Agro-Inundation For Maximizing Crop Yield and Water Efficiency

Ashriti Jain, Akshay Asati, Aman Kumar, Sanhita Mahajan, Mrunalee Dhone
Computer Science and Engineering
G. H. Raisoni College of Engineering, Nagpur, India
jain_ashriti.cs@ghrce.raisoni.net, asati_akshay.cs@ghrce.raisoni.net, kumar_aman.cs@ghrce.raisoni.net, mahajan_sanhita.cs@ghrce.raisoni.net, mrunali.dhone@raisoni.net

*Abstract***— Population is increasing at an alarming rate which calls for the improvement of the food resource yield obtained via farming. Irrigation requires tremendous amounts of water, so it is imperative to use water wisely. For this purpose, there is a need for smart irrigation techniques which are designed and proposed in such a manner that the techniques depict the water level and its requirements in a precise and recommended way. As of now, farmers couldn't find a reliable way to estimate the water needs for their crops. The model proposed uses machine learning techniques to forecast crop water requirements based on the weather attributes for a particular day. In order to make efficient use of this dataset during the prediction and extraction of similar patterns, the features that determine crop water requirements were examined and pre-processed based on our understanding of irrigation and machine learning. Correspondingly, a decision tree technique was applied to predict water requirements to evaluate our data preprocessing technique. We have built a training dataset using the weather attributes of the year 2015. The potential daily water consumption in this training dataset was determined using the Evapotranspiration method. Therefore, this machine learning-based solution is essential for farmers to utilize the available water resources to the fullest.**

Keywords—Machine learning, data preprocessing, decision tree, irrigation, evaporation, transpiration, evapotranspiration.

I. INTRODUCTION

Agriculture depends heavily on the supply of water. Functions of various hydrological models and approaches related to data mining are becoming essential for effective water management in irrigation. Crop production does not always make efficient use of the majority of the water given for irrigation. Typically, studies suggest that approximately 45% of water is being only utilized by field crops and the rest 15% is wasted by agriculture supply channels, 15% is wasted during transportation, and 25% is wasted due to inefficient water governance [1]. Farm Work itself is responsible for the majority of water wastage due to incompetent water management practices. So, a smart irrigation technique can be useful to manage water efficiently [2]. Thus Agro-inundation aims to predict crop water requirements based on the present weather attributes for a particular day, using a particular formulation. The weather attributes of the year 2015 have been taken while designing the model.

II. ARCHITECTURE

Every computing system or program always has an architecture of the system. It defines the life cycle of the system. Basically, architecture is the design that demonstrates the best exemplary view of the system. So to visualize the system the workflow is split into 5 phases.

A. Overview of system architecture

The entire model of agro-inundation is divided into five segments mainly collecting data from various reliable sources and then preparing a high-quality dataset file by preprocessing the collected data. Pre-processing of data consists of two sub-parts, attribute selection and construction of the data set. This process provides us with our dataset containing input parameters and output which is the crop water requirement. The third step includes selecting the appropriate formulation for our model from carrying out various surveys which will be discussed in the subsequent section. This formulation is automated through R programming code. After that, an IBM Watson Machine Learning API service is deployed for checking the accuracy of our model against the formulation-based output. This Watson service uses a C5.0 algorithm, a type of decision tree to evaluate the model. After getting the desired accuracy of the model, a web-based application is developed for taking inputs from the user's end and returning the result, which is the crop's water requirement (in litres) for the given input weather attributes.

