IUBAT Review

A Multidisciplinary Academic Journal

Md. Rashedul Islam*, Md. Sakibul Islam, Md Sabbir Talukdar & Utpal Kanti Das 2021. **"A temperature-based machine learning model to predict water needs and to minimize water waste during irrigation"** *IUBAT Review* 4 (1): 19-31. iubat.net/journal/

A temperature-based machine learning model to predict water needs and to minimize water waste during irrigation

Md. Rashedul Islam[1]*, Md. Sakibul Islam[1], Md Sabbir Talukdar[1] and Utpal Kanti Das[1]

[1]Department of Computer Science and Engineering, International University of Business Agriculture and Technology (IUBAT), Dhaka 1230, Bangladesh

*Corresponding author: E-mail: rashed@iubat.edu

ABSTRACT: *While the use of machine learning algorithms in the irrigation sector is not a new approach, the method of applying machine learning algorithms in the irrigation sector in terms of minimizing the wastage of water is new in Bangladesh. Irrigation of any crops is heavily dependent on the groundwater source in our country. To meet the goal of precious agriculture, it is a matter of necessity that we effectively find a novel approach to address the issue of water wasting during irrigation and ensure the proper use of the water sources. The origin of our research is based on this issue where we have applied decision tree and random forest regressor to build a model that is assisting us to irrigate a precise amount of water during harvesting season. Reference Crop Evapotranspiration (ETo) and Crop Coefficient (Kc) are the parameters to calculate the water need for a particular crop. The Hargreaves method is used to find out the Reference Crop Evapotranspiration (ETo) by importing a python-based library named PyEto. To maximize the use of our dataset, we logically fragmented our dataset into two sets providing 180 instances each. A respective logical fragment contains the Crop Coefficient (Kc) value of Tomato crops of its four different growing stages. The multiplication result between Reference Crop Evapotranspiration (ETo) and Crop Coefficient (Kc) compute the target variable that is, Crop Water Needs (ETCrop). The data split in 90:10 ratio to train and test the model respectively. Our model provides significantly higher accuracy after comparing the training and testing of the data. The high accuracy of our model will lead to irrigating the exact amount of water required to harvest tomato crops and save the unnecessary wastage of the water.*

KEYWORDS: *Precious Agriculture, Decision Tree, Random Forest, Irrigation, Crop Water Needs, Weather*

1. Introduction

A significant part is played by the weather and climate in the agriculture productivity throughout the world. Agriculture is, without any doubt, a vulnerable major sector that is affected by climate change such as changes in the rainfall patterns, temperatures, the rise of floods and sudden droughts. The crop productivity of Bangladesh is heavily dependent on the rain, surface or groundwater. Rain-fed farming is the natural way to get the rain and use it directly on the crop fields. But recently we are experiencing the delay of the monsoon season in our country. And if this continues, then the production of rice and wheat may fall by eight per cent and 32 per cent respectively by 2050 (Rezvi, 2018). Due to the high density of the population, as has been observed, there is a declining trend in the groundwater level in many areas in Bangladesh. This is also affecting other surrounding areas. The overall reason for the groundwater level declining trend is overexploitation and very negligible recharge due to changes in rainfall patterns. As a result, the ground level water is permanently declining in many areas of Bangladesh, and the recharging of the groundwater level is nearly zero percent (WASA, 2020). The opposite scenario is regularly observed in our country. From the survey of the Needs Assessment Working Group (NAWG), in 2020, 21 districts of Bangladesh were affected by the flood, and the impact was moderate to severe in 16 districts. As a result, the destruction of crops and inundation of farmland due to sudden overflow decreased agricultural production every year.

Due to the changing nature of the climate, it's definitely not a fair idea to depend only on the ground or surface water or even on the rainfall. Farmers need to depend on the irrigation system to cultivate their lands that mostly depend on the groundwater. However, the farmers waste a substantial amount of water during irrigation in Bangladesh (Mainuddin et al., 2020). Around 800 liters of irrigation-water is misused by the farmers across the country to produce a kilogram of the paddy (Hoque, 2018). As a matter of fact, our current crop cultivation approaches and techniques are failing to meet the potential advantages of precision agriculture (Pierce & Nowak, 1999). It is a noticeable situation that we need to find, adopt and apply new approaches and methods to cultivate our crops without wasting any cultivation resources such as water, pesticides or even the sunlight.

The adaptation and application of machine learning algorithms in our irrigation system may lead us to the zero per cent wastage of water during cultivation.

