Ensemble machine learning based model to estimate irrigation water requirement for wheat crop

Satendra Kumar Jain, Anil Kumar Gupta
Department of Computer Science and Applications, Barkatullah University, Bhopal, India

Article Info

Article history:

Received Sep 15, 2025
Revised Nov 18, 2025
Accepted Jan 1, 2026

Keywords:

Crop coefficient
Gradient boosting regressor
Irrigation water requirement
Machine learning
Reference evapotranspiration

ABSTRACT

India faces a serious water shortage issue, as its population grows faster than the percentage of fresh water available, with only 4% of the world's fresh water available to 18% of the world's population. Agriculture sector is more water-consuming sector in India. India's irrigation system still faces two significant problems: low irrigation efficiency and a lack of optimization during irrigation. To address these problems, agriculturists ought to be aware of the water requirements for crops beforehand. Innovative fields like machine learning, a branch of artificial intelligence, have a big potential to improve irrigation. Verifying the suitability of the gradient boosting regressor machine learning algorithm-based model for estimating irrigation water requirements (IWR) is the aim of this research. The experiment is conducted in Ludhiana, a city in Central Punjab, India, with a hot, semi-arid climate that features scorching summers and chilly winters. The results demonstrate the remarkably high accuracy rate with coefficient of determination (R^2) =0.98 for predicting IWR. The suggested model, which is based on a gradient boosting regression, allows the stakeholders to accurately estimate the amount of water needed for irrigation, the number of irrigation applications for the growing season of wheat crops, and the interval between irrigations.

Corresponding Author:

Satendra Kumar Jain
Department of Computer Science and Applications, Barkatullah University
Bhopal, Madhya Pradesh, India
Email: satendra.k.jain@gmail.com

1. INTRODUCTION

India is in a unique position to produce wheat globally. It is grown in most northern Indian states, including Punjab, Haryana, Madhya Pradesh, and Rajasthan. Wheat is a rabi crop which is sown in October and matures in March. It is believed that wheat requires a sufficient amount of water to grow. In India, it has been observed that farmers are generally unaware of the amount of water that is sprayed on crops during irrigation.

Significant water loss during irrigation has also been noted as a result of spills, leaks, and evaporation. Less irrigation efficiency results from it. These problems inspired us to suggest a model based on artificial intelligence that gives us remarkable precision regarding the amount of water used during irrigation. Calculating the frequency of irrigation applications and the time between irrigations may also be made easier by the model. To accurately determine the amount of water needed for irrigation, a number of meteorological variables must be considered, including temperature, wind speed, humidity, sunshine hours, precipitation, and reference evapotranspiration (ETo). The literature contains a number of direct and indirect techniques for calculating ETo.

A commonly used technique that depends on climatic parameters is the Penman-Monteith equation, Allen *et al.* [1] suggested by the Food and Agriculture Organization (FAO), United Nations. The water requirement of a crop, expressed as crop evapotranspiration (ETc), is calculated by multiplying its crop coefficient (Kc) by ETo. Together with effective rainfall, ETc is a crucial parameter for determining the amount of water needed for irrigation.

There have been numerous authors who have contributed significantly to the estimation of irrigation water requirements (IWR) through the use of direct, indirect, and machine learning-based methods, according to the literature review. Saggi and Jain [2] completed a thorough investigation into smart irrigation scheduling decision support systems. Satpute *et al.* [3] evaluated the net and gross water irrigation requirements for sixteen distinct crops in Central Punjab, both with and without effective rainfall. Akbar *et al.* [4] used CropWat software to estimate crop water requirements and irrigation scheduling for wheat in Punjab State, Pakistan, across various climatic zones. Faruq *et al.* [5] estimated the crop water requirement for rice in Bandarban, Bangladesh, using the same software. Solangi *et al.* [6] calculated the amount of water needed for the irrigation of wheat, cotton, bananas, and sugarcane.

Sharma *et al.* [7] proposed varying Kc values for every stage of wheat crop growth in Jalandhar, Punjab, under drip irrigation. A lot of research has been done in [8] to see how the ETo, moisture index, and aridity index change in Punjab State's various agroclimatic zones. In order to determine the optimal option for estimating ETo, Islam and Alam [9] evaluated the performance of fifteen empirical models. Irmak *et al.* [10] suggested solar and net radiation-based equations to estimate ETo. Saggi and Jain [11] proposed a fuzzy-genetic, regularization, and random forest-based model to estimate ETc for maize and wheat crops. Saggi and Jain [12] presented deep learning based model to estimate ETo. Pagano *et al.* [13] utilized multi-layer perceptron and random forest to forecast daily actual evapotranspiration. Meenal *et al.* [14] discussed the applicability of random forest for evapotranspiration modelling. Ravindran *et al.* [15] tested deep neural network-based model built on different climatic parameters to predict ETo. Kumar *et al.* [16] applied an artificial neural network for estimation of ETo of the western Himalayan region. Granata [17] created three models based on various climate parameters and used support vector machines, regression trees, bagging, and random forests to estimate ETo. Khosravi *et al.* [18] evaluated the capacity of adaptive neuro-fuzzy inference systems and different data mining algorithms to estimate ETo. Ferreira *et al.* [19] estimated ETo in Brazil using support vector machines and artificial intelligence. Hu *et al.* [20] have contrasted estimated ETo with physical-based, data-driven-based, and hybrid modeling.

To predict ETo, models based on artificial neural networks, random forests, lasso ridge, and gradient boosting regressors were used in [21]. Kumar *et al.* [22] estimated ETo using support vector machines, random forests, gradient boosting regressors, and long short-term memory neural networks. Hu *et al.* [23] suggested an IoT-based model that estimates ETo using a machine-learning algorithm. Nagappan *et al.* [24] investigate how well deep neural networks work for scheduling irrigation.Bastam *et al.* [25] proposed a hybrid ensemble framework for smart irrigation.

This study's goal is to create an artificial intelligence-based model for estimating irrigation water needs to save fresh water for the next generation. As agriculture sector is extracting huge underground water for irrigation, we want to reduce it by precise prediction using proposed model. The remaining portion of the article is explained in the following order: the dataset of the site under study, the machine learning-based model that is suggested to estimate the amount of water needed for irrigation, the gradient boosting regressor algorithm, the results that are discussed, and the study's conclusion.

2. METHOD

2.1. Dataset

Punjab, a state in northern India, is well known for its contribution to the green revolution. It significantly boosts the nation's production of rice and wheat. A sufficient amount of water is necessary for the growth of every crop. Numerous climatic parameters, crop characteristics, and geographic data about the study site are needed to complete this task. The district chosen for this study is Ludhiana in the Punjab state, which has hot, semi-arid agroclimatic zone. Its average elevation is 240 meters, and its coordinates are 30°54'N 75°51'E. The average amount of precipitation observed each year is 500-750 mm. From mid-March through the last week of June, the heat waves arrive. The south-west monsoon provides 78% of the precipitation that falls between July and September. During the non-monsoon months, western disturbances account for the remaining 22% of precipitation.

