Egypt. J. Soil Sci. **Vol. 65**, No. 3, pp: 1313 - 1330 (2025)

Egyptian Journal of Soil Science
http://ejss.journals.ekb.eg/

Long-Term Meteorology Data Analysis and Machine Learning-Driven Evapotranspiration Prediction for Optimized Irrigation Management

Yasser E. Arafa [1*], **Ali A. Abd El-Aziz** [2] and **Abed Alataway** [3]

[1] *Agricultural Engineerin gricultural Engineering Department, Faculty of Agriculture, Ain Shams University, Shoubra El-Kheima, Cairo, Egypt*
[2] *Soil Chemistry and Physics Department, Desert Research Center, 1 Mathaf El-Matareya Street, Cairo, Egypt*
[3] *Prince Soltan Institute for Environmental,Water and Desert Research, King Saud University, KSA*

CLIMATE-INDUCED stress on irrigation water usage is escalating rapidly across the globe, especially in arid and semi-arid regions like Egypt, where climate change and water scarcity create unsustainable conditions for agriculture. Notably, the Development of innovative and practical methods for estimating reference evapotranspiration (ETo) is important for efficient irrigation scheduling. This study evaluated the efficacy of six machine learning (ML) algorithms, Linear Regression (LR), K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Decision Tree (DT), Random Forest (RF), and XGBoost, in estimating ETo, utilizing long-term meteorological variables from publicly available datasets. Three established empirical models served as baseline comparisons: FAO Penman-Monteith (PM), Hargreaves (HA), and Blaney-Criddle (BC). Each model was assessed using historical daily meteorological data retrieved from the NASA POWER database, which provides reliable long-term climate records relevant to agricultural applications. Model performance was evaluated based on a test using three statistical metrics: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). When comparing the ETo estimation methods, the Blaney-Criddle (BC) equation used in combination with ML models displayed the most accurate predictions. In terms of ML algorithms, Random Forest (RF) consistently outperformed other algorithms with RC^2 = 0.98 and RMSE =0.12 mm/day when using the BC equation during testing. Support Vector Regression (SVR) performed well for all models as well. RF appeared to be the best ML algorithm, and the BC equation was the best ETo model for the study area and conditions studied. It supports the use of ML models to enhance ETo estimation with limited meteorological data, particularly evident in water-scarce surroundings such as Egypt. The current study aims to fill the gap of localized ETo estimation models in Egypt by comparing ML predictions to traditional empirical models using long-term climatic data, thus providing a valuable contribution to the body of research on precision irrigation through the use of imperfect use of data. A specific focus on the Behera Governorate of Egypt determines a relationship between prioritized adaptive irrigation models where temperature rises and variable rainfall, complexifying irrigation demands, and increasing evaporation. The main aim of this study is to evaluate and compare traditional ETo estimation equations and machine learning algorithms to determine the most accurate and robust method for ETo prediction in arid climates with limited data availability.

Keywords: Machine learning; Evapotranspiration estimation; Random Forest; Blaney-Criddle; Arid agriculture; Data-driven irrigation.

1. Introduction

Water scarcity is a significant problem in many regions, and the combination of factors like poor standards of water management, climate change, and population growth makes it difficult to produce crops sustainably - especially so in arid and semi-arid regions like Egypt (Emran et al., 2024). Working towards this requires maximizing water use efficiency, which ultimately relies on accurately estimating reference evapotranspiration (ETo). ETo is an important variable in irrigation scheduling, crop productivity, and long-term agricultural water management (Agyeman et al., 2024; Rashad, 2024). In conditions limited by water, such as Egypt, this means estimating ETo as accurately as possible to maximize the effective use of available water resources (Raza et al., 2020; Allen et al., 1998). ETo is defined as the potential water loss by plant transpiration and soil evaporation for a hypothetical reference crop under standard conditions (Allen et al., 1998; Chouaib et al., 2022). The development of ETo estimation methods can be empirical, relying on physical models in either case, which are based on the analysis of meteorological data. Direct measurements of ETo, such as lysimeters and pan evaporation, estimate actual ET (ETa) rather than ETo, which is beyond the scope of this study. Pan evaporation

*Corresponding author e-mail: yasser_arafa@agr.asu.edu.eg
Received: 25/04/2025; Accepted: 16/06/2025
DOI: 10.21608/EJSS.2025.378816.2137

is a popular hydrological method that allows for direct measurements; hence, it is not considered in this discussion of indirect estimation methods.

The FAO-56 Penman-Monteith (PM) equation, although an indirect method, is the most widely accepted because it has been proven effective in various climates (Allen et al., 1998). The advantage of the PM is that it only requires key weather variables (temperature, relative humidity, solar radiation, and wind speed) to estimate ETo. Zhang et al. (2023) noted that temperature and solar radiation were the most significant variables in estimating ETo using PM. In areas with limited meteorological data coverage, simpler methods, such as the Hargreaves-Samani (HA) and Blaney-Criddle (BC) equations, are often employed (Hargreaves & Samani, 1985; Thongkao & Kongchu, 2022). These models require fewer data inputs and are beneficial as alternatives to the more data-intensive models when available climate data is not comprehensive. However, they do not include key climatic factors, such as humidity and wind speed. They may not reliably provide estimates of ETo in certain climates, which can result in an underestimation of ETo in conditions of high wind (Shiri & Yaseen, 2021). Efforts have been made to enhance the regional applicability of the HA method, such as incorporating climate coefficients applicable to Iran (Ogunrinde et al., 2022). Similarly, the BC method has reasonable accuracy for temperate climates; however, it demonstrates less reliability in arid environments or those with rapid changes (Momen & Abdelatti, 2021; Mobilia, 2021).