2. Literature Review

Machine learning algorithms can be used in various sectors of agriculture such as plant disease detection, yield prediction, managing crop quality, checking the weather condition, crop growth monitoring and so on (Devi et al., 2020). The random forest regression (RFR) approach is used to monitor the health of the crops with the help of satellite data (Maimaitijiang et al, 2020). Detection of Maize Leaf Disease Detection and Classification was implemented by applying different supervised machine learning algorithms such as Naive Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM) (Panigrahi et al., 2020). The Support Vector Regression (SVR) and Linear Regression (LR) were used to predict the crop yield (Haque, Abdelgawad et al., 2020). There are several well-known methods available such as the FAO-56 Penman-Monteith method, The Hargreaves equation and the Thornthwaite equation to calculate the evapotranspiration (ETo), from where we can calculate the amount of water required for a specific crop. The FAO-56 Penman-Monteith method requires site location, air temperature, humidity, radiation and wind speed data for daily, weekly, ten-day or monthly ETo calculations (Zotarelli et al., 2010). The Hargreaves equation requires minimum, maximum and mean temperature and extraterrestrial radiation to calculate the ETo (Samani & Pessarakli, 1986). The Thornthwaite equation (Thornthwaite, 1948) is one of the oldest equations to calculate ETo from the mean monthly air temperature and mean daily daylight hours for each month.

As our research approach is to use machine learning algorithms to predict the exact amount of water required at different growth levels of a crop with a possible minimum weather learning, so we are going to apply The Hargreaves equation to calculate the Reference Evapotranspiration (ETo) with different machine learning algorithms to get a precious output for the Crop Water Need (ETcrop).

Various machine learning-based approaches were applied to accomplish the task of predicting or forecasting the crop water need. In a few cases, satellite-based information such as satellite images and weather-related information were analyzed as a parameter to increase the accuracy of the water need prediction. By collecting field and weather data, classifying soil moisture and analyzing the correlation between variables to develop an artificial intelligence model to predict crop water need in (Pasha et al., 2020). In (Khan et al., 2011, pre-processed datasets along with five different data mining methods and traditional approaches are used to decide the water need prediction. According to their applied techniques, the SysFor data mining approach performs well and gives 97.5% of accuracy. The Internet of Things (IoT) based applications were used to predict the required water supply for the irrigation process (Goap et al., 2018). Sensor-based data like soil moisture, soil temperature, environmental conditions, air temperature, Ultraviolet (UV) light radiation and relative humidity from the crop field were analyzed to improve the prediction of crop water need. The hybrid particle swarm optimization (PSO-ELM) with the artificial neural network (ANN) and Random Forest were combinedly used to predict the reference evapotranspiration (ETo) value (Zhu et al., 2020). The ANN, Random Forest and PSO-ELM predicted well with a minimum number of climate data such as radiation and temperature. Collecting soil data with the help of Wireless Sensor Network (WSN) and sending it to a common Data Server to predict the water demand supported by the machine learning (ML) algorithms were developed for minimizing the irrigation water loss (Kwok, & Sun, 2018). Along with the WSN and ML algorithms, the backbone of the system was the Internet of Things (IoT) for providing a comparatively cheaper and precise solution. To address the case of real-time optimal water pumping scheduling, the study focused on developing the ANN model to prepare the water demand schedule twenty-four-hour and one week ahead (Bata et al., 2020). The developed two ANN time-series-based models, a nonlinear autoregressive with exogenous (NARX) model and a nonlinear autoregressive (NAR) model helped to drop the error level by 30%. The continuous monitoring of the weather conditions like temperature, air pressure, sunlight and rainfall and then relaying the collected data through a sensor network were analyzed through Multiple Linear Regression algorithms and developed a model to forecast the water needed in the irrigation field (Vij et al., 2020).

The traditional machine learning-based approaches such as Linear Regression, Support Vector Machine, Decision Tree, Random Forest, Extra Tree Regressor, Adaboost and Gradient Boosting were applied to predict the crop water needs (Sidhu et al., 2020). The Random Forest and Decision tree performed well on their adopted dataset. They divided their dataset variables into four different sets by analyzing the correlation coefficients among the variables. And perform the ML algorithms on each set by following a similar trend of years. In total, fourteen variables were used to run the ML models such as maximum temperature, minimum temperature, gross minimum temperature, dry bulbs at 6 am, 2 pm, the wet-bulb at 6 am, 2 pm, humidity, relative humidity, soil temperature at different depth of the surface, sunshine hours, wind speed, wind direction and rainfall. But our research target was to get better results than those found by Sidhu et al. (2020) in terms of model accuracy of predicting the crop water needs by selecting the minimum number of weather-based features.

The accuracy obtained by (Sidhu et al., 2020) in Se1, Set2, Set3 and Set4 for Decision tree were 70.5%, 67%, 65% and 65% respectively. The accuracy of each set for Random Forest were 64%, 66.5%, 69% and 68%. Securing higher accuracy in each set of data applying decision tree and random forest is one of our research targets. To address our research goal, we have used only four features which are minimum, maximum and mean temperature. Other researchers have not taken extraterrestrial radiation as a parameter to determine the crop water needs. However, the extraterrestrial radiation analyzed through the Random Forest and Decision Tree algorithms to predict the crop water needs in our research.