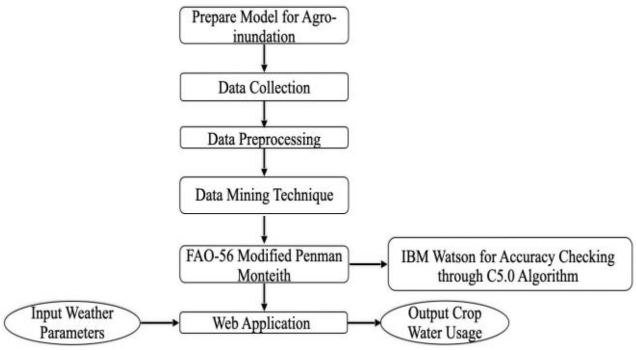

Fig. 1. System Architecture

B. Parameters for Input and Output

A user-interactive web application serves as the system's input. These inputs are fed to the formulation system developed and return the desired output. The input and output parameters are described below:

Input to the system:

1. Maximum and minimum Temperature (in degrees).
2. Rainfall (in mm)
3. Solar Radiation (in MJ/day)
4. Humidity (in %)
5. Wind speed (km/day)

Output to the system:

1. Daily crop water Requirement (in litres)

III. LITERATURE SURVEY

To estimate the crop's water requirement, different parameters such as the crop water level, the atmosphere of the targeted area, and different ways to calculate the water requirement must be discovered.

A. Iirrigation Requirement in India

In 2020, around 35% of India's agricultural area had effective irrigation. Monsoons are a necessary and dependable source of water for about two-thirds of the cultivated land in India [1]. Irrigation systems serve to increase agricultural production, decrease reliance on monsoons, offer rural employment possibilities, stabilize the crop cycle and provide food security [3]. The steady supply of water for irrigation provides a sense of stability to the farmer and stimulates him to utilize contemporary farming techniques and patterns to boost production.

B. Crop Water Requirement

The water need of a crop is the measure of water, regardless of source, needed for typical crop production and development over a certain time period at a place. It can be provided through irrigation, precipitation, or a combination of both. Plant metabolism, evaporation, and transpiration - collectively known as consumptive use (CU) - all require water in order to function. As a plant's metabolic processes use a minimal quantity of water (below 1% of the total water flowing through a plant), evaporation and transpiration (ET) correlate directly to consumptive consumption (CU) [4]. The phrase "water requirement" (WR) also refers to ET and includes the quality of water required for special procedures including water required for recycling, transplanting, and preparing the land, as well as any losses experienced.

During irrigation application in the fields. Application losses and the water required for special activities are added to CU to calculate WR. Therefore, Water Requirement (WR) = Irrigation (IR) + Effective Rainfall (ER) + Soil type(S)

The provision of irrigation, soil profile, and rainfall, together with shallow water tables, would determine the water requirement and the corresponding supply demand.

C. Study Region

Nagpur district in Maharashtra is the focus of the project. The Indian Meteorological Department provided us with reliable weather data, which drove our decision to select Nagpur. The Nagpur area is known for its orange production, so this was a natural project subject. As a result of the National Research Center for Citrus' presence in Nagpur, extensive knowledge about Orange Crops was gained there.

D. Various techniques for calculating agricultural water needs

There are a number of methods for predicting crop water needs. Of all possible methods, around seven appeared to be useful after a survey of all possible methods. Comparative studies were conducted on all these methods. A survey found that the FAO 24 Penman-Monteith method proved to be more accurate in humid climates, while the FAO 56 Hargreaves method proved to be more accurate in summer climates. Aside from that, the Turc method and Priestley Taylor both performed well in Kharif and Rabi. Method Thornthwaite showed the most Accurate results in the dry region. And only if air-temperature datasets are available for a site, the Blaney-Criddle results are superior. The studies concluded that all the above methods depended on climatic conditions and operated differently, but FAO 56 Modified Penman-Monteith is clearly the most accurate and efficient [5].

TABLE I. DIFFERENT METHODS AND THEIR SUITABLE CONDITIONS

S.no	Method	Condition in which performing well
1	FAO-56 Modified Penman Monteith	All types of season and different climates
2	FAO-24 Penman Monteith	Humid Climate
3	FAO-56 Hargreaves	Summer season
4	Turc	Kharif season
5	Thornthwaite	Dry Region
6	Blaney-Criddle	Only air-temperature datasets are available for a site
7	Priestley Taylor	Rabi season

Using the table, we conclude that the FAO 56 Modified PM method is the best formulation for our Agro-inundation model, as it provides accurate results in various weather conditions as well as in different climate conditions. Although there were methods which required minimal amounts of data in their formulation, they couldn't beat the challenge of providing accurate results in diverse climatic conditions which made the FAO 56 Modified Penman-Monteith method to be a standard approach for calculating Evapotranspiration which varies depending on the region [5].