To overcome the limitations of conventional approaches, particularly in situations where data may be missing or the conditions are complex, researchers are increasingly using machine learning (ML) approaches to estimate ETo. ML models offer flexibility because they can learn complex, non-linear relationships in data and adapt to noisy or missing data (Farooque et al., 2022a; Chen & Guestrin, 2016). Several algorithms (Linear Regression (LR), K-Nearest Neighbors (KNN), Decision Trees (DT), Support Vector Regression (SVR), Random Forest (RF), Extreme Gradient Boosting (XGBoost)) have been proven useful in numerous agro-climatic environments. Several studies have characterized ML algorithms, particularly Random Forest and Support Vector Regression, as having the potential to accurately and reliably provide ETo predictions under different environmental conditions (Gong et al., 2021a; Kar et al., 2021; Kumar & Mishra, 2024a,b; El Azhar et al., 2024a). Other methods, including certain forms of KNN and XGBoost, also provided very high accuracy, provided proper hyperparameter tuning was performed. Granata (2019) and Granata & Nunn (2021) have also employed ML methods for ETo modeling in data-limited regions, reporting compelling results despite considerable climatic differences. Validation methods, such as walk-forward validation, were frequently used to assess the reliability of predictions over the specified time results (Farooque et al., 2022b). This approach enables models to retrain iteratively with new data, facilitating continuous evaluation in dynamic settings and allowing ETo models to transition toward more reliable predictions. While my models are extremely flexible, they present modeling aspects and, therefore, data preprocessing issues, as well as adequate algorithm selection and parameter tuning that need to be considered to provide reliable predictions (Dos Santos Farias et al., 2020). These modeling conditions would also be present in traditional ETo equations and would remain relevant when applied to separate ecological zones from different climatic regions.

This study set out to assess the accuracy and reliability of traditional methods for estimating ETo (Penman-Monteith, Hargreaves-Samani, and Blaney-Criddle) in Egypt, which has a predominantly arid climate, and compare these estimates with six machine learning models (LR, KNN, DT, RF, SVR, and XGBoost) trained on local meteorological datasets. The intention was to identify the best-performing method that would produce ETo estimates that were both precise and practical in water-limited agricultural regions.

2. Materials and methods

2.1. Experimental

The study was conducted on a farm located on reclaimed land in the Behera Governorate of Egypt (30° 38'29.24"N, 30° 0'38.99" E). As seen in Figure 1, the farm grows three-year-old Valencia orange trees spaced 5 × 4 meters apart and irrigated using a drip system (GR) with emitters delivering water at 4 L/hour positioned 50 cm from each tree. Long-term climatic data (1985–2022) were sourced from the NASA POWER database (Sparks, 2018; White et al., 2020). The statistical description of key weather parameters was as follows: minimum, maximum, and mean air temperatures were 14.6°C, 28.4°C, and 20.7°C, respectively; relative humidity averaged 59.0% with a standard deviation of 6.3%; wind speed averaged 2.9 m/s; and net solar radiation at the surface was 20.0 MJ/m²/day (with a top-of-atmosphere value of 31.4 MJ/m²/day). Annual precipitation ranged between 13.9 mm (in 2010) and 252.9 mm (in 2020), with a long-term mean of 73.3 mm and a coefficient of variation (CV) of 58.2%. Reference evapotranspiration (ETo) values are not presented in this section, as detailed results and equations used (e.g., FAO Penman-Monteith) are provided later in the Results section.

2.2. Data acquisition

Meteorological data were obtained from the NASA POWER database, a long-term global climate database suitable for agro-climate modeling (Sparks, 2018; White et al., 2020). The dataset consists of daily data for air temperature, relative humidity, wind speed, solar radiation, and precipitation from 1985 to 2022. Data were downloaded from the NASA website (Figure 2), where data are available for download, and users select the desired measurements by geocoding coordinates. The data used in this study are also visible via NASA POWER https://power.larc.nasa.gov. These daily data were then used to calculate reference evapotranspiration (ETo) using ETCalc, an easy-to-use online calculator that employs the FAO Penman-Monteith equation for ETo calculations, as well as the Hargreaves and Blaney-Criddle methods for specific climates (Danielescu, 2022a; Danielescu, 2022b; Schomberg et al., 2023). Several studies, including those conducted in semi-arid climates, have validated ETCalc outputs (Danielescu, 2022a; Schomberg et al., 2023). Despite not conducting a lysimeter comparison, Behera's climate could be likened to previous ETCalc validations. Therefore, this reinforces its use under Egyptian conditions. For further applications, researchers may consider using REF-ET generated by Idaho University, which offers more functionality for estimating ETo in various environments. See links for more details about REF-ET software https://www.uidaho.edu/cals/kimberly-research-and-extension-center/research/water-resources/ref-et-software) and ETCalc software https://www.etcalc.com.

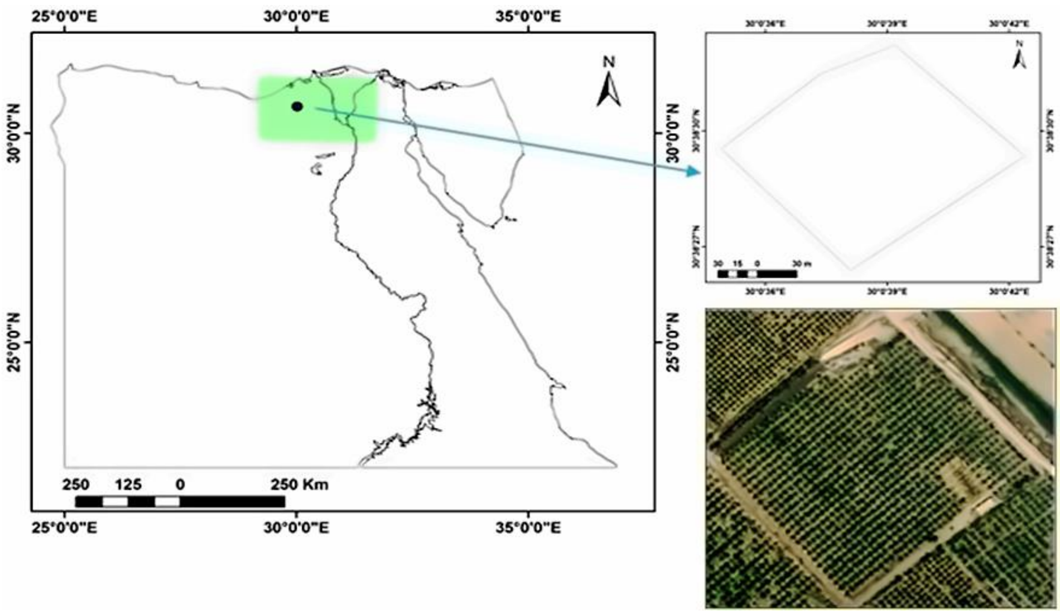

Fig. 1. The location of study area.

Fig. 2. NASA's POWER Data Access Viewer.