A dataset spanning twenty-two years, from 2000 to 2021, has been considered for experimentation. The dataset is furnished with six parameters: temperature (both low and high), humidity, wind speed, sunshine hours, and precipitation. Annual mean values of these parameters are shown in Table 1, and the mean monthly variation is depicted in Figure 1.

Table 1. Ludhiana's annual mean climate data values

Year	Max. temperature (°C)	Min. temperature (°C)	Relative humidity (%)	Wind (m/s)	ETo (mm/day)
2000	34.38	19.30	33.20	2.06	6.18
2001	33.99	19.32	34.74	1.87	5.92
2002	34.90	20.07	33.12	1.91	6.09
2003	33.08	19.05	41.39	1.96	5.70
2004	34.07	19.71	37.54	1.93	5.91
2005	32.67	18.90	41.64	1.90	5.64
2006	32.98	19.67	44.67	1.92	5.62
2007	33.31	19.56	41.78	1.89	5.70
2008	31.80	18.58	47.81	1.86	5.35
2009	33.69	19.54	38.23	1.87	5.83
2010	33.37	19.67	41.97	1.86	5.71
2011	32.06	18.92	46.77	1.78	5.37
2012	32.82	18.75	39.97	2.08	5.87
2013	32.26	18.65	47.11	1.98	5.59
2014	32.27	18.52	44.54	2.00	5.70
2015	32.00	18.49	47.71	1.94	5.44
2016	33.69	19.50	39.62	1.93	5.89
2017	32.70	18.96	44.36	1.90	5.64
2018	32.07	18.67	47.03	1.92	5.55
2019	30.81	18.37	53.93	1.82	5.18
2020	30.87	18.19	54.51	1.92	5.14
2021	32.10	18.76	49.99	1.91	5.46

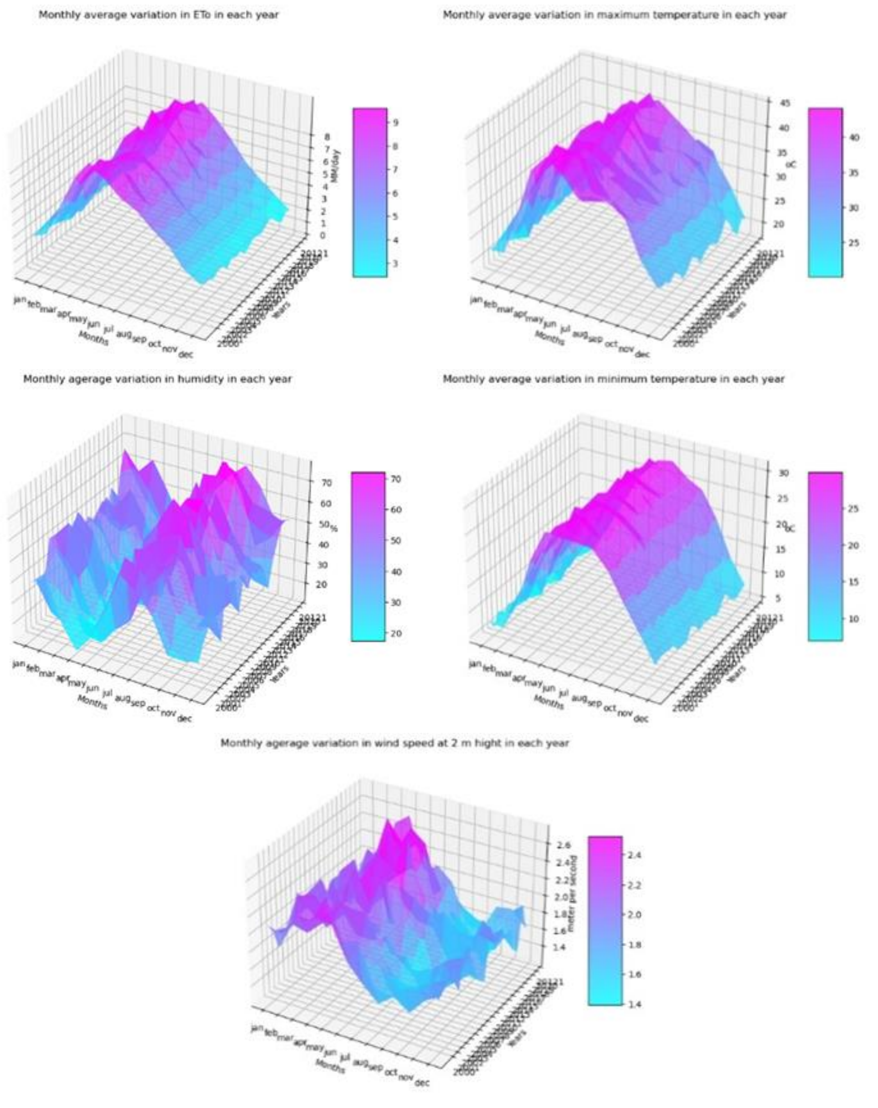


Figure 1. Average monthly changes in Ludhiana's ETo, humidity, temperature, and wind speed

In (1) is used to calculate ETo values. FAO recommended ETo values based on various agroclimatic zones. When temperatures are higher than 30 degrees, it varies between 6 and 8 millimeters per day in arid and semi-arid regions and 5 to 7 mm per day in the humid region. This range is confirmed by Table 1's mean ETo values. Punjab is a major wheat producer in India. For many people in India, wheat is a staple diet. Moreover, it is regarded as a crop that can be used to produce a wide range of food items. Many types of wheat variety, including PBW 826, PBW 869, PBW 824, PBW 803, and others, are produced in this region. To grow wheat, a large amount of land is needed. As a result, irrigating the cropland requires a significant amount of fresh water. It encourages us further to create a machine learning-based model that predicts how much water crops will need to grow. The amount of water needed for irrigation for the wheat crop is considered in this study. The standard timeframe for growing and harvesting wheat is from the last week of October to the last week of March. It takes roughly 145 to 150 days in total. Table 2 contains the information on wheat, which is used in this study at the time of estimating crop water requirement.

$$ET_0 = \frac{0.408\times(R_n - G) + \gamma\times(\frac{900}{T+273})\times u2\times(e_s - e_a)}{\Delta+\gamma\times(1+.34\times u2)} \tag{1}$$

Table 2. Wheat crop data

Crop	Duration	Date of sowing	Date of harvesting	Soil type	Growing stages	Length of growth in each stage
Wheat	146 days	28-Oct	21-22 March	Sandy loam	Intimal stage	25 days
					Developing stage	41 days
					Mid-season	51 days
					Late-season	29 days

2.2. Proposed model

The proposed model is presented in Figure 2. It is divided into four phases. The first stage involves gathering the necessary data. The FAO concept is used in the second stage to calculate observed ETo, ETc (crop water requirement), and IWR. The third step separates the combined data (climate data along with observed ETo received from second stage) into two groups: training (80%) and test (20%) datasets. The trained gradient boosting algorithm anticipates the new ETo, ETc, and IWR values in this stage. At the fourth stage, the model's performance is evaluated based on observed IWR (obtained from second stage) and predicted IWR (obtained from third stage).