3. Crop Water Need (ETCrop) Calculation:

We adopted the methods and procedures that are approved by the Food and Agriculture Organization (FAO) (Allen et al., 1998). FAO is a specialized agency of the United Nations (UN). The goal of FAO is to defeat hunger and achieve food security for all. FAO works in over 130 countries worldwide. According to the study, the major climate factors which influence the Crop Water Needs (ETCrop) are the sunshine period, temperature, humidity and wind speed. The requirement of ETCrop increases with the temperature. If the temperature becomes scorching, the demand of ETCrop goes up. Simultaneously, the demand of ETCrop drives down while the scale of temperature goes downwards. The same sort of relationship is valid for humidity, wind speed and sunshine. In a nutshell, we can come out with a statement that the areas which are hot in temperature, humidity is relatively dry, windy and sunny, highest ETCrop is required for those particular areas.

The influence from the climate and surface on crop water needs is the Reference Crop Evapotranspiration (ETo). The unit of ETo is represented in millimeters per unit of time, for example, mm/day, mm/month or mm/season. ETo is the estimation of the evapotranspiration from a “reference surface” (Brouwer, & Heibloem, 1986). In our research, the “reference surface” area in Basel is a city in northwestern Switzerland on the bank of the river Rhine. Our entire dataset is populated from Basel, and the training & testing of our models were based on this dataset.

Another vital factor that affects the demand for water needs is the Crop Coefficient (Kc). Crop Coefficient is the value of a particular crop that can be determined during the growing period of a crop, specifically changes in vegetation and the ground cover means. The value of Kc differs in the main three different growth stages of a crop such as the initial stage (Kcini), the mid-season stage (Kcmid) and the late season or end-stage (Kcend) (Allen et al.,1998). For our research purpose, we have used the Kc value of different stages of the Tomato plant. Because the tomato is a perennial plant and the maximum growing period of the tomato is 180 days, that is why we can easily divide our dataset into two groups assigning 180 days each per year for ensuring maximum use of our dataset. The Kc values of the Tomato plant are shown in the following table. (Allen et al., 1998),

Table 1: Values of the Kc for Various Growth Stage of Tomato Plant (Brouwer & Heibloem, 1986)

Crop	Kc_{ini}	Kc_{dev}	Kc_{mid}	Kc_{end}
Tomato	0.45	0.75	1.15	0.80

The determination of Crop Water Needs (ETCrop) can be calculated effectively for a particular crop from the Reference Crop Evapotranspiration (ETo) and Crop Coefficient (Kc) using the following formula (Allen, Pereira, Raes, & Smith, 1998),

$$ET\ crop = ETo \times Kc \qquad \text{Eq. 1}$$

Till now we have known the values of Kc for a particular crop, and in our case, it is the tomato plant. But we are yet to find the value of ETo. There are several methods that are well established to calculate the ETo. But our aim is to use minimum climate data for the prediction of the ETCrop; for that reason, we are applying Hargreaves Equation to estimate the ETo.

4. The Hargreaves Equation to determine ETo

George H. Hargreaves and Zohrab A. Samani combinedly came out with a famous equation for the measurement of ETo in 1985. That equation is called The Hargreaves equation, where ETo is calculated from the daily or mean values of maximum and minimum temperature with the value of extraterrestrial radiation of a particular region (Hargreaves, & Samani, 1985).
The other well-established methods such as the FAO Penman-Monteith method and the Thornthwaite method can also be used to determine the ETo. But why we have chosen the Hargreaves method will be highlighted in the following discussion.
The Hargreaves method is recommended by the FAO (Allen et al., 1998) and widely used to calculate the ETo. The equation is formulated as follows (Hargreaves, & Samani, 1985),

$$ETo = 0.023\ (0.408)\ (T_{mean} + 17.8)\ (T_{max} - T_{min})0.5\ R_a \qquad \text{Eq. 2}$$

Where, ETo is the Reference Crop Evapotranspiration, T_{max} = maximum air temperature (^oC), T_{min} = minimum air temperature (^oC), T_{mean} = mean air temperature (^oC) and R_a = extraterrestrial radiation (MJ.m-2). Additionally, 0.408 is a factor to convert MJ.m-2 to mm.
The extraterrestrial radiation (R_a), can be estimated based on the location's latitude and the calendar day of the year by follows (Fisher, & Pringle, 2013),

$$R_a = \frac{24(60)}{\pi} G_{sc} d_r\ [\omega_s \sin(\varphi) \sin(\delta) + \cos(\varphi)\ \cos(\delta) \sin(\omega_s)] \qquad \text{Eq. 3}$$

Where, Gsc is the solar constant (0.0823 MJ.m-2.min-1), φ is the latitude (radians) and the term 24(60) is a factor to convert min to day.