2.3. The FAO-Penman-Monteith (FAO-PM) equation

ETo is usually computed using the FAO-PM formula. (Allen et al., 1998) provide it as an equation 1.

$$\mathbf{ETo} = \frac{0.408\Delta(Rn-G) + \gamma \dfrac{900}{T+273}[es-ea]U_2}{\Delta + \gamma\,(1+0.34 * U_2)} \qquad (1)$$

Where:

ETo : reference crop evapotranspiration, mm/day.
Rn : net radiation, $MJ/m^2/day$.
G : soil heat flux $MJ/m^2/day$.
T : average daily air temperature at the height of 2 m, °C.
u_2 : wind speed at a height of 2 m, m/s.
es : saturation vapor pressure, kPa.
ea : actual vapor pressure, kPa.
es – ea : vapor pressure deficit, kPa.
Δ : slope of the saturation vapor pressure- temperature curve, kPa /°C.
γ : is the psychometric constant, kPa/ °C.

2.4. Hargreaves equation (ETo_HA)

The Hargreaves model is one of the more often used variants of a prior evapotranspiration model (Tabari, 2010). As described by Hargreaves and Allen 2003, this model's form is shown by equation 2.

$$\mathbf{ETo = 0.0023Ra\,(Ta + 17.8)\,(Tmax - Tmin)^{0.5}} \quad (2)$$

Where:

Ra : water equivalent of extraterrestrial radiation, mm/day.
Ta : mean air temperature, °C.
T_{max} : daily maximum temperature, °C.
T_{min} : daily minimum temperature, °C.

2.5. Blaney-Criddle equation (ETo_BC)

The original Blaney-Criddle equation (Doorenbos, J.; Pruitt, 1977) determines ETo using the mean daily air temperature and The equation used to determine the average hourly distribution throughout the day is given as follows:

$$\mathbf{ETo = a + b\,[p(0.46T + 8.13)]} \quad (3)$$

Where:

a , b : calibrated constants.
p : the average daily percentage of total annual daytime hours.
T : the average daily air temperature, °C.

2.6. Dataset Pre- processing

Data pre-processing converts unstructured data into a consistent, clear form to improve model accuracy and performance, a fundamental step in machine learning. This stage comprises various processes: feature transformation, feature scaling, feature selection and the handling of outliers and missing values. The pre-processing phase prepares the training and testing datasets required for model building and evaluation. Successful machine-learning projects depend on a well-planned pre-processing step (Eke et al., 2020).

2.6.1 Handling missing values, outliers, and normalization

Missing data can distort forecasting models and increase the sensitivity to outliers, necessitating effective management techniques (Lin & Tsai, 2020). Imputation techniques (such as mean or median imputation) can substitute missing data (Khan & Hoque, 2020). Outliers differ significantly from the remaining data, which can cause distortions in analysis. Being able to identify and address outliers using robust methods, such as the Z-score and IQR, is essential when managing data (Singh & Patna, 2022). Lastly, normalization is the process of ensuring that the data values fall within a common range (0-1) using the min-max method, allowing the model to learn efficiently and minimizing the severity of weighting larger values (Soomro et al., 2022).

$$X_{norm} = \frac{X_n - X_{min}}{X_{max} - X_{min}} \qquad (4)$$

Where:
X_{norm} : the homogenized value.
X_n : the adjusted value.
X_{max} : the highest value.
X_{min} : the lowest value of the details.

2.7. Train test split

In time series analysis, time is a crucial factor. The train-test split involves dividing the dataset into two groups: a training dataset (historical) and a testing dataset (future). This division method is frequently employed in outdoor conditions forecasting (e.g., weather and energy). When modeling a time series, the testing dataset consists of future data to validate the model's performance. In contrast, the training dataset consists of past data that enables the model to learn. In the current research, 80% of the data was used for training (January 1, 1985 to December 31, 2015), and 20% was used for testing (January 1, 2016 to December 31, 2022) (Gul et al., 2022). The split is illustrated in Figure 3.

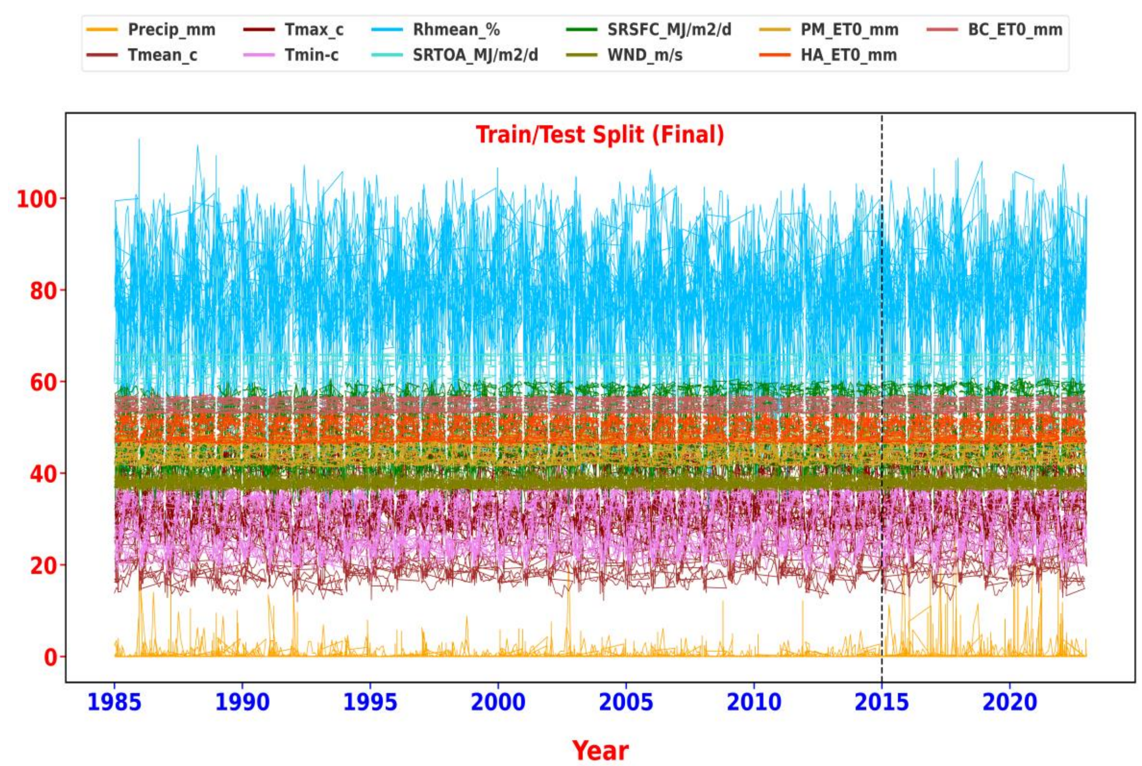

Fig. 3. illustrates the data split into training and testing sets.