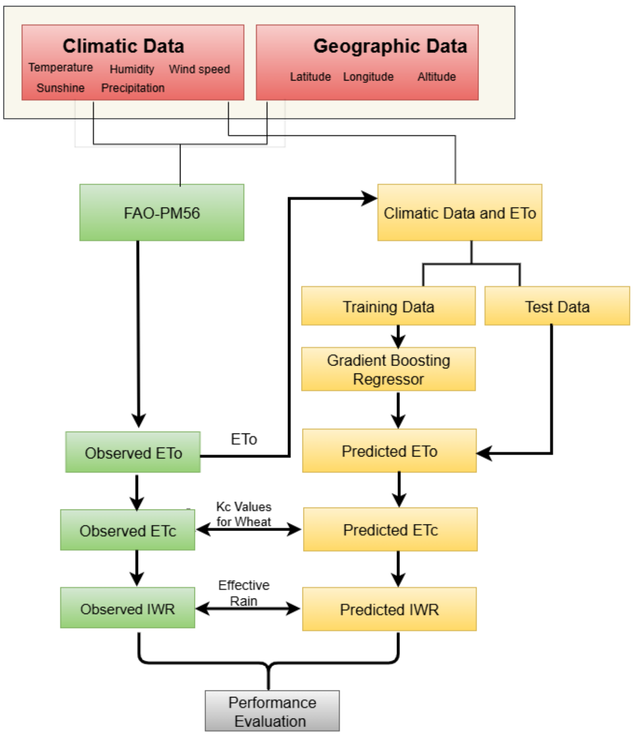


Figure 2. Ensemble machine learning based proposed model

To calculate the amount of water needed for irrigation, ETo is computed first. It is a process where water is lost from the soil and plants at the same time. Geographical information (latitude, longitude, and altitude) and the meteorological information (temperature, humidity, wind speed, and sunshine hours) are needed when estimating it for a given area. The ETo of Ludhiana is determined in this study using the MS. Excel application, which includes all necessary formulas as recommended for the FAO-56 PM approach shown in (1) and validated by the FAO-recommended CropWat 8.0 software. ETc values are derived using (2), where Kc is the crop coefficient, which comprises the crop's characteristics. The crop duration is divided into four stages, namely initial stage, development stage, mid-season stage, and late season stage. Each stage has own specific duration and Kc values. FAO suggested the standard Kc values for different crops. Wheat is taken into consideration in this study with the characteristics listed in Table 2. The Kc values are inherited from [26] in this study. Effective rainfall plays crucial role in calculating IWR. A fraction of rainfall that is needed to grow the crop and retained in the soil is called effective rainfall. The United States Department of Agriculture (USDA) soil conservation service is used to calculate the effective rainfall given in (3) and (4). This study calculates it every ten days since the crop's duration began, and every month also. If the ETc is less than the effective rainfall, no irrigation water is needed. On the other hand, IWR is equivalent to the crop water requirement when there is no effective rainfall.

$$ET_c = ET_o \times K_c \tag{2}$$

$$Effective_{rain} = Total_{rain} \times \frac{125-0.2\times Total_{rain}}{125}\ for\ Total_{rain} < 250\ mm\ per\ month \tag{3}$$

$$Effective_{rain} = 125 + 0.1 \times Total_{rain} for\ Total_{rain} >> 250\ mm\ per\ month \tag{4}$$

In addition to the traditional approach, this study measures the machine learning algorithm's applicability in estimating IWR. A new area of computer science called machine learning helps identify patterns in a wide range of fields. Among these, the fields of agriculture and irrigation are the ones that draw researchers to use artificial intelligence to improve crop productivity and optimize irrigation to conserve water. One popular ensemble-learning technique that performs well on various kinds of data is gradient boosting regressor. It can effectively classify and predict the data. A dataset comprising weather data and ETo values is used to train the gradient boosting regressor algorithm in the suggested model. The grid search method is used to determine the ideal number of trees. When test data is used, new ETc values are anticipated. Effective rain is deducted from the predicted ETc values to predict IWR values shown in (5). When supplying the observed and predicted values of IWR, the model's accuracy is verified using the root mean squared error (RMSE) and coefficient of determination (R^2) performance matrices represented in (6) to (9).

$$IWR = ETc - Effective_{rain} \tag{5}$$

2.3. Gradient boosting regressor

The gradient boosting regressor is an ensemble machine learning technique that builds decision tree one after the other sequentially so that the current one corrects the previous one's error. Gradient decent optimization is its foundation, where a user-defined differentiable loss function is minimized. Mean squared error is employed as a loss function for prediction problems. A gradient boosting regressor is renowned for its high scalability and precision. It is extensively used in a variety of fields, including health risk assessment, credit risk assessment, medical diagnosis, recognition of photos, and travel time estimation. This motivates us to verify its applicability to estimate the IWR, which is rarely found in the literature.

For this, several hyperparameters have been defined. To control the learning process of the model, hyperparameters play an important role. These are the configuration variables that can be tuned to get their optimum value to enhance the efficiency of the algorithm and avoid the under-fitting and over-fitting issues. In the context of gradient boosting regressor, numerous hyperparameters have been defined, such as n_estimator (number of decision trees), the depth of the tree and learning_rate, which are important ones that influence the accuracy and efficiency of the gradient boosting regressor algorithm.

The number of trees (n_estimator) and the tree's depth (max_depth) are two crucial hyperparameters. In the current study, we use the grid search strategy to tune the n_estimator while taking into account that the max_depth is 5. The following steps summarize the gradient boosting regressor method and are represented in Figure 3.