Remaining factor, the dr determined from (Fisher, & Pringle, 2013),

$$d_r = 1 + 0.33 \cos(\frac{2\pi}{365} J) \qquad \text{Eq. 4}$$

where, d_r = inverse relative distance from earth to sun and J=calendar day of the year.
The solar declination (δ) determines from (Fisher, & Pringle, 2013),

$$\delta = 0.409 \sin(\frac{2\pi}{365} J - 1.39) \qquad \text{Eq. 5}$$

Sunset hour angle (ω_s) can be formulated as (Fisher, & Pringle, 2013),

$$\omega_s = arccos\ (-\ tan\ (\varphi)\ tan\ (\delta)) \qquad \text{Eq. 6}$$

As we are using the Python programming language to train and test our model, so we are using a python-based library called PyETo. PyETo library provides different methods such as FAO-56 Penman-Monteith, Hargreaves and Thornthwaite, using those we can calculate ETo simply loading a dataset.
Among the mentioned methods, Thornthwaite is the oldest one that was proposed in 1948 (Palmer, & Havens, 1958). The Thornthwaite equation requires mean monthly air temperature and mean daily daylight. The FAO Penman-Monteith method requires site location, air temperature, humidity, radiation and wind speed data for daily, weekly, ten-day or monthly ETo calculations. If we compare the Thornthwaite method with the FAO Penman-Monteith method, it is observed

that FAO Penman-Monteith always performs better than the Thornthwaite method (Chen et al., 2005).
On the other hand, in the case of evaluation of ETo based on temperature, the Hargreaves method performs best and while due to requirements of detailed meteorological data, the application of FAO Penman-Monteith is often constrained (Lang et al., 2017). By studying the above works of literature and factors, we have chosen the Hargreaves method over the FAO Penman-Monteith and Thornthwaite method.

To import the PyETo library and to get ETo using the Hargreaves equation, we have to write the python commands as follows (Richards, 2015),
First, convert latitude to radians and the date to day of the year (Julian day):

> >>> lat = pyeto. deg2rad (latitude of Basel) # Convert latitude to radians

>>> day_of_year = datetime.date(dates from the dataset).timetuple().tm_yday

>> import datetime, pyeto

To estimate extraterrestrial radiation, we first need to calculate solar declination, sunset hour angle and inverse relative distance Earth-Sun:
>>> sol_dec = pyeto.sol_dec(day_of_year) # Solar declination

>>> sha = pyeto.sunset_hour_angle(lat, sol_dec)

>>> ird = pyeto.inv_rel_dist_earth_sun(day_of_year)

>>> et_rad = pyeto.et_rad(lat, sol_dec, sha, ird) # Extraterrestrial radiation

Finally, we can estimate ETo by using the hargreaves() method of the pyeto library ,

>>> Hargreaves (minimum temperature, maximum temperature, mean temperature, et_rad)
The obtained values are stored in dataset under the ETo variable and multiplied with the Crop Coefficient (Kc) of different stages of the Tomato plant to compute the Crop Water Need (ET_{Crop}).

5. Data Collection, Splitting and Processing

The historical weather data of the Basel, a city in Switzerland, is freely available at https://www.meteoblue.com/en/weather/week/basel_switzerland_2661604. The dataset is populated from 01 January 2008 to 12 September 2020 (around 12 years). The features of the dataset are minimum temperature (oC), maximum temperature (oC), mean temperature (oC), minimum relative humidity (%), maximum relative humidity (%), mean relative humidity (%), mean cloud cover (%), minimum – maximum – mean soil moisture (m3/m3).
The total maximum harvesting duration (in days) of tomato crops are 180 days (Brouwer, & Heibloem, 1986). In the following table duration (in days) of initial, development, mid-season and end stages are shown,

Table 2: Duration of different growth stages of tomato crops

Crops	Total	Initial Stage	Crop Development Stage	Mid-Season Stage	Late Season Stage
Tomato	180	35	45	70	30

By considering the duration of the different growing stages of tomatoes, we have logically divided the dataset into 4 groups. Each logical group of data associated with the values of Kc at different growing stages of the tomato crop. The Kc of the initial growing stage of tomato crops is 0.45. According to the dataset, from 01 January

2008 to 04 February 2008 is the first logical set of data to be considered as the initial stage. The next 45 days of the development stage are the data from 5 February 2008 to 20 March 2008. The mid-season and late-season sage require 70 and 30 days respectively. From 21 March 2008 to 29 May 2008 designated as the mid-season stage and from 30 May 2008 to 28 June 2008 days are designated for the late-season growing stage of the tomato crops. This is the SET1 of our dataset that consists of the first 180 days of a particular year those are the growing period of the tomato crops in first timestamp.

Our second set, the SET2, started from 29 June 2008 to 26 December 2008, which is exactly the second 180 days of the year 2008. Our SET 2 is also logically divided into four groups based on the four growing stages of tomatoes. By maintaining the same pattern, each year from 2009 to 2019 is separated into two equal sets consisting of 180 days each to train and test our model.

6. Model Selection

From the adopted models in Sidhu et al. (2020), the researchers came out with a conclusion that, on this kind of historical dataset where we need to determine the Crop Water Needs (ETCrop), the Decision Tree and Random Forest regressor perform well.

omitting other machine learning algorithms such as Support Vector Machine (SVM), Linear Regression, Extra Tree, Ada Boost and Gradient Boosting.

For classification problems, a decision tree is always among the good choices. Basically, the decision tree creates a node using each variable of the dataset and generates a classification output. From the dataset we can observe that based on a specific temperature, humidity, cloud covering and soil moisture, the ETCrop value varies. That is the reason we can apply the decision tree to determine and predict our output or target variable.

