr}$ of each individual cluster C= cr based on vectors in their respective rth cluster and is defined in the equation (5) as

$$\bar{vr} = \{\{1/|cr| \times \sum_{e=1}^{|cr|} cre\}, \forall cre \in cr\} \ldots\ldots\ldots(5)$$

Where, | cr | denotes the number of similar plant monitoring data vectors and cre represents the eth vector in the rth cluster which belongs into cluster set C. The process of updating centre is iterated until a situation where centres do not change anymore or criterion function becomes lesser and the clustering stage algorithm is described in the below sub section.

K-Means Algorithm

Input: Plant monitoring dataset $\bar{X} = xij$ with n vectors, centroid vector set $\bar{V}=\bar{vr}$ with k vectors

Output: Produced k distinct clusters C = {c1,c2,……,ck} Begin

1. Randomly select k number of centroid vectors over

ISSN [Online]: 2583-2654

the plant monitoring dataset $\overline{X}$ with n vectors and assigned into $\overline{V}=\overline{vr}$ for r = 1,2,3,……..,k

2. Iteratively computes the Euclidean distance between centroid vector set $\overline{V}=vr$ monitoring data vector $\overline{X}=$ xi using Equations (2) and (3)¯and plant

3. Place the each individual vector in the plant monitoring data vector set $\overline{X}=$ xi into its closest cluster centroid cluster based on higher similarity using Equation (4)

4. Update the each individual rth cluster centroid $(\overline{vr})$ based on elements or vectors of the rth cluster by Equation (5)

5. Repeat the steps from 2 to 4 until the present iteration cluster centroid is similar to previous iteration.

End

F Classification

The process of classifying input data into predetermined classes or categories based on its attributes is referred to as the classification stage in machine learning. Learning a mapping between input feature and class labels is the aim in order to allow the model to predict the class of incoming, unseen data. The kmeans model is fed labeled data during the training phase, where each data point is linked to a predetermined class label. The linkages and patterns between the input feature labels and their matching class labels are taught to the kmeans model. The class labels of fresh, unseen data points can be predicted by the kmeans model once it has been trained. This stage is commonly known as the inference stage. The model predicts the class label by using the features of the new data item as input. There are two types of classification tasks: binary, which involves only two classes, and multi-class, which involves more than two classes. This is a binary classification that determines, in accordance with the needs of the plants, whether to switch on or off the motor.

IV. EXPERIMENTAL SETUP

The Fig 3 shows our water requirement prediction system's top view for mixed plants i.e., ragi, rose and pothos. Our system used Arduino Mega as a microcontroller, Temperature and soil moisture sensors along with relay and pumps to water the plants as per the individual plant requirement.

Fig 3: Water Requirement Prediction System Top View

The Fig 4 shows our water requirement prediction system's front view for mixed plants i.e., ragi, rose and pothos. Our system used Arduino Mega as a microcontroller, Temperature and soil moisture sensors along with relay and pumps to water the plants as per the individual plant requirement.

Fig 4: Water Requirement Prediction System Front View

V. RESULTS

The below Fig 5 shows the graphical interface of main page. This consists of a dropdown menu for selecting the available COM ports, a button to update COM ports in case they are not shown and a connect button. This also displays the type of plants used.

ISSN [Online]: 2583-2654

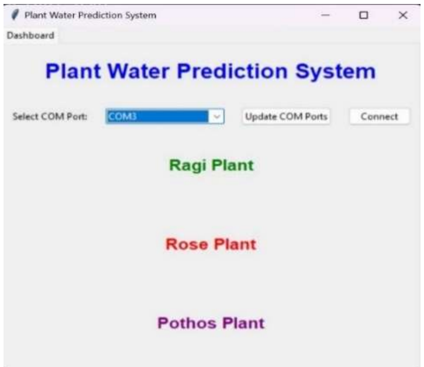

Fig 5: Graphical Interface of Main Page

The Fig 6 shows the data collected by the temperature and soil moisture sensors for ragi, rose and pothos which are represented as 1, 2 and 3 respectively. This data is taken by the trained kmeans model to predict the time to turn on the motor for each plant.

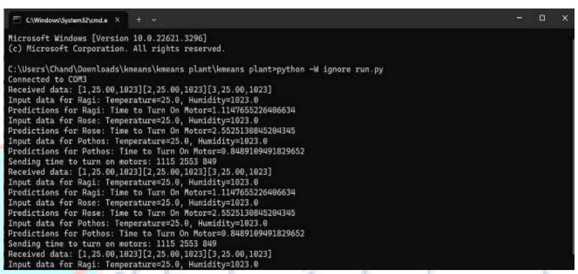

Fig 6: The Data Taken by Sensors and Predictions to Turn on Motor for Plants

VI. CONCLUSION

The aim of the project is to create a system that will automatically water varieties of plants without human interruption using machine learning techniques. The proposed system (APWS) works in this aspect to achieve the aim of the project. In initial stage, we have done with hardware setup. Provided sensors are used to collect the data. The Arduino Mega microcontroller is used to collect the data from the sensors in turn which is connected to Arduino IDE platform to access the stored data. Preprocessing is done for all the plant datasets and observed that there are no missing values. We have implemented K-Means clustering for training purpose and have classified the water requirements of each plant. We have also developed the graphical interface of main page. This consists of a dropdown menu for selecting the available COM ports, a button to update COM ports and a connect button. This also displays the type of plants used. Then one of the available COM ports are selected in order to make a connection with the Arduino Mega by clicking on the connect button. This facilitates to collect the sensors data for three types of plants and to make a decision for watering plants. The decision of predicting amount of time to turn on motor is made by analyzing temperature and soil moisture sensors data by a kmeans model. This turns on the pump according to the requirements of the respective plants.