Step1: Base model is created by predicting the target value y with
$\mathrm{initial_{model}F_0} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Step2: For m=1 to M weak learners (number of trees), repeat following steps until the desired outcome is reached.
Step3: Residual is calculated by reducing the predicted value from the target value
residual $rs_i = y_i - F_{m-1}$
Step4: Build the weak learner on dataset $D(x, rs_i)$ that makes prediction P_m.
Step5: Add the prediction of new learner to the previous weak learner's prediction to update it.
$F_m = F_{m-1} + \text{learning\_rate} \times P_m$

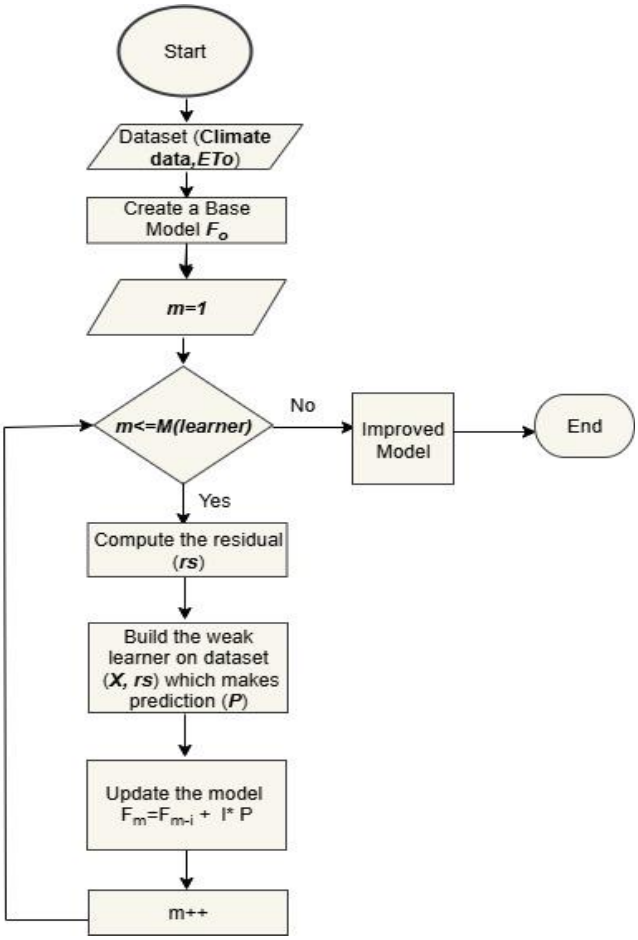


Figure 3. Flow chart of gradient boosting regressor model

2.4. Performance evaluation

Two statistical measures are used in this study to measure the capability of proposed model:

i) RMSE is the square root of the squared average of residuals:

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{6}$$

ii) R^2 is calculated as follows and shows how data points nearly lie across the regression line:

$$R^2 = \frac{sum\_of\_squared\_regression}{sum\_of\_squared\_total} \tag{7}$$

$$sum\_of\_squared\_regression = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{8}$$

$$sum\_of\_squared\_total = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{9}$$

3. RESULTS AND DISCUSSION

When calculating a crop's IWR, ETo is a significant factor. Estimating IWR for the wheat crop in Ludhiana, Punjab, is the study's goal. A variety of programs, including MS. Excel and Python libraries (Scikit-learn and Matplotlib), are employed to complete this task. Using the FAO-56 PM equation, the ETo, ETc, and IWR values are observed first. Maximum and minimum temperatures, humidity, wind speed at two meters above the ground, and sunshine hours are the input columns in the Excel sheet. The FAO-PM56 formulas are used to determine the vapor pressure deficit, net radiation, shortwave radiation, longwave radiation, and extraterrestrial radiation. To estimate the ETo of a certain region, these parameters are essential. Table 1 displays the observed annual average ETo values. Given that Ludhiana is located in a hot, semi-arid climate. In May and June, high ETo values are observed, reaching over 10 mm per day. The dataset under evaluation exhibits annual mean ETo values of roughly 5 mm/year. Daily variation of ETo is shown in Figure 4.

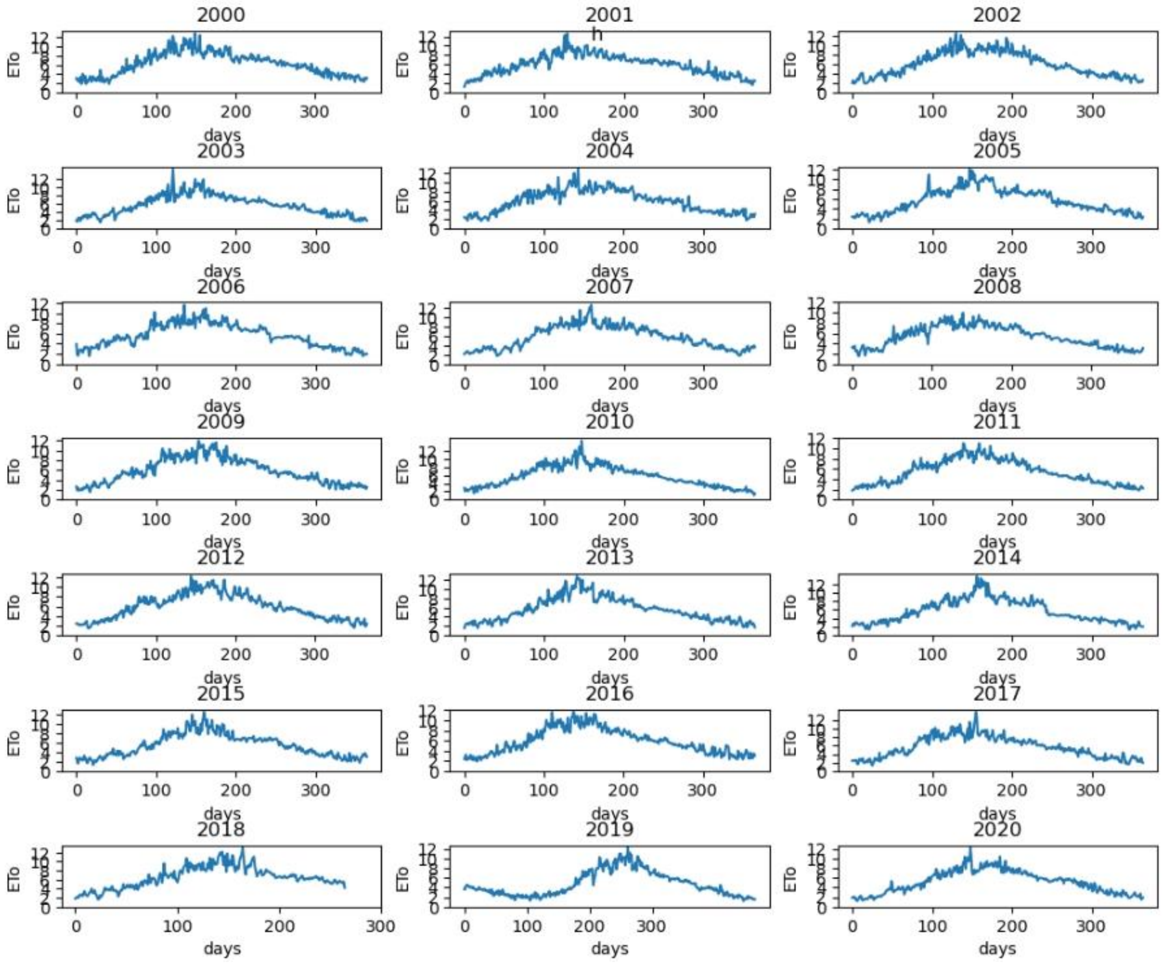


Figure 4. Daily variation in ETo values at Ludhiana

Since wheat is a crop that grows in the winter. Harvesting takes 120 to 150 days. There are four distinct stages throughout the wheat crop's entire growing season: the early, development, mid, and late stages. Each stage has a unique Kc value. The literature has shown that the Kc values for crops sown early and those sown late vary. The Kc values for wheat crop are taken from [26] and used in this investigation. The inherited Kc values are 0.39, 1.26, and 0.36, respectively, for the initial stage, mid-season stage, and the late-season stage. ETc values for the development stage are calculated using the (10).