IV. DATA COLLECTION AND DATA PREPROCESSING

Different data types are available from different sources and every source has its own procedure. Subsequently, the data should be preprocessed to remove the noises and the null values.

A. Data Source

Once parameters are identified, we need to find the source from where we can get the historical data of these parameters. The following are the sources of our data:

1. Indian Meteorological Department, Nagpur
2. National Solar Radiation Database, USA
3. National Research Centre for Citrus, Nagpur

B. Attribute Selection

The data for temperature, rainfall, average humidity, and wind speed were collected from the Indian Meteorological Department, Nagpur. The data regarding solar radiation was collected from NSRDB. The monthly crop water usage pattern for different ages of orange plants was obtained from the National Research Centre for Citrus, Nagpur. The data received from the above-mentioned sources is in the form of an Excel spreadsheet.

Once the attribute selection is done the next step is to remove noise from the data set. Removal of NA values can be done in 3 ways.

1. Remove all rows from the dataset that have null values
2. Replace the value of the attribute that has a null value with zero.
3. Replace null values with the mean of that attribute.

In our case, we have replaced all null values with the mean of that attribute. This is because we want the data for all days, so we must maintain accuracy in results [6].

C. *Data Set Construction*

Before using the final dataset to train the IBM model, it is preprocessed with R language. The daily reference evapotranspiration was applied to compute the crop's daily water use. Evaporation and transpiration are the two terms that are simultaneously used to refer to the loss of water occurring from the land in the presence of sunlight. These terms give us ET0, which frequently describes both of these processes, defined as the speed at which water instantly evaporates from soil from a specific vegetation surface area [7].

ET0 computed as –

$$ETo = \frac{0.408\,\Delta(Rn - G) + \gamma\,\frac{900}{T+273}\,u2(es - ea)}{\Delta + \gamma(1 + 0.34\,u2)} \tag{1}$$

TABLE II. Input data set

YEAR	MONTH	DATE	TEMP_MAX	TEMP_MIN	R/F	SOLAR_RADIATION	AVG_HUMIDITY	WIND_SPEED	ATMP_PRESSSURE
2015	1	1	21	17	4.06	17.712	74	10	101.2
2015	1	2	19	17	0.76	17.712	90	5	101.5
2015	1	3	21	18	0	17.712	88	2	101.6
2015	1	4	22	17	0	17.712	77	3	101.7
2015	1	5	18	11	0	17.712	58	5	101.7
2015	1	6	17	6	0	17.712	53	2	101.6
2015	1	7	18	7	0	17.712	57	2	101.5
2015	1	8	18	8	0	17.712	52	3	101.6
2015	1	9	17	6	0	17.712	54	2	101.8
2015	1	10	17	5	0	17.712	51	2	101.8
2015	1	11	18	8	0	17.712	51	3	101.8
2015	1	12	18	7	0	17.712	54	5	101.7
2015	1	13	18	7	0	17.712	49	3	101.6
2015	1	14	18	6	0	17.712	59	2	101.5
2015	1	15	18	7	0	17.712	54	3	101.6
2015	1	16	18	8	0	17.712	52	3	101.7
2015	1	17	18	9	0	17.712	51	3	101.7
2015	1	18	18	9	0	17.712	52	3	101.7
2015	1	19	18	8	0	17.712	49	3	101.7
2015	1	20	18	8	0	17.712	51	3	101.7

Daily crop water usage is calculated as -

$$Xi = \frac{ET_0^i}{\sum_{i=1}^{n} ET_o^j} \tag{2}$$

$Wi = Xi * W_T$ using (2)