2.8. Machine Learning Models for Predicting reference evapotranspiration (ETo)

The ML in this study utilized inputs of maximum and minimum air temperature, relative humidity, solar radiation, and wind speed, all of which directly affect ETo. The same variables were consistently used across models. Temperature influences evaporation; relative humidity affects the moisture in the air; solar radiation provides the energy for evaporation, and wind speed promotes the transport of moisture away. Hyperparameter tuning with GridSearchCV using 10-fold cross-validation was performed for each season to minimize overfitting. The data was divided into 10 sets, with the models trained on nine sets of the data and validated on the other. Hyperparameters such as the number of trees for Random Forest (n_estimators), maximum depth for RF (max_depth), C, and epsilon for Support Vector Regression (SVR) were tuned based on past literature and evaluation. Models were evaluated based on RMSE and R^2 and optimized to select the combination with the lowest RMSE and the highest R^2 from a set of trials. Tuning was performed based on the climate in the study region to optimize the model for agricultural practices further.

2.8.1. Linear Regression

Linear regression is a popular statistical technique and machine learning method used to determine a linear relationship between one or more independent variables and a dependent variable. Linear regression is simple, common, and effective in both data analysis and machine learning (Maulud & Abdulazeez, 2020). There are two versions of linear regression: simple regression (with one independent variable) and multiple linear regressions (with multiple independent variables) (Maulud & Abdulazeez, 2020). Linear regression has been demonstrated to be effective for modeling reference evapotranspiration (ETo) (Belayneh et al., 2014; Kisi and Kim, 2015; Jain et al., 2022), as well as for forecasting hydrology and agriculture.

2.8.2. K-Nearest Neighbours

The K-Nearest Neighbors (KNN) approach can distinguish data when uncertainty is present. The KNN method requires you to select an appropriate K-value, which determines the number of nearest data points to consider in our prediction. If we apply a K-value that is too high, we may create uncertainty; if we apply a K-value that is too low, we may overlook the noise. We also recommend using an odd K-value to mitigate errors (Bansal et al., 2022). KNN has been applied to the modeling of reference evapotranspiration (ETo). Shiri et al. (2012), Emamgholizadeh et al. (2014), and Ahoojelay et al. (2021) provide evidence in support of the KNN method as a promising approach for forecasting hydrologic variables in any global climate.

2.8.3. Extreme Gradient Boosting

Boosting is an ensemble method that enhances accuracy by combining the predictions of weak learners to create stronger ones. XGBoost is a gradient-boosting methodology that develops an ensemble of decision trees (Chen & Guestrin, 2016) by assigning weights to features based on their importance in the model. This method evaluates variance about optimally fitting the model to the tree ensemble (Wu & Fan, 2019). The effectiveness of XGBoost stems from its ability to leverage parallel computing and forgoing overfitting in decision trees (Ghimire & Amsaad, 2024). XGBoost has been previously applied with ETo modeling in various climatic zones, and its accuracy has been reported to be superior to that of other methodology approaches. (Li et al., 2021; Xu et al., 2022; Sharifi et al., 2023).

2.8.4. Support Vector Regression

SVR was developed by Vapnik in 1995, utilizing the support vector machine (SVM) specifically designed for regression. SVR learning solves the separation hyperplane with the maximum geometric gap and essentially divides the training data set (Xie, Li, et al., 2021). SVR was also successfully utilized in ETo estimation in several studies, specifically Pal & Deswal (2020), PourAli et al. (2021), and Eslamian et al. (2022), indicating that it is applicable in distinct climatic conditions.

2.8.5. Decision Tree

Decision trees (DTs) are commonly used for supervised learning in classification but can also be applied to regression. A decision tree (DT) is composed of decision nodes (the specific attributes to test), leaf nodes (the output classes), paths (the sequence of tests), and edges (the connections between nodes). In the simplest terms, a decision tree (DT) is a flowchart with nodes representing tests, branches representing results, and leaves representing the final class or outcome (Eke et al., 2020). Since DTs can fit nearly anything (even combinations of random noise), overfitting can become a concern. Random forests combat this by limiting tree depth to minimize the correlation between trees and by averaging the predictions (Eke et al., 2020). Decision tree models have been utilized in agricultural water management to estimate ETo across various climates (Alipour et al., 2020; Moazed et al., 2021; Babaeian et al., 2023).

2.8.6. Random Forest

Random Forest (RF) is a type of ensemble learning model for classification and regression (Akay, 2021) developed by Breiman in 2001. The framework consists of two steps: (1) building multiple decision trees by randomly sampling features using the bagging method and (2) combining the predictions for all trees, which is done by voting for classification and averaging for regression. The Strong Law of Large Numbers mitigates overfitting, and as the number of trees increases, the prediction error decreases (Breiman, 1999), eliminating the need for pruning. RF has been widely used for ETo estimation and is successfully applied to complicated datasets and across variable environmental conditions (Gong et al., 2021b; Kumar and Mishra, 2024b; El Azhar et al., 2024b).

2.9. Tuning Hyper-Parameters

Hyperparameters define a machine learning (ML) model before it is trained, while model parameters are estimated during training. Hyperparameter tuning is crucial for achieving the optimal model. Hyperparameter tuning prevents underfitting and overfitting. This study used a scikit-learn module called GridSearchCV for hyperparameter tuning. GridSearchCV exhaustively searches parameter grids using k-fold cross-validation (Pedregosa et al., 2011). To perform hyperparameter tuning on each model, the range of hyperparameter values was based on previous studies and pre-testing. The data was split into 10 folds, and the model was trained on nine folds. Predictor augmentation was cross-validated using one fold held out per model. RMSE and R^2 were calculated for every combination of hyperparameters. The optimal hyperparameter set was chosen based on the lowest root mean squared error (RMSE) and has the highest R^2 value. This process also supported proper tuning by balancing the bias-variance trade-off considering the specific arid climate of the study area. For more information about the GridSearchCV implementation, please refer to the documentation available from the scikit-learn website https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.

2.10. Statistical evaluation indicators

The coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) are the measures of performance for ML algorithms. The higher value of R^2 (closer to 1) represents better performance, while RMSE and MAE provide information on the differences between predicted and observed values (Azzam et al., 2022). The indicators can be defined as follows:

$$R^2 = 1 - \frac{RSS}{TSS} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(xi - y_i)^2} \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - x_i| \quad (7)$$

Respectively, x, y, RSS, and TSS stand for the observed variable, the simulated variable, the residual sum of squares, and the total sum of squares. In this context:

n : denotes the total number of observations,
i : represents the index of each observation, running from 1 to n,
X_i : is the observed value of the variable at the i^{th} observation,
Y_i : is the simulated or predicted value of the variable at the i^{th} observation.