The observed ETo and ETc values for the different growth stages are shown in Table 3 and depicted using the web chart in Figure 5. One significant finding is that ETo values are higher than ETc during the first, development, and end of the season. However, in the middle of the season, the opposite is true, with

ETc values exceeding ETo. It indicates that the crop loses more water during the mid-season and has a longer lifespan, whereas ETo contributes more in the initial, development, and end season.

$$K_{c\_of\_ith\_day} = Kc_{previous\_stage} + \left(\frac{(ith\_day - \sum length_{previous\_stage})}{length_{all\_tage}}\right) \times \left(K_{c_{next\_stage}} - K_{c_{previous\_stage}}\right) \quad (10)$$

Table 3. Observed ETo and ETc values at various wheat crop growth stages

Duration from Oct 28 to Mar 21-22	ETo_initial stage	ETc_initial stage	ETo_devp stage	ETc_devp stage	ETo_mid season	ETc_mid season	ETo_end season	ETc_end season
2000-01	4.2	1.6	3.3	2.7	3.6	4.5	5.6	2.0
2001-02	4.2	1.6	3.0	2.4	3.3	4.1	5.1	1.8
2002-03	3.7	1.5	3.1	2.5	3.0	3.8	4.8	1.7
2003-04	3.7	1.4	2.8	2.2	3.0	3.7	6.1	2.2
2004-05	3.6	1.4	3.0	2.5	2.6	3.3	4.3	1.6
2005-06	4.0	1.6	3.0	2.5	3.6	4.5	4.9	1.8
2006-07	3.4	1.3	2.5	2.0	2.9	3.6	4.0	1.5
2007-08	3.9	1.5	3.0	2.5	3.1	3.9	5.8	2.1
2008-09	3.4	1.3	2.7	2.2	2.9	3.7	5.4	1.9
2009-10	3.6	1.4	2.8	2.3	3.2	4.1	5.5	2.0
2010-11	3.4	1.3	2.3	1.9	2.8	3.6	4.8	1.7
2011-12	3.3	1.3	2.5	2.0	2.8	3.5	5.3	1.9
2012-13	3.5	1.4	2.9	2.3	2.9	3.6	4.7	1.7
2013-14	3.3	1.3	2.8	2.3	2.8	3.5	4.5	1.6
2014-15	3.6	1.4	2.7	2.1	2.8	3.6	3.8	1.4
2015-16	3.5	1.4	2.8	2.4	3.1	3.9	5.0	1.8
2016-17	3.6	1.4	3.2	2.7	2.8	3.5	4.8	1.7
2017-18	3.1	1.2	2.8	2.3	3.2	4.0	5.1	1.8
2018-19	3.0	1.2	2.3	1.8	2.3	2.9	3.7	1.3
2019-20	3.0	1.2	1.9	1.6	2.3	2.9	3.7	1.3
2020-21	3.4	1.3	2.4	2.0	2.7	3.4	5.6	2.0

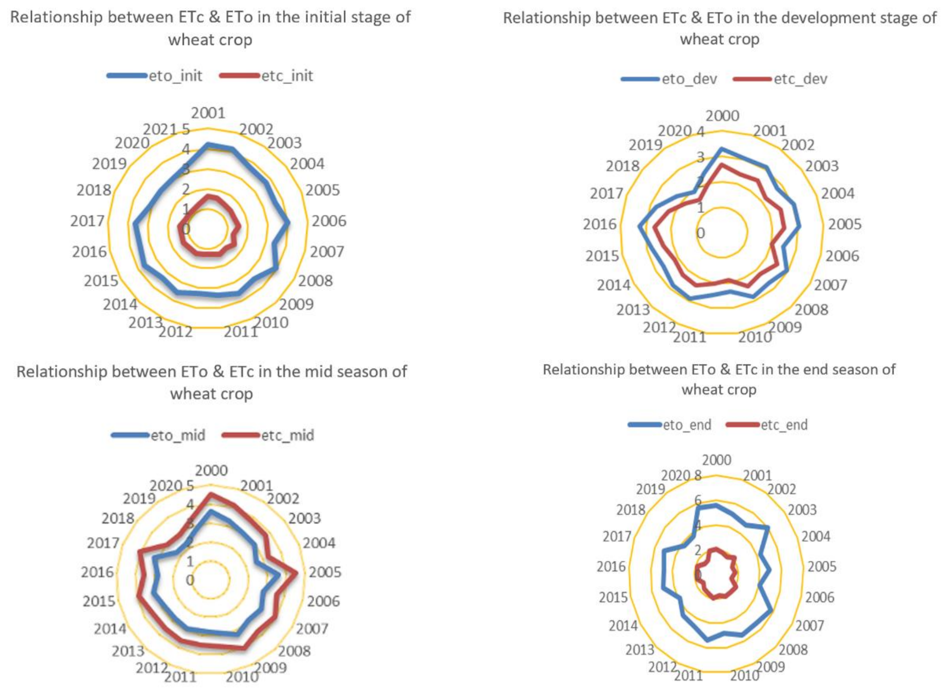


Figure 5. Relationship between ETo and ETc in the initial, development, mid, and late season

Crop growth and productivity are impacted by crop water stress, which is when a plant does not receive enough water to grow. Irrigation and rainfall can help reduce it. Some precipitation-related water evaporates and runs off. The soil does not absorb all of the precipitation, and the crop does not use it to grow. Effective water is the water that is absorbed by the soil and utilized by the crop for growth.

In this research, we examine 21 wheat-crop durations (planting to harvesting) from 2000 to 2021. Table 4 lists total rainfall, effective rainfall, observed ETc, and observed IWR information yearly for eachwheat-crop-duration. It can be noticed in the table that the greatest amount of precipitation

(207.3 mm/season) occurred between October 28, 2019, and March 21, 2020. The ETc value observed for this period is 280.2 mm/season, whereas effective rain is computed as 180.7 mm/season. The lowest amount of precipitation (21.3 mm/season) occurred between October 28, 2000, and March 22, 2001. Observed ETc value for this period is 439.2 mm/season, whereas the computed effective rain is 20.6. The IWR values for these periods are 99.6 and 418.6, respectively. An interesting finding is that crop irrigation quantity is influenced by effective rainfall. Low effective rainfall, which results from less rainfall, necessitates more water for irrigation, whereas high effective rainfall, which results from more rainfall, requires less water.