As we can see, the decision tree uses the tree representation in terms of input and output features (Bonaccorso, 2017). Where the input features construct the decision nodes and on the other hand, the output feature lies at the leaf node. If we consider a set of dataset X as follows,

$$X = \{x_1, x_2, \dots, x_n\} \text{ where } x_i \, \varepsilon \, \mathbb{R}^m \qquad \text{Eq. 7}$$

Here every vector is made up of m features. The structure of the tree may change based on the feature and threshold. Threshold basically determines the left and right branches of a tree. Formally we can write (Bonaccorso, 2017),

$$\sigma = \langle i \,|\, tk \rangle \qquad \text{Eq. 8}$$

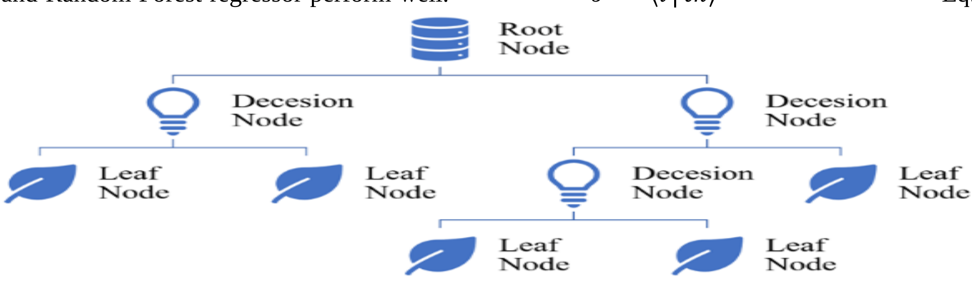

Fig 1. The inner structure of a decision tree

That is why, by considering their verdict we have also chosen the decision tree and random forest for predicting the ETCrop by

where, i is the index of the feature that we want to split and tk is the threshold. Random forests is a classifier consisting of a collection of tree-structured classifier {h (x, Θk), k=1,…}, where {Θk} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x (Breiman, 2001).

7. Feature Optimization and Performance Evaluation of the Model

In terms of evaluating of the model, we considered mean square error (MSE), mean absolute error (MAE), explained variance score (EVS) and coefficient of determination (R square score).

Calculating the average of square of errors is called MSE (Sidhu et al., 2020). MSE can be determined as following,

$$\text{MSE} = \frac{\sum_{i=1}^{n}(pi-ai)^2}{n} \qquad \text{Eq. 9}$$

Where a is the actual value of the target variable, p is the predicted value and n is the number of instances.

MAE can be formulated as follows (Sidhu et al., 2020),

$$\text{MAE} = \frac{\sum_{i=1}^{n}(pi-ai)}{n} \qquad \text{Eq. 10}$$

and the parameter of MAE is same as MSE. To get a better explanation of the variation of the dataset we are going to use EVS. EVS can be generated from the following equation (Sidhu et al., 2020),

$$\text{EVS} = 1 - \frac{Var(p-a)}{Var\ (a)} \qquad \text{Eq. 11}$$

Where a and p represent the actual value and the predicted value respectively.

R^2 explains the number of variances explained by the model. We can say a model is performing well if the value is nearer to 1 or equal to 1 (Sidhu, Kumar, & Rana, 2020). R^2 can be found by following formula,

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad \text{Eq. 12}$$

Where SST represents Sum of Square Total of the predicted value, SSR represents Sum of Square Regression of predicted value, and SSE represents the Sum of Square Error of the predicted value.

8. Results and Discussion

In our work we have made the dataset split into two parts. The initial dataset consisted of 12 years' worth of temperature, humidity, cloud cover, precipitation etc. values. In total 4700 instances were spilt into 90:10 ration. 90% data for model training and 10% data

	Dataset	Model	MSE	RMSE	MAE	EVS	R2	ACC
0	set1	DTree	0.003133	0.055970	0.013335	0.998865	0.998865	79.003361
1	set1	RForest	0.002038	0.045146	0.009428	0.999262	0.999262	83.003551
2	set2	DTree	0.001150	0.033914	0.007865	0.998641	0.998636	77.016034
3	set2	RForest	0.001125	0.033537	0.006901	0.998675	0.998666	81.016029

Fig. 2. Performance of the models

for

model testing stage. The model was tested with 470 values. We filtered some instances and reduced instances of our test data to provide better fitness values.

For prediction of the Crop Water Needs (ET_{Crop}), other supervised learning methods such as Support Vector Machine, Neural Network, Kernel ridge regression and other Linear models as well as Deep learning were considered. While the prior models were unable to give a satisfactory prediction, the Decision Tree and Random Forest regressor have given better prediction. The performance of Decision Tree regressor, Random Forest regressor models are evaluated in terms of various parameters and represented in figure 2.