TABLE III. Final Dataset

	YEAR	MONTH	DATE	TEMP_MAX	TEMP_MIN	R.F	SOLAR_RADIATION	AVG_HUMIDITY	WIND_SPEED	water_use
1	2015	1	1	21	17	4.06	17.712	74	10	7519.376167727
2	2015	1	2	19	17	0.76	17.712	90	5	4405.08908429512
3	2015	1	3	21	18	0	17.712	88	6	4950.29267991176
4	2015	1	4	22	17	0	17.712	77	3	5938.74844594978
5	2015	1	5	18	11	0	17.712	58	5	7333.36775106894
6	2015	1	6	17	6	0	17.712	57	2	5265.54604380109
7	2015	1	7	30	7	0	17.712	57	2	7542.08941483588
8	2015	1	8	18	8	0	17.712	52	3	6538.72799959081
9	2015	1	9	17	6	0	17.712	54	2	5441.39681449848
10	2015	1	10	17	5	0	20	51	2	5790.89142571757
11	2015	1	11	18	8	0	17.712	51	3	6633.72967946353
12	2015	1	12	18	7	0	17.712	54	5	7379.61956714573
13	2015	1	13	18	7	0	17.712	49	3	6710.51729286771
14	2015	1	14	18	6	0	17.712	59	2	5379.07466810493
15	2015	1	15	18	7	0	17.712	54	3	6377.69284622215
16	2015	1	16	18	8	0	17.712	52	3	6609.21874063803
17	2015	1	17	18	19	0	17.712	51	3	7999.28880419953
18	2015	1	18	18	9	0	17.712	52	3	6714.58463259701
19	2015	1	19	18	8	0	17.712	57	3	6259.86709326009
20	2015	1	20	18	8	0	17.712	51	3	6721.59417657629

V. ALGORITHM

Data mining and machine learning algorithms are useful tools for discovering information and patterns from a huge amount of data. Moreover, they provide us with various approaches & algorithms that help increase output accuracy and find mining rules. Since classification offers a clear hierarchy that is simple to understand and employs to make judgments, the decision tree-based approach is used for classification [8].

A. Decision tree

Decision trees have been successful when it comes to working with a supervised learning dataset. It divides the dataset into groups based on similarity. It takes a categorical dataset as input and designs a tree where leaf nodes are known as classes and non-leaf nodes are known as tests based on which a target field is obtained. The paths are built from the root node until all leaf nodes are verified, and each node provides a choice for the field. We will be using the C5.0 algorithm which is a type of decision tree using Watson ML API to build our model.

B. C5 Algorithm

C5 Algorithm steps -

1. Collecting training data: The raw data collected is the first pre-process. Data preprocessing is a process in which pointless features are removed from data and missing values are handled by replacing them with the mean value in the distribution of that attribute.
2. Attribute Selection: The information gain & gain ratio is used to select the most relevant attributes so as to transform data to make it suitable for modelling. The attributes are selected such that they have one categorical value as an output field and at least one input field.
3. Tree Building: The C5.0 decision tree algorithm is used to generate decision trees. The algorithm will start with a root node. The attribute with the highest gain ratio is selected as the root node. It will recursively split data into smaller subsets based on the attributes that provide the most information gain.
4. Pruning: To prevent overfitting, the algorithm prunes the tree after building the decision tree and testing the last level split. Pruning involves removing branches of

the tree that do not contribute much to the accuracy of the model.

5. Model Evaluation: This is the final step in which we pass updated data through the decision tree to make predictions. The tree will output a predicted value based on the input fields.

We have used the C5.0 algorithm using Watson ML API due to its strengths including saving us from performing long model training. It is also efficient when faced with missing data values and large amounts of input data.