Table 4. Seasonal IWR for wheat in Ludhiana

Duration from Oct 28 to Mar 21-22	Total rain (mm)	Effective rain (mm)	Observed ETc (mm/season)	Observed IWR (mm/season)
2000-01	21.3	20.6	439.2	418.6
2001-02	43.4	41.1	402.1	361.0
2002-03	114.0	93.4	380.4	287.0
2003-04	64.5	57.4	379.9	322.5
2004-05	160.4	141.0	348.9	227.9
2005-06	52.0	48.1	421.8	379.7
2006-07	173.8	141.3	340.9	228.7
2007-08	34.2	32.8	402.2	369.3
2008-09	61.4	57.9	369.4	311.5
2009-10	25.4	24.8	395.4	370.6
2010-11	54.2	50.1	343.1	293.0
2011-12	37.5	35.7	348.2	312.5
2012-13	125.2	114.1	362.0	263.8
2013-14	85.2	79.8	353.2	273.4
2014-15	191.9	153.8	344.7	233.4
2015-16	69.4	62.5	382.2	331.4
2016-17	111.2	97.4	374.7	281.6
2017-18	73.5	67.8	382.0	315.3
2018-19	153.0	131.3	294.7	166.2
2019-20	207.3	180.7	280.2	151
2020-21	63.1	59.9	346.3	287.9

A gradient boosting regressor machine learning algorithm, could be a replacement of FAO-56 PM approach to predict the IWR. As shown in Table 5, the data set, which comprises 8036 samples (spanning the years 2000–2021), is divided into training (7890 samples) and test (146 samples) dataset. The test dataset includes samples of wheat crops grown from October 28, 2020, to March 22, 2021. The grid search approach is used in this study to initially determine the best value of the hyperparameter, n_estimator (number of decision tree). The mean test score of hyperparameter tuning is represented in Table 6. The best available hyperparameter is used to train the gradient boosting regressor using a training dataset. The model's prediction is summarized in Table 7 for the test data set.

For RMSE and R^2, the gradient boosting regressor predicts ETo values with an accuracy of 0.22 mm/day and 0.98, respectively. Regression analysis between observed ETo and predicted ETo is shown in Figure 6. Figure 6(a) presents the regression line Y=1×X+0.08, while the corresponding residual graph is shown in Figure 6(b). The Kc and predicted ETo values are taken into account when predicting ETc values. The gradient boosting regressor predicts ETc values with an accuracy of 0.19 mm/day and 0.98 for RMSE and $R^{2,}$ respectively. Regression analysis between observed ETc and predicted ETc is shown in Figure 7. Figure 7(a) presents the regression line Y=0.97×X+0.12, while the corresponding residual graph is represented in Figure 7(b). Finally, IWR is estimated with an accuracy of 0.18 mm/day and 0.98 for RMSE and $R^{2,}$ respectively. Regression analysis between observed IWR and predicted IWR is shown in Figure 8. Figure 8(a) presents the regression line Y=0.98×X+0.11, while the corresponding residual graph is represented in Figure 8(b). Table 7 shows the relationship between observed IWR and predicted IWR. It can be noticed that for the wheat crops grown from October 28, 2020, to March 22, 2021, the observed IWR is 287.9 mm/season, whereas predicted IWR is 296.0 mm/season. The accuracy demonstrated by gradient boosting regressor is summarized in Table 8.

Table 5. A division of training and test data

Dataset duration	Total samples	Training dataset	Test dataset
01-Jan-2000 to 31-Dec-2021	8,036	7,890 (01-Jan-200 to 27-Oct2020) and (23-Mar-2021 to 31-Dec-2021)	146 (28-Oct-2020 to 22-Mar-2021)

Table 6. Results of adjusting hyperparameters for varying tree counts

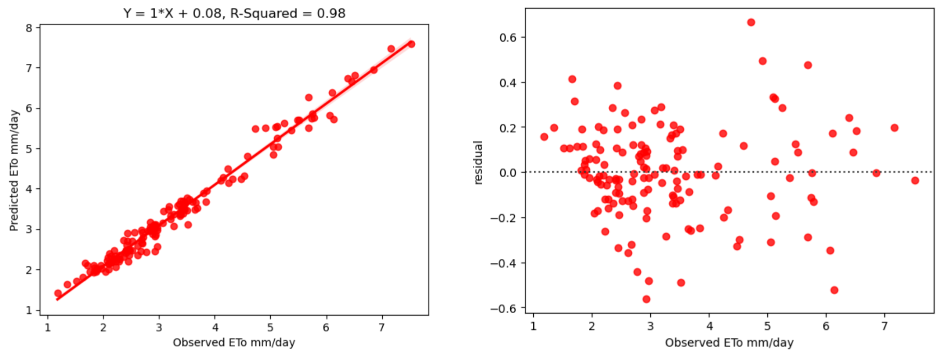
n_estimator	10	20	30	40	50	60	70	80	90	100
max-depth	5	5	5	5	5	5	5	5	5	5
Mean test score	0.834	0.949	0.968	0.973	0.975	0.976	0.977	0.977	0.977	0.977

Table 7. Predicted IWR for wheat crop in Ludhiana

28 Oct 2020 to 22 Mar 2021	Decade	Rain (mm/dec.)	Effective rain (mm/dec.)	Observed ETc (mm/dec.)	Predicted ETc (mm/dec.)	Observed IWR (mm/dec.)	Predicted IWR (mm/dec.)
Oct	1	0.0	0.0	6.4	7.1	6.4	7.1
Nov	1	0.0	0.0	13.9	14.1	13.8	14.0
	2	13.6	12.7	11.3	11.4	0.0	0.0
	3	6.6	6.3	13.1	14.0	6.8	7.6
Dec	1	0.5	0.5	19.1	20.6	18.6	20.1
	2	8.4	8.1	20.6	21.3	12.5	13.3
	3	3.9	3.8	26.7	27.8	22.9	24.0
Jan	1	17.5	16.1	21.4	24.4	5.4	8.3
	2	0.0	0.0	29.2	29.5	29.2	29.5
	3	1.1	1.1	37.3	33.9	36.2	32.8
Feb	1	5.6	5.5	39.0	41.5	33.5	36.1
	2	0.0	0.0	44.8	45.2	44.8	45.2
	3	1.1	1.1	17.2	17.5	16.1	16.5
Mar	1	1.2	1.2	20.2	20.6	19.0	19.4
	2	0.7	0.7	21.4	21.3	20.7	20.6
	3	2.8	2.8	4.9	5.1	2.1	2.3