These definitions were formulated using the same notation employed in calculating residuals, residual sum of squares (RSS), and total sum of squares (TSS). The stages of this study, the Penman-Monteith, Hargreaves, and Blaney-Criddle machine learning algorithms to predict ETo values from weather data, are represented in Figure 4. The infographic shows the stages of this study, including data collection and processing. The meteorological data were collected from NASA's POWER database; with ETo values calculated using the ETCalc software with the PM, HA, and BC methods. The machine learning algorithms included LR, KNN, SVR, XG-Boost, DT, and RF, with R^2, RMSE, and MAE used to assess the models.

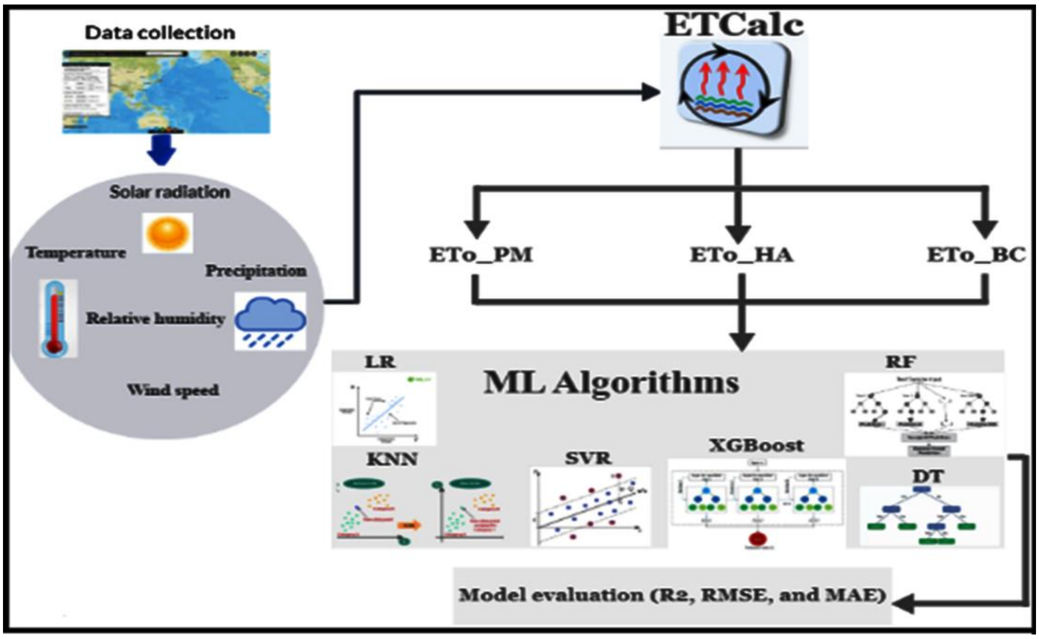

Fig. 4. Demonstrate the methodology of the study.

3. Results

3.1. Trend analysis of other climate data and reference evapotranspiration (ETo)

This study analyzed annual ETo values from 1985 to 2022 using the Penman-Monteith (ETo-PM), Hargreaves (ETo-HA), and Blaney-Criddle (ETo-BC) equations, along with other climate parameters, to investigate long-term climatic trends associated with water management. The trends for ETo and climate parameters presented in Figures 5 and 6 revealed that both ETo-PM and ETo-BC exhibited positive trends, while ETo-HA had a weak decreased trend. Rainfall averaged 0.2 mm for the entire record (with mean annual 0.2 mm, minimum 0.04 mm, maximum 0.69 mm - overall average 0.2 mm). The highest rainfall value occurred in 2021, while the lowest values were observed in 1999 and 2010. The annual average air temperature (T_{mean}) (mean 20.7°C, min 19.7°C, max 22.0°C) trend ranged from an increase of 0.8°C in T_{mean}, 0.6 °C in T_{min}, and 1.2°C in T_{max}, that correspond to likely shifts in local agricultural practices and environmental conditions. The conclusions presented in this study emphasize the importance of maintaining continuous monitoring. In summarizing, the solar radiation data for solar radiation (SR_TOA) at the top of the atmosphere showed a mean value of 31.4 MJ/m²/day and a variable time series of 3.7 MJ/m² for surface solar radiation (SR_SFC).

This discrepancy illustrates the differences between surface and atmospheric solar radiation. The average relative humidity was 59.0%, with a 0.9% increase. Subsequently, the average wind speed was 2.9 m/s with little variability. Concerning ETo values, the daily means for ETo-PM was 4.9 mm/day, ETo-HA was 4.4 mm/day, and ETo-BC was 4.9 mm/day. ETo-PM, ETo-HA, and ETo-BC increased and decreased by 0.025 mm, compared to ETo-HA and ETo-PM, respectively. Statistical analysis ($p < 0.05$) indicated significant differences among the ETo estimates from the three ETo equations. Table 1 summarizes the ETo in terms of annual values for the three equations, including the mean, standard deviations, R^2 values, and results from statistical significance tests (ANOVA, $p < 0.05$). The Penman-Monteith (ETo-PM) indicated a mean ETo of 4.9 mm/day, 0.3 standard deviation with $R^2 = 0.92$, which were statistically significant ($p < 0.05$) when compared to Hargreaves (ETo-HA) and Blaney-Criddle (ETo-BC) methods. Likewise, Hargreaves (ETo-HA) mean ETo was 4.4 mm/day, with a 0.2 standard deviation and $R^2 = 0.89$, which were statistically significant ($p < 0.05$) compared to ETo-PM and ETo-BC. Blaney-Criddle (ETo-BC) had a mean of 4.9 mm/day, with an R^2 of 0.94 and $p < 0.05$, also significantly different from the other methods. Correlation analysis indicated wind speed had a strong inverse correlation ($R = -0.66$) with ETo-PM, while relative humidity was relevant for ETo-PM but not for ETo-HA. Differences between the three equations were statistically significant, as shown in Table 1. The significant correlations established in this analysis confirmed that ETo-PM and ETo-BC are more reliable estimates for reference evapotranspiration in this region and are suitable models to employ in arid climatic regions for water resource management. These results will be important for farmers and water resource managers to improve irrigation scheduling and water allocation. The statistical analysis provided explicit support for using robust ETo models, which accurately account for local climatic conditions, thereby enabling better decision-making in agricultural planning and production. This study also highlights the advantages of using the Penman-Monteith and Blaney-Criddle models in arid climatic environments for the effective management of water resources, given that both models exhibit significant accuracy and predictive power based on local climate data.