As we observe in the figure 2, The MSE, RMSE, MAE and EVS for Set 1 of both DTR and RFR provide us the satisfactory outcome and on the other hand the R^2 value is impressively nearer to the 1 for both regressor. The overall performance for Set 2 of DTR and RFR underperformed than the Set 1. In terms of R^2 value, the Set 2 performed similarly as Set

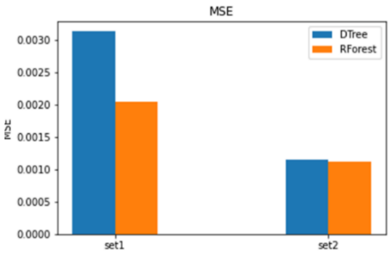

Fig. 3. The MSE Validation results for Set1 and Set 2

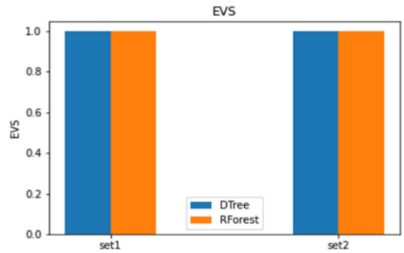

Fig. 4. The EVS Validation results for Set1 and Set 2

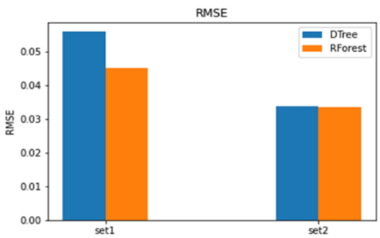

Fig. 5. The RMSE Validation results for Set1 and Set 2

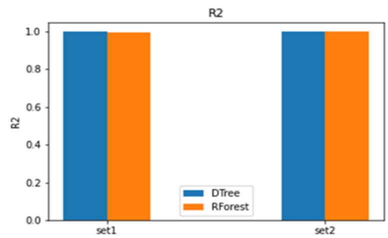

Fig. 6. The R^2 Validation results for Set1 and Set 2

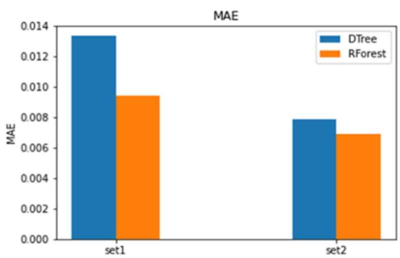

Fig. 7. The MAE Validation results for Set1 and Set 2

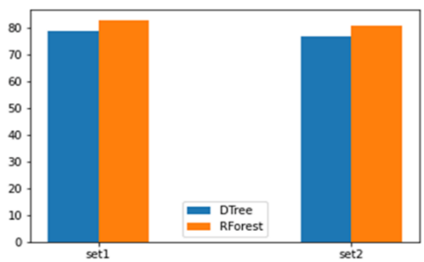

Fig. 8. The Model Accuracy results for Set1 and Set 2

In the above presented figures 3 to 8, we used blue color bar to represent the performance of decision tree regressor and the orange color bar representing random forest regressor. The outcome of the validation result for MSE, RMSE, MAE, EVS and R^2 is shown in figure 9 to 12. The overall accuracy of our model is shown in figure 8. The accuracy values for both Set1 and 2 for DTR and RFR shows better accuracy than (Sidhu et. al., 2020).

Performance of the both model is analyzed over two dataset Set1 and Set2. **Firstly,** for the Set1 dataset application of the DTR (Decision Tree Regressor) gives us the best performance with a slightly higher R^2 value and low MSE values. As the variance score is also better in DTR, the model satisfies multiple fitness criteria with higher values. **Secondly,** for the Set2 dataset application RFR (Random Forest Regressor) obtained highest value of R^2 value and explained variance score but least error score. A low MSE and MAE values along with a higher R^2 values means our models performance satisfies three distinct fitness benchmarks with higher values.

Thirdly, while predicting the Crop Water Needs (ET_{Crop}), our proposed methodology provides an accurate analysis under a lower computation time. More depth of the DTR and RFR will provide better fitness score as well.