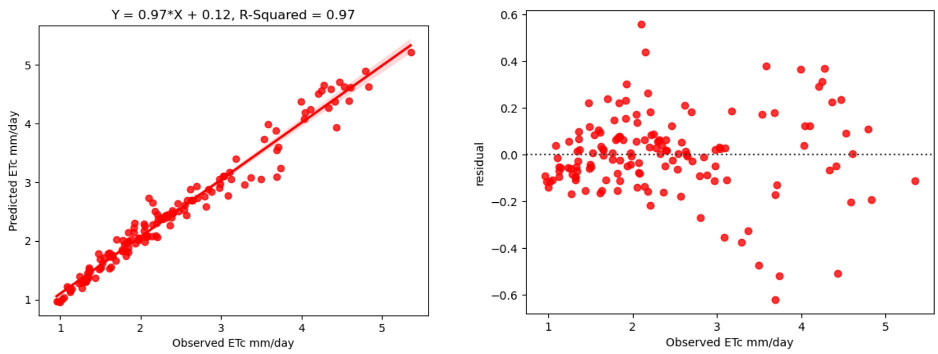
Table 8. Gradient boosting regression's accuracy in predicting ETo, ETc, and IWR

Parameter	ETo	ETc	IWR
RMSE	0.22 mm/day	0.19 mm/day	0.18 mm/day
R^2	0.98	0.98	0.98



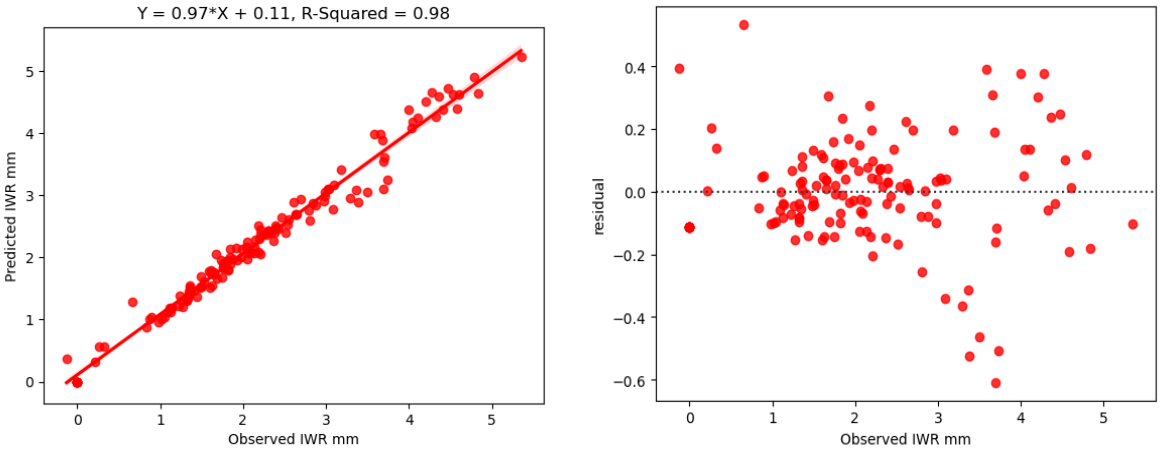
(a) (b)

Figure 6. Relationship between observed and predicted ETo: (a) regression line Y=1×X+0.08 and (b) corresponding residual graph



(a) (b)

Figure 7. Relationship between observed and predicted ETc: (a) regression line Y=0.97×X+0.12 and (b) corresponding residual graph



(a) (b)

Figure 8. Relationship between observed and predicted IWR: (a) regression line Y=0.98×X+0.11 and (b) corresponding residual graph

ETc is a requisite to estimate IWR, which needs the ETo and Kc. In the proposed model, the ETo is estimated by the FAO-PM, which is very dependent on climatic parameters such as temperature and wind speed. Solar radiation and humidity. Still, many locations, especially under developed countries across the world, are not furnished with modern sensors or tools that generate the accurate values of the mentioned climatic data. Unreliable and incomplete data affect the accuracy of the proposed model. Apart from climatic data, the crop type, its height, and the crop canopy covering the surface are also the influencing factors that affect the model's accuracy. These crop characteristics are incorporated in the form of Kc. Since FAO suggested standard Kc values for the specific crop. It is said that these values may be affected by the local climatic conditions. Adjustments to these values are required according to the climatic conditions of the study site. The suggested model has been evaluated on authentic data and can be used for different crops and climatic zones as long as it has a trustworthy dataset. The proposed model is suitable for crops having a precise duration of their growth life, along with Kc and favorable climatic conditions. Since the proposed model is tested for wheat. It would be applied to more water-intensive crops like rice, sugarcane, and cotton.

4. CONCLUSION

India's largest employer, the agriculture sector, accounts for about 18% of the nation's GDP. This industry might be important to the government's vision for New India. The state of Punjab is the primary producer of wheat. Due to the massive extraction of groundwater for irrigation, the state is currently experiencing a water crisis. Irrigation is one of the many factors that affect a crop's productivity. India's irrigation system still faces two significant problems: low irrigation efficiency and a lack of optimization during irrigation. This study assesses the suitability of machine learning, a branch of artificial intelligence, in light of these issues. To estimate the IWR for the wheat crop in Ludhiana, Punjab, a model based on a gradient boosting regressor is suggested. Experimental results show that the suggested model has an impressive accuracy R^2 of 0.98 in predicting evapotranspiration, crop water requirements, and IWR. The stakeholders can precisely estimate the quantity of water required for irrigation, the number of irrigation applications for the growing season of wheat crops, and the time between irrigations with the help of the proposed model, which is based on a gradient boosting regression. Enhancing crop productivity and water management will boost the country's economy and put an end to the problem of starvation in developing countries. The suggested model will be used in the future to optimize the scheduling and application of irrigation for the numerous water-consuming crops, such as rice, cotton, and sugarcane.

ACKNOWLEDGMENTS

We would like to thank the Department of Computer Science and Applications, Barkatullah University, Bhopal, for providing research facilities to carry out this research work.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Satendra Kumar Jain	✓		✓	✓	✓			✓	✓		✓			
Anil Kumar Gupta		✓				✓	✓			✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data for this study has been taken from the Indian Meteorological Department, Pune, India.

REFERENCES

[1] R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, *Crop evapotranspiration: guidelines for computing crop water requirements*. Rome, Italy: FAO - Food and Agriculture Organization, 1998.

[2] M. K. Saggi and S. Jain, "A survey towards decision support system on smart irrigation scheduling using machine learning approaches," *Archives of Computational Methods in Engineering*, vol. 29, no. 6, pp. 4455–4478, 2022, doi: 10.1007/s11831-022-09746-3.

[3] S. Satpute, M. C. Singh, and S. Garg, "Assessment of irrigation water requirements for different crops in central Punjab, India," *Journal of Agrometeorology*, vol. 23, no. 4, pp. 481–484, 2021, doi: 10.54386/jam.v23i4.183.

[4] S. Akbar, A. Riaz, T. Mahmood, A. J. Hashmi, and D. H. Kazmi, "A guide to irrigation requirement of wheat crop in FAO-developed agro-climatic zones of Punjab, Pakistan," *Discover Agriculture*, vol. 3, no. 1, 2025, doi: 10.1007/s44279-025-00189-5.