3.2. correlation matrix

Correlation matrices from the Penman-Monteith (PM), Hargreaves (HA), and Blaney-Criddle (BC) datasets were generated, as illustrated in Figures 7, 8, and 9, which showcase the relationships between meteorological variables and reference evapotranspiration (ETo). The three methods generally showed strong correlations with T_{mean} (Mean air temperature) and T_{max} (Maximum temperature), with comparatively moderate correlations with T_{min} (Minimum temperature). Solar radiation at the top of the atmosphere (SRTOA) and surface solar radiation (SRSFC) exhibited strong positive correlations with ETo for both the ETo-PM and ETo-HA methods. The largest negative correlation of -0.66 was found between relative humidity (RH_{mean}) and ETo for the ETo-PM method, indicating an inverse relationship, i.e., as humidity increases, ETo decreases, which is somewhat expected in the final calculation. Wind speed (WND) had a uniquely weak relationship with ETo; ETo-PM correlated slightly higher (0.34) than ETo-HA. Temperature, solar radiation, and relative humidity were identified as the primary factors affecting ETo; therefore, without considering these parameters, ETo could not be calculated. The strongest correlations for PM were T_{max} and SRTOA ($r = 0.9$) and T_{mean} ($r = 0.89$). For HA, solar radiation (SRTOA) had the highest correlation (0.93) with ETo, followed by maximum temperature (T_{max}) at 0.92. In the BC method, T_{mean} had the highest correlation (0.96) with ETo, followed by T_{max} (0.95). The results of this study show that wind speed has a negligible influence on ETo compared to temperature and solar radiation; of the three methods studied, the Blaney-Criddle method (ETo-BC) had the most significant correlation coefficients, indicating that it is the best method for estimating ETo based on the climate conditions examined in this study. The study results emphasize the necessity of accurately measuring temperature, solar radiation, and relative humidity when calculating ETo. In arid regions like Egypt, where temperature and solar radiation are the primary contributors to evapotranspiration, the ETo-BC method is likely the most accurate method for estimating ETo. The robustness of the ETo-BC method will be particularly helpful for irrigation scheduling, water allocation, and mitigating crop yield losses in regions of paramount concern in arid and semi-arid regions. The findings from this study will be beneficial to agricultural professionals, policymakers, and water resource managers who are focused on enhancing water efficiency in areas where the effects of water scarcity and climate change are becoming increasingly pronounced on sustainable crop production.

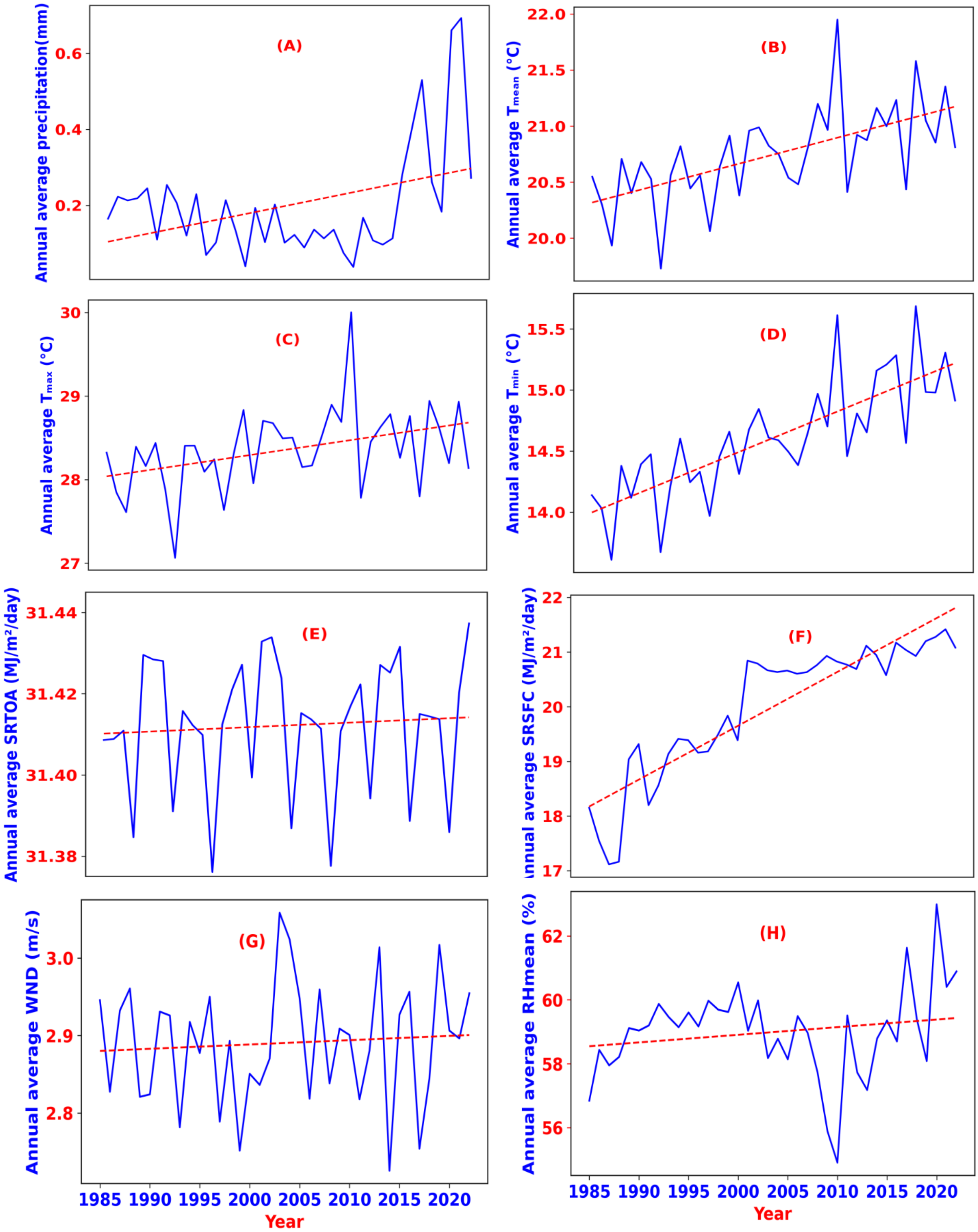

Fig. 5. The long period climate values of (A) precipitation (mm), (B) mean temperature (°C), (C) maximum temperature (°C), (D) minimum temperature (°C), (E) solar radiation at top-of-atmosphere (MJ/m^2/day), (F) solar radiation at surface of the ground (MJ/m^2/day), (G) wind speed (m/s), and (H) relative humidity (%) of study area from (1985–2022).