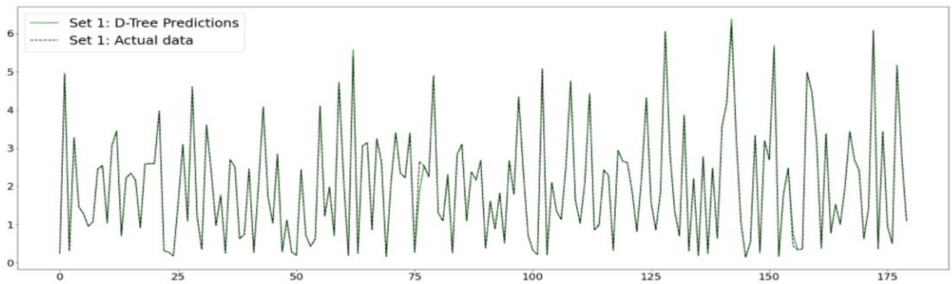

Fig. 9. Comparison between Training and Testing data for Decision Tree of Set 1

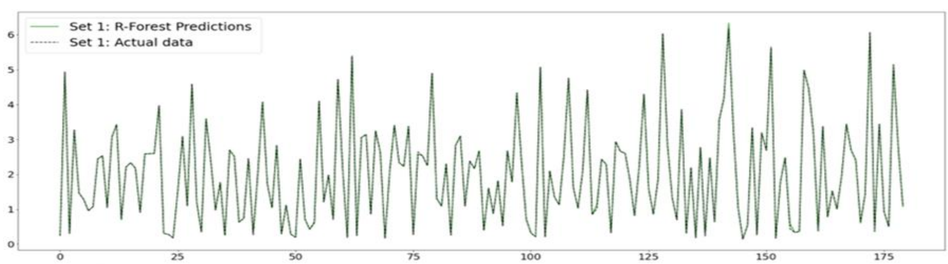

Fig 10. Comparison between Training and Testing data for Random Forest of Set 1

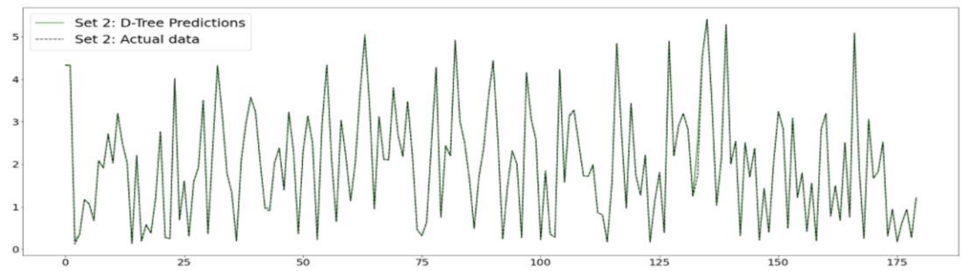

Fig 11. Comparison between Training and Testing data for Decision Tree of Set 2.

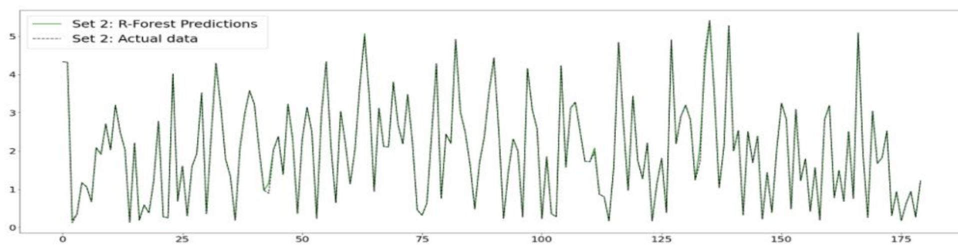

Fig. 12. Comparison between Training and Testing data for Random Forest of Set 2

As seen in Figure 9, the accuracy of our decision tree is quite similar to the actual data of Set 1. The model slightly underperforms around 70-80th and after 150th instance. Excluding those instances, our model secured 79.003361% accuracy that is relatively higher than our target accuracy. After performing Random Forest regressor on the Set 1, the model accuracy leads to 83.003351%, that is obviously crossing our target accuracy in figure 10. There is a slight deviation can be observed at 60th, 145th and 152nd instances.

On our Set 2, the accuracy of Decision Tree regressor is lower than the Random Forest regressor. But in both of the cases the obtained accuracies cross our target threshold. A slight fault of prediction is observed around 130th instance in figure 11.

Same sort of deviation in terms of prediction is also visible in figure 7.10 at 4th, 44th, 130th instances of figure 12. In despite of all these underperformances, in both cases the availed accuracies are 77.016034 and 81.016029, those are higher than our target accuracy.

9. Conclusion

For the sustainable development of a nation, adaptation of precious agriculture is mandatory. While irrigation is required for growing different crops, on the other hand, making sure of proper use of water is also an utmost necessity. To minimize the wastage of water during the irrigation, we applied a decision tree regressor and random forest regressor model from machine learning algorithms. We divided our data set into two different sets and apply decision tree and random forest regressors. Excluding a few minor deficiencies, our model performed with 79.003361% and 83.003351% accuracy for the first data set using decision tree and random forest regressor respectively. Both models performed with 77.016034% and 81.016029% accuracy for the second data set. In both cases, the obtained accuracy exceeding our target threshold and performing as expected will lead us to reach the minimum loss of water during irrigation. We can extend our present work by considering the humidity, soil moisture, and cloud coverage as the input variables that may lead us to higher accuracy within the same amount of execution time. The same kind of approach can be carried out with the dataset that will be populated in Bangladesh in the future.

Acknowledgment

We wish to express my sincere appreciation to **Miyan Research Institute** of **International University of Business Agriculture and Technology**, in terms of providing the required funding and support to carry out our research work.

References

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome, 300(9), D05109.

Bata, M. T., Carriveau, R., & Ting, D. S. K. (2020). Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. Smart Water, 5, 1-18.

Bonaccorso, G. (2017). Machine learning algorithms. Packt Publishing Ltd.

Brouwer, C., & Heibloem, M. (1986). Irrigation water management: irrigation water needs. Training manual, 3.

Chen, D., Gao, G., Xu, C. Y., Guo, J., & Ren, G. (2005). Comparison of the Thornthwaite method and pan data with the standard Penman-Monteith estimates of reference evapotranspiration in China. Climate research, 28(2), 123-132.