[5] M. O. Faruq, S. A. Turash, N. Fattah, and I. K. Nahid, "Estimation of crop water requirement for boro rice in Bandarban, Bangladesh based on CropWat 8.0 model," *Asian Journal of Plant and Soil Sciences*, vol. 6, no. 1, pp. 31–38, Aug. 2021, doi: 10.56557/ajopss/2021/v6i147.

[6] G. S. Solangi *et al.*, "Investigation of irrigation water requirements for major crops using CropWat model based on climate data," *Water*, vol. 14, no. 16, 2022, doi: 10.3390/w14162578.

[7] V. Sharma, N. M. Changade, D. A. Madane, K. K. Yadav, and S. B. Tarate, "Development of growth-stage specific crop coefficient for drip irrigated wheat crop grown in climatic conditions of Jalandhar, Punjab," *Journal of Agrometeorology*, vol. 26, no. 3, pp. 356–361, 2024, doi: 10.54386/jam.v26i3.2628.

[8] S. Aatralarasi, P. K. Kingra, R. K. Setia, and S. P. Singh, "Variability in reference evapotranspiration and moisture availability in three agroclimatic zones of Punjab, India," *Journal of Agrometeorology*, vol. 23, no. 2, pp. 154–162, Jun. 2021, doi: 10.54386/jam.v23i2.61.

[9] S. Islam and A. K. M. R. Alam, "Performance evaluation of FAO Penman-Monteith and best alternative models for estimating reference evapotranspiration in Bangladesh," *Heliyon*, vol. 7, no. 7, 2021, doi: 10.1016/j.heliyon.2021.e07487.

[10] S. Irmak, A. Irmak, R. G. Allen, and J. W. Jones, "Solar and Net radiation-based equations to estimate reference evapotranspiration in humid climates," *Journal of Irrigation and Drainage Engineering*, vol. 129, no. 5, pp. 336–347, Oct. 2003, doi: 10.1061/(asce)0733-9437(2003)129:5(336).

[11] M. K. Saggi and S. Jain, "Application of fuzzy-genetic and regularization random forest (FG-RRF): Estimation of crop evapotranspiration (ETc) for maize and wheat crops," *Agricultural Water Management*, vol. 229, Feb. 2020, doi: 10.1016/j.agwat.2019.105907.

[12] M. K. Saggi and S. Jain, "Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning," *Computers and Electronics in Agriculture*, vol. 156, pp. 387–398, 2019, doi: 10.1016/j.compag.2018.11.031.

[13] A. Pagano *et al.*, "Machine learning models to predict daily actual evapotranspiration of citrus orchards under regulated deficit irrigation," *Ecological Informatics*, vol. 76, 2023, doi: 10.1016/j.ecoinf.2023.102133.

[14] R. Meenal, P. K. Jala, R. Samundeswari, and E. Rajasekaran, "Crop water management using machine learning-based evapotranspiration estimation," *Journal of Applied Biology & Biotechnology*, vol. 12, no. 2, pp. 198–203, 2024, doi: 10.7324/JABB.2024.155791.

[15] S. M. Ravindran, S. K. M. Bhaskaran, and S. K. N. Ambat, "A deep neural network architecture to model reference evapotranspiration using a single input meteorological parameter," *Environmental Processes*, vol. 8, no. 4, pp. 1567–1599, 2021, doi: 10.1007/s40710-021-00543-x.

[16] N. Kumar, S. Maharshi, A. Poddar, and V. Shankar, "Application of artificial neural networks for estimating reference evapotranspiration in Western Himalayan Region," in *Roorkee Water Conclave 2020*, Roorkee, Uttarakhand, India: Indian Institute of Technology Roorkee and National Institute of Hydrology, 2020, pp. 1–9.

[17] F. Granata, "Evapotranspiration evaluation models based on machine learning algorithms—a comparative study," *Agricultural Water Management*, vol. 217, pp. 303–315, 2019, doi: 10.1016/j.agwat.2019.03.015.

[18] K. Khosravi *et al.*, "Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: a case study in Iraq," *Computers and Electronics in Agriculture*, vol. 167, 2019, doi: 10.1016/j.compag.2019.105041.

[19] L. B. Ferreira, F. F. da Cunha, R. A. de Oliveira, and E. I. F. Filho, "Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – a new approach," *Journal of Hydrology*, vol. 572, pp. 556–570, 2019, doi: 10.1016/j.jhydrol.2019.03.028.

[20] X. Hu, L. Shi, G. Lin, and L. Lin, "Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation," *Journal of Hydrology*, vol. 601, 2021, doi: 10.1016/j.jhydrol.2021.126592.

[21] A. Bijlwan, S. Pokhriyal, R. Ranjan, R. K. Singh, and A. Jha, "Machine learning methods for estimating reference evapotranspiration," *Journal of Agrometeorology*, vol. 26, no. 1, pp. 63–68, 2024, doi: 10.54386/jam.v26i1.2462.

[22] U. Kumar, "Modelling monthly reference evapotranspiration estimation using machine learning approach in data-scarce North Western Himalaya region (Almora), Uttarakhand," *Journal of Earth System Science*, vol. 132, no. 3, 2023, doi: 10.1007/s12040-023-02138-6.

[23] Z. Hu, R. N. Bashir, A. U. Rehman, S. I. Iqbal, M. M. A. Shahid, and T. Xu, "Machine learning based prediction of reference evapotranspiration (ET0) using IoT," *IEEE Access*, vol. 10, pp. 70526–70540, 2022, doi: 10.1109/access.2022.3187528.

[24] M. Nagappan, V. Gopalakrishnan, and M. Alagappan, "Prediction of reference evapotranspiration for irrigation scheduling using machine learning," *Hydrological Sciences Journal*, vol. 65, no. 16, pp. 2669–2677, 2020, doi: 10.1080/02626667.2020.1830996.

[25] M. Bastam, D. H. Majeed, and M. Babagoli, "A hybrid ensemble framework for smart irrigation: optimizing water management in precision agriculture," *International Journal of Engineering*, vol. 39, no. 3, pp. 705–726, 2026, doi: 10.5829/ije.2026.39.03c.12.

[26] J. Kaur, K. K. Gill, S. Kaur, and R. Aggarwal, "Estimation of crop coefficient for rice and wheat crops at Ludhiana," *Journal of Agrometeorology*, vol. 19, no. 2, pp. 170–171, 2017, doi: 10.54386/jam.v19i2.716.

BIOGRAPHIES OF AUTHORS

Satendra Kumar Jain is a Ph.D. research scholar of the Department of Computer Science and Application at Barkatullah University. His research interest is in the field of machine learning and its applications in various domains. He can be contacted at email: satendra.k.jain@gmail.com.

Anil Kumar Gupta is an associate professor of the Department of Computer Science and Applications at Barkatullah University, Bhopal, India. He did a Ph.D. in Computer Science. His research interest is in machine learning and image processing. He can be contacted at email: akgupta_bu@yahoo.co.in.