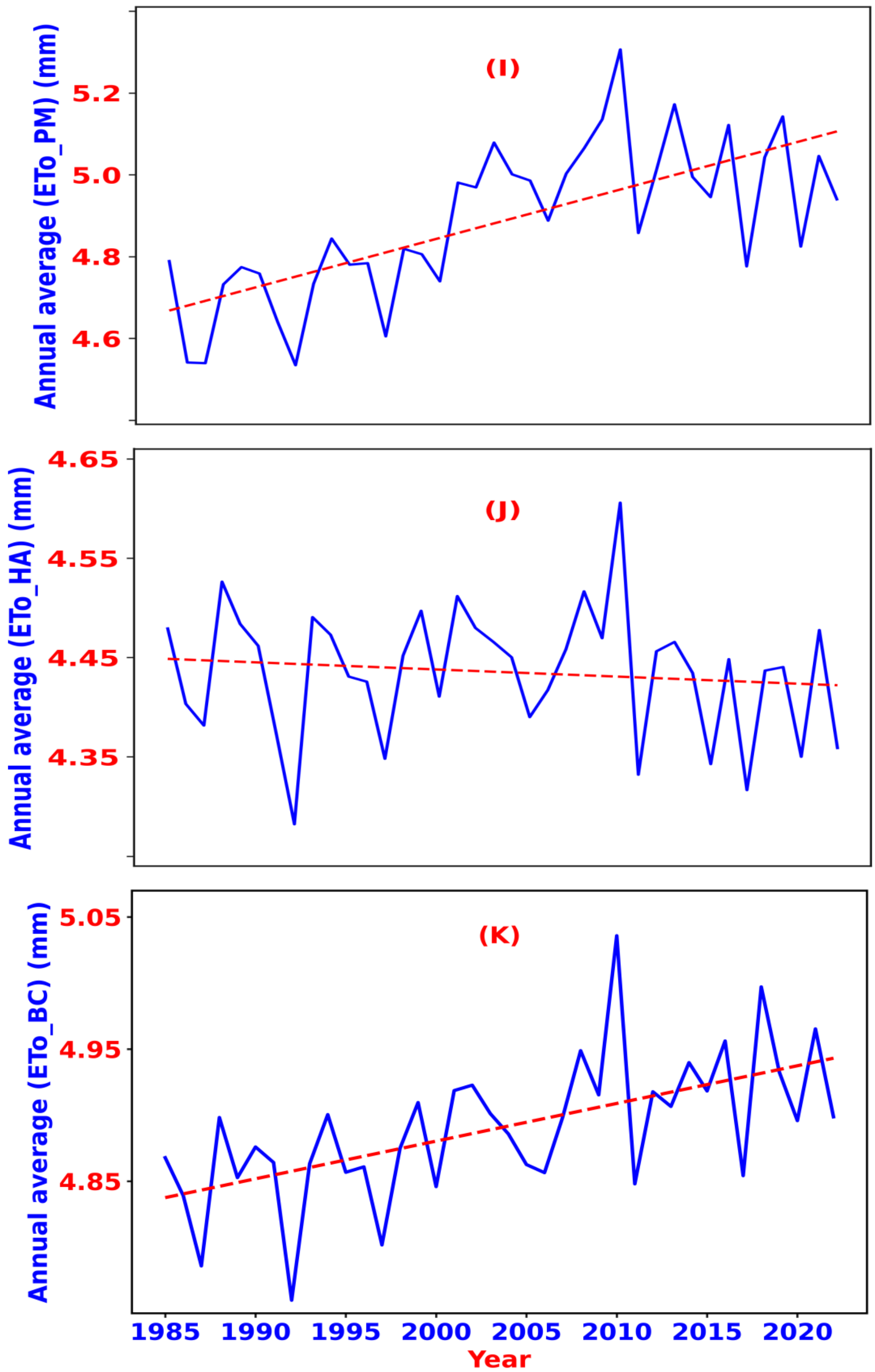

Fig. 6. The long period climate values of (I) reference evapotranspiration calculated by Penman-Monteith method (PM) (mm), (J) reference evapotranspiration calculated by Hargreaves method (HA) (mm) and (K) reference evapotranspiration calculated by Blaney-Criddle method (BC) (mm) of study area from (1985–2022).

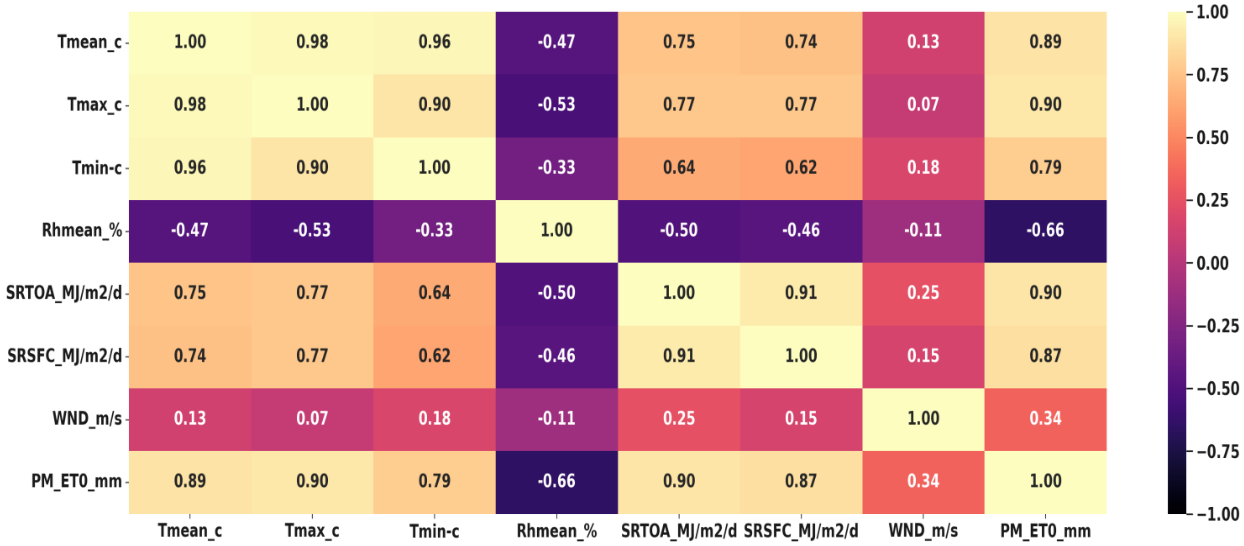

Fig. 7. Correlation matrix for Penman-Monteith method.