Devi, Y. S. S., Prasad, T. K. D., Saladi, K., & Nandan, D. (2020). Analysis of precision agriculture technique by using machine learning and IOT. In Soft Computing: Theories and Applications (pp. 859-867). Springer, Singapore.

Fisher, D. K., & Pringle III, H. C. (2013). Evaluation of alternative methods for estimating reference evapotranspiration. Agricultural Sciences, 2013.

Forests, R. (2001). by Leo Breiman. Machine learning, 45(1), 5-32.

Goap, A., Sharma, D., Shukla, A. K., & Krishna, C. R. (2018). An IoT based smart irrigation management system using Machine learning and open source technologies. Computers and electronics in agriculture, 155, 41-49.

Haque, F. F., Abdelgawad, A., Yanambaka, V. P., & Yelamarthi, K. (2020, June). Crop yield analysis using machine learning algorithms. In 2020 IEEE 6th World Forum on Internet of Things (WF-IoT) (pp. 1-2). IEEE.

Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. Applied engineering in agriculture, 1(2), 96-99.

Hoque, M. (2018). Why Farmers in Bangladesh Waste 800 Liters of Water to Produce 1 Kg of Paddy.

Khan, M. A., Islam, M. Z., & Hafeez, M. (2011). Irrigation water requirement prediction through various data mining techniques applied on a carefully pre-processed dataset. Journal of Research and Practice in Information Technology, 43(22), 1-17.

Kwok, J., & Sun, Y. (2018, January). A smart IoT-based irrigation system with automated plant recognition using deep learning. In Proceedings of the 10th international conference on computer modeling and simulation (pp. 87-91).

Lang, D., Zheng, J., Shi, J., Liao, F., Ma, X., Wang, W., ... & Zhang, M. (2017). A comparative study of potential evapotranspiration estimation by eight methods with FAO Penman–Monteith method in southwestern China. Water, 9(10), 734.

Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A. M., Erkbol, H., & Fritschi, F. B. (2020). Crop Monitoring Using Satellite/UAV Data Fusion and Machine Learning. Remote Sensing, 12(9), 1357.

Mainuddin, M., Maniruzzaman, M., Alam, M. M., Mojid, M. A., Schmidt, E. J., Islam, M. T., & Scobie, M. (2020). Water usage and productivity of Boro rice at the field level and their impacts on the sustainable groundwater irrigation in the North-West Bangladesh. Agricultural Water Management, 240, 106294.

Momin, B.M. Abdul, Bangladesh Water Development Board: Ground Water Processing Branch. (2020). *Ground Water Table Hydrograph of 38 districts for the year 2008 to 2018*. Processing and Flood Forecasting Circle.

Palmer, W. C., & Havens, A. V. (1958). A graphical technique for determining evapotranspiration by the Thornthwaite method. Monthly Weather Review, 86(4), 123-128.

Panigrahi, K. P., Das, H., Sahoo, A. K., & Moharana, S. C. (2020). Maize leaf disease detection and classification using machine learning algorithms. In Progress in Computing, Analytics and Networking (pp. 659-669). Springer, Singapore.

Pasha, M. F. K., Srinivasamurthy, N., Singh, D., Yeasmin, D., & Valenzuela, G. (2020, May). An Artificial Intelligence Model to Predict Crop Water Requirement Using Weather, Soil Moisture, and Plant Health Monitoring Data. In World Environmental and Water Resources Congress 2020: Water Resources Planning and Management and Irrigation and Drainage (pp. 9-14). Reston, VA: American Society of Civil Engineers.

Pierce, F. J., & Nowak, P. (1999). Aspects of precision agriculture. Advances in agronomy, 67, 1-85.

Rezvi M. Rahman, (2018, Feb 14). Impact of climate change on agricultural. *The Financial Express.*

https://www.thefinancialexpress.com.bd/views/impact-of-climate-change-on-agricultural-1518619832

Richards, Mark (2015). Hygreveous Equation. PyEto.

https://pyeto.readthedocs.io/en/latest/index.html

Samani, Z. A., & Pessarakli, M. (1986). Estimating potential crop evapotranspiration with minimum data in Arizona. Transactions of the ASAE, 29(2), 522-0524.

Sidhu, R. K., Kumar, R., & Rana, P. S. (2020). Machine learning based crop water demand forecasting using minimum climatological data. Multimedia Tools and Applications, 79(19), 13109-13124.

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. Geographical review, 38(1), 55-94.

Vij, A., Vijendra, S., Jain, A., Bajaj, S., Bassi, A., & Sharma, A. (2020). IoT and machine learning approaches for automation of farm irrigation system. Procedia Computer Science, 167, 1250-1257.

Zhu, B., Feng, Y., Gong, D., Jiang, S., Zhao, L., & Cui, N. (2020). Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. Computers and Electronics in Agriculture, 173, 105430.

Zotarelli, L., Dukes, M. D., Romero, C. C., Migliaccio, K. W., & Morgan, K. T. (2010). Step by step calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). Institute of Food and Agricultural Sciences. University of Florida