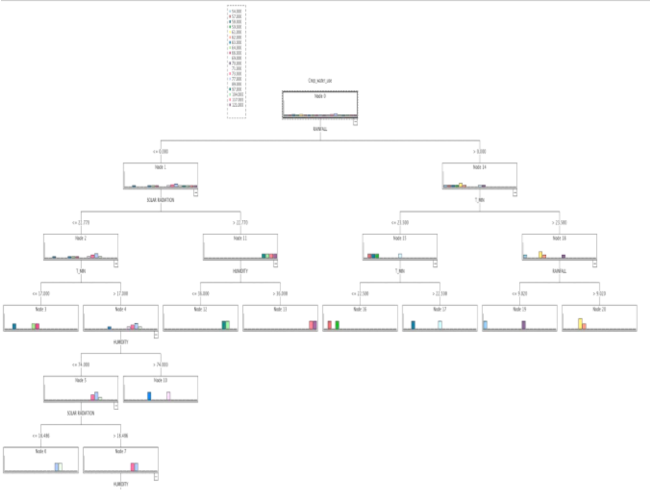

Fig. 2. Decision Tree

VI. Development and Deployment Model

The development of this project is divided into two parts.Firstly, the production of the website and secondly the accuracy check of the calculated value from the website.

A. Structure

1. Frontend: HTML, CSS and JavaScript form the front end of the application. JavaScript was chosen because of its simple integration and use for our purposes.

2. Backend: NodeJS, is used as a common backend JavaScript framework. Communication between frontend and backend takes place via API calls.

3. Service: IBM Watson service was used with our model to check the accuracy.

B. Platform

As mentioned earlier, the preprocessed data is used to create a decision tree using the C5.0 algorithm. IBM Watson Machine Learning API is utilized in the creation of the decision tree. To produce decision trees, software called SPSS Modeler is employed.

C. IBM Watson Machine Learning API

IBM Watson is IBM's flagship machine learning service, which provides sophisticated functionalities for various machine learning problem statements, including analytics software SPSS. A predictive analytics tool called IBM SPSS Modeler enables users to quickly create precise prediction models and communicate that insight to groups, people and systems. It offers a variety of cutting-edge algorithms and analyses to produce insights in a timely manner. The modeler has several models, but we use the C5.0 model [9].

D. Steps for using IBM Watson Machine Learning API

1. Download SPSS Modeler.

2. Select a data source node from the list, which is excel in our case, and upload the pre-processed data excel file.

3. Select a type specifier node from the list to specify the types of different attributes as continuous or discrete and to read the value from the source node.

4. Select the C5.0 model from the various options in the SPSS modeler. Select the daily crop water usage column as the target and input features to include weather variables such as rainfall, wind speed, humidity, minimum and maximum temperature, and solar radiation.

5. Run the model just created. A golden nugget will be created which is the trained pre-processed data, a decision tree.

6. Select an input file type from the source list. Upload a CSV file of our pre-processed data. Connect it to the type specifier as in step 3 and connect it to the golden nugget.

7. Connect the golden nugget to a table from the output list.

8. Select the branch created in steps 6 and 7, right click and select use as a scoring branch.

9. Save the model.

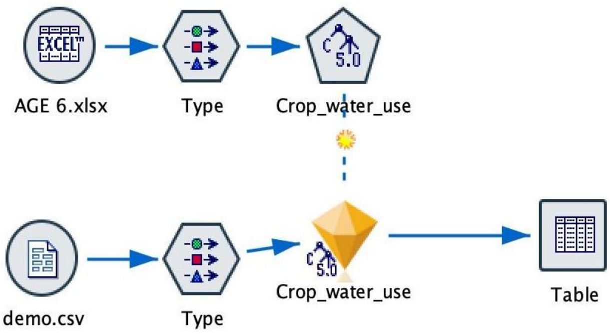

Fig. 3. Decision tree model built using SPSS Modeler

VII. Results

The training dataset was constructed to test our model. Based on the Reference Evapotranspiration Coefficient, the potential water consumption for each day independently was determined through this training dataset. After this, we got a pleasing result with an accuracy of 79.145% against the manually calculated result.