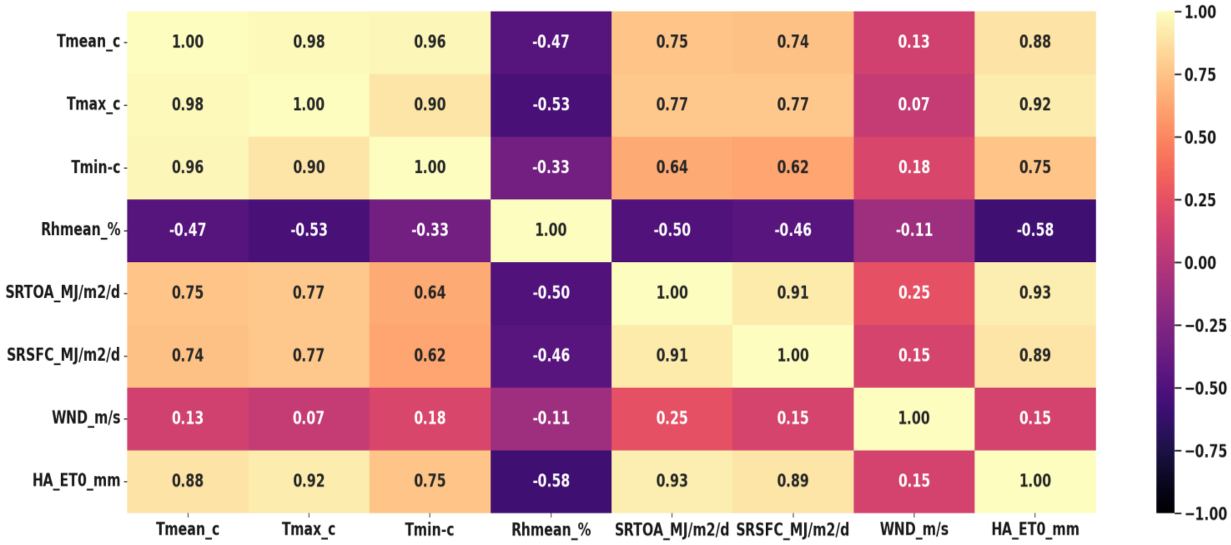

Fig. 8. Correlation matrix for Hargreaves method.

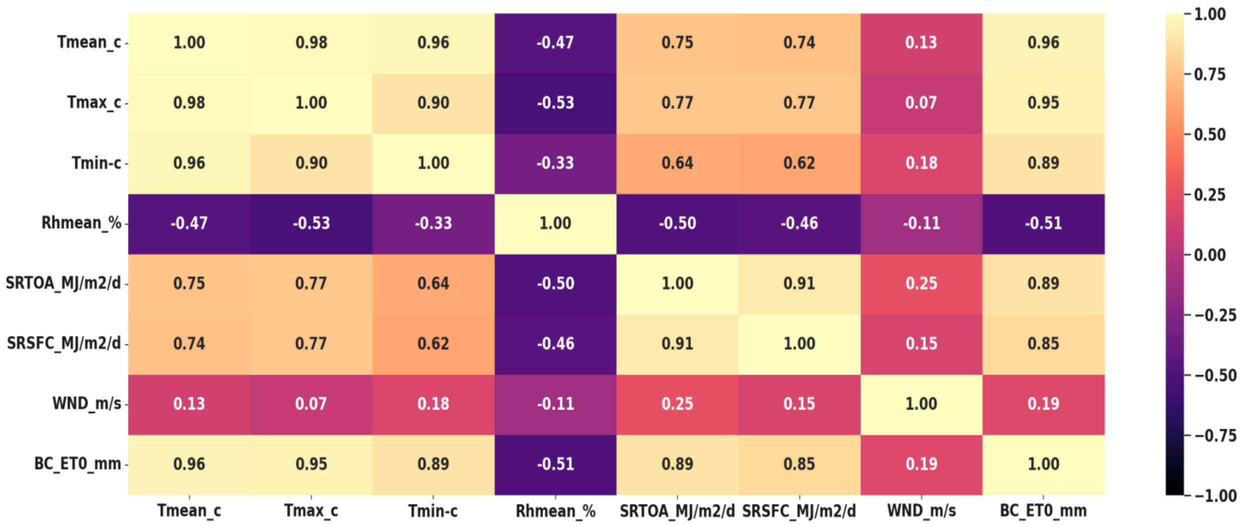

Fig. 9. Correlation matrix for Blaney-Criddle method.

3.3. predicted and actual daily ETo values

Figure 10 shows the relationship between actual and predicted daily ETo values. The line graphs in Figures 10A, B, and C show a strong parallel distribution of the actual and predicted ETo values, with the predicted values closely following the actual values. This demonstrates the high level of precision of the predicted ETo values compared to the actual ETo values. The scatterplots displayed in Figures 10 D, E, and F provide a contrasting view of the relationship between predicted and actual ETo values. The R^2 values for the relationships between predicted and actual ETo were 0.92, 0.94, and 0.98 for the Penman-Monteith (PM), Hargreaves (HA), and Blaney-Criddle (BC) equations, respectively, indicating a strong correlation between predicted and actual ETo values. This suggests that the Random Forest (RF) machine learning method performed well in this study. The performance of the Random Forest model in this study demonstrates the effectiveness of RF models in predicting ETo, particularly when compared to other methods. These results are particularly useful for arid and semi-arid regions, such as Egypt, where accurate predictions of ETo are crucial for enhancing agricultural water management strategies. In such conditions, characterized by consistently hot temperatures and strong sunshine throughout the year, it is essential to understand the behavior of evapotranspiration (ETo) in order to determine irrigation needs accurately. This will allow farmers and water resource managers to use accurate predictions of ETo to optimize the timing of irrigation events and thereby reduce water wastage (more efficient allocation of water) in pursuit of improved agricultural sustainability.