REFERENCES

[1] Kasara Sai Pratyush Reddy, Y Mohana Roopa, Kovvada Rajeev L N and Narra Sai Nandan, "IoT based Smart Agriculture using Machine Learning", Proceedings of the Second International Conference on Inventive Research in Computing Applications, DOI: 10.1109/ICIRCA48905.2020.9183373, pp. 130-134, Sep-2020.

[2]M Senthil Vadivu, M Purushotham Reddy, Kantilal Rane, Narendra Kumar, A Karthikayen and Nitesh Behare, "An IoT-Based System for Managing and Monitoring Smart Irrigation through Mobile Application",Journal of Machine and computing, Vol.3,no:3 DOI:10.53759/7669/jmc202303018, pp. 196-205, July - 2023.

[3] Narmadha S, Sheeja S,Mohana Priya T,Rajesh Kanna R, Cecil Donald5, Suresh K and Abdalla Ibrahim Abdalla Musa, "IOT Based Smart Agriculture To Avoid Post Harvest Losses Using Machine Learning Algorithms", Journal of Survey in Fisheries Sciences, Vol. 10, no: 4, DOI: 10.17762/sfs.v10i4S.1163,pp. 1130-1138, April -2023.

[4] Muhammad Ibrar, Javed Iqbal Bangash, Muhammad Tariq, Iqtidar Ali, Sana Zahir and Nida Gul, "Smart Irrigation Management System for Water Channels Using IoT", 1st International Conference on Computing Technologies, Tools and Applications, pp. 84-89, July 2023.

[5] Sameer Qazi, Bilal A. Khawaja and Qazi Umar Farooq,"IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends", IEEE Access, Vol.10, DOI:10.1109/ACCESS.2022.3152544,pp.21219-21235, Feb 2022.

[6] Dhanashri Jadhav, Chaitali Jundhare, Shreya Koshti, Vaishnavi Unde and D. P. Mahurkar, "Automated Smart Irrigation System using IoT", International Journal of Innovative Research in Science, Engineering and Technology,Vol.11,no:5,DOI:10.15680/ijirset.2022.1105159, pp. 5518-5520, May 2022.

[7] Keyurbhai A. Jani and Nirbhay Kumar Chaubey, "A Novel Model for Optimization of Resource Utilization in Smart Agriculture System Using IoT (SMAIoT)", IEEE Internet of Things Journal, Vol. 9, no: 13, DOI: 10.1109/jiot.2021.3128161 , pp. 1-9, Nov 2021.

[8] Susan A. O'Shaughnessy, Steven R. Evett, Paul D. Colaizzi and Terry A. Howell, "A Crop Water Stress Index and Time Threshold for Automatic Irrigation Scheduling of Grain Sorghum", USDA-Agriculture Research DOI:10.1016/j.agwat, pp. 122-132, Jan 2012.

[9] Amarendra Goap, Deepak Sharma, A.K. Shukla and C. Rama Krishna, "An IoT Based Smart Irrigation Management System using Machine Learning and Open Source Technologies", CSIR-Central Scientific Instruments Organisation, Vol. 155, DOI: 10.1016/j.compag.2018.09.040, pp. 41-49, Dec 2018.

[10] Anat Goldstein, Lior Fink, Amit Meitin, Shiran Bohadana, Oscar Lutenberg and Gilad Ravid, "Applying Machine Learning on Sensor Data for Irrigation Recommendations: Revealing the Agronomist's Tactic Knowledge", Springer Science and Business Media New York, Vol.19, DOI:10.1007/s11119-017-9527-4, pp. 421-444, May 2017.

[11] Et-Taibi Bouali, Mohamed Riduan Abid, El-Mahjoub Boufounas, Tareq Abu Hamed and Driss Benhaddou, "Renewable Energy Integration into Cloud and IoT- based Smart Agriculture", IEEE Access, Vol.10,DOI:10.1109/ACCESS.2021.3138160,pp.1175-1191, Dec 2021.

[12] M Gayathri, D Arun Shunmugan and A Ishwariya, "Smart Irrigation System using IoT", International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences, Vol. 9, no: 3, DOI: 10.37032/IJIRHPS.2021.V09103.027,pp. 149-154, June 2021.

[13] Youness Tace, Mohamed Tabaa, Sanaa Elfilali, Cherkaoui Leghris, Hassna Bensag and Eric Renault, "Smart Irrigation System Based on IoT and Machine Learning", Elsevier, Vol. 8, no: 9, DOI: 10.1016/j.egyr.2022.07.088, pp. 1025-1036, May 2022.

[14] Kamal Elhattab, Karim Abouelmehdi and Abdelmajid Elmoutaouakkil, "Internet of Things(IOT) for Smart City, Agriculture and Health care", Journal of theoretical and applied information technology, Vol. 100, no:4, pp. 1104-1112, Feb 2022.

[15] Amine Dahane, Rabaie Benameur and Bouabdellah Kechar, "An IoT Low-Cost Smart Farming for Enhancing Irrigation Efficiency of Smallholders Farmers", Wireless Personal Communications, Vol. 127, DOI: 10.1007/s11277-022-09915-4, pp. 1-38, August 2022.

[16] Archana Gupta, Dharmil Nagda, Pratiksha Nikhare and Atharva Sandbhor, "Smart Crop Prediction using IoT and Machine Learning", International Journal of Engineering Research and Technology, Vol. 9, no: 3, pp. 18-21, 2021.

[17] Sameer.M. Patel, Mittal. B. Jain, Sarvesh.S. Pai and Sagar. D. Korde, "Smart Agriculture using IoT and Machine Learning", International Research Journal of Engineering and Technology, Vol. 8, no: 4, pp. 4533-4536, April 2021.