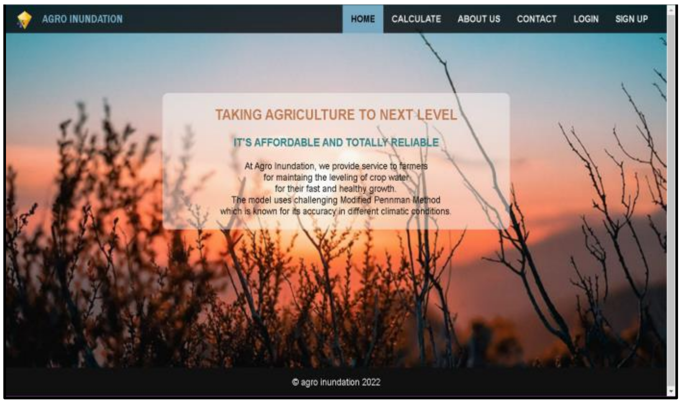

Fig. 4. Home page of our current application

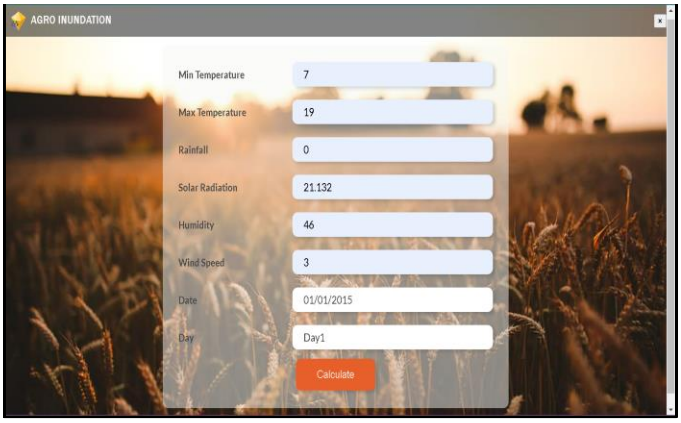

Fig. 5. Current online application screenshot

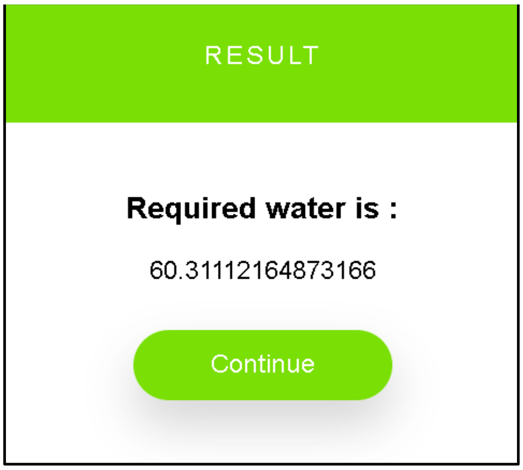

Fig. 6. Current online application showing results.

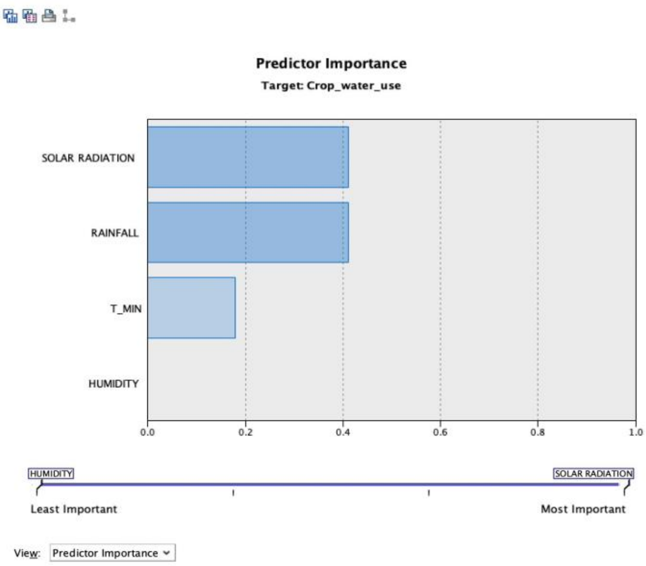

Fig. 7. Dependence of daily crop water usage on input parameters

VIII. CONCLUSION

Farmers are using the traditional way of irrigation system in which they themselves need to provide water to their crops which results in over-watering or under-watering of crops. The lack of scientific tools for predicting crop water requirements has been a significant challenge for farmers. However, the development of the agro-inundation model has provided an effective solution to this problem. By using this model, farmers can accurately provide the optimal amount of water to their crops, resulting in better crop yield and reduced water waste. Moreover, the use of agro-inundation models can save farmers from investing in expensive IOT-based irrigation systems. Therefore, the agro-inundation model holds great promise for the future of agriculture.

REFERENCES

[1] Pulido Calvo, I., Roldan, J., Lopez Luque, R. and Gutierrez Estrada, J.C. (2003): Demand Forecasting for Irrigation Water Distribution Systems. Journal of Irrigation and Drainage Engineering 129(6):422-431.

[2] Smart Farming: A Step towards Techno-Savvy Agriculture, N. Shashwathi, Priyam Borkotoky, Suhas K

[3] Robert C. J. Koo, Water Requirements of Citrus And Response To Supplemental Irrigation

[4] Richard G. Allen, Luis S. Pereira, Dirk Raes, Martin Smith: Crop evapotranspiration Guidelines for computing crop water requirements

[5] EzAnkit K. Kulkarni, Ravichandra Masuti, V. S. Limaye, Comparative study of evaluation of evapotranspiration methods and calculation of crop water requirements at chaskaman command area in Pune region, India

[6] Mahmood A.khan, Md. Zahidul Islam, Mohsim hafeez (2011): Irrigation water Demand Forecasting - A data pre-processing and data mining approach based on spatiotemporal data.

[7] Lincoln Zotarelli, Michael D. Dukes, Consuelo C. Romero, Kati W. Migliaccio, and Kelly T. Morgan, Step by Step Calculation of the Penman- Monteith Evapotranspiration (FAO-56 Method)

[8] A comparative study of decision tree ID3 and C4.5 : Badr HSSINA, Abdelkarim MERBOUHA, Hanane EZZIKOURI, Mohammad ERRIT

[9] Juraj Collinaszy, Marek Bundzel, Iveta Zolotova : Implementation of Intelligent Software Using IBM Watson and Bluemix

[10] National Renewable Energy Laboratory," National Solar Radiation Database," [Online]. Available: https://nsrdb.nrel.gov/.[Accessed:Sep 12, 2022]

[11] India Meteorological Department, "IMD Nagpur," [Online]. Available: http://www.imdnagpur.gov.in. [Accessed: Sep. 12, 2022]

[12] National Research Centre for Citrus, Nagpur, "Home - NRCC," [Online]. Available: https://nrccitrus.nic.in/. [Accessed: Sep. 12, 2022].

[13] V. Kanhekar, T. Deshbhratar, Y. Matey, K. Kalbande and A. Deshmukh, "Hydroponic Farming using IoT," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 583-586, doi: 10.1109/ICECAA55415.2022.9936366.

[14] K. Kalbande, S. Choudhary, A. Singru, I. Mukherjee and P. Bakshi, "Multi-Way Controlled Feedback Oriented Smart System for Agricultural Application using Internet of Things," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2021, pp. 96-101, doi: 10.1109/ICOEI51242.2021.9452946.

[15] P. Kolhe, K. Kalbande and A. Deshmukh, "Internet of Thing and Machine Learning Approach for Agricultural Application: A Review," 2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22), Nagpur, India, 2022, pp. 1-6, doi: 10.1109/ICETET-SIP-2254415.2022.9791751.

[16] P. Kolhe, A. Baseshankar, M. Murekar, S. Sadhankar, K. Kalbande and A. Deshmukh, "Smart Communication System for Agriculture," 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, India, 2022, pp. 1122-1126, doi: 10.1109/ICICICT54557.2022.9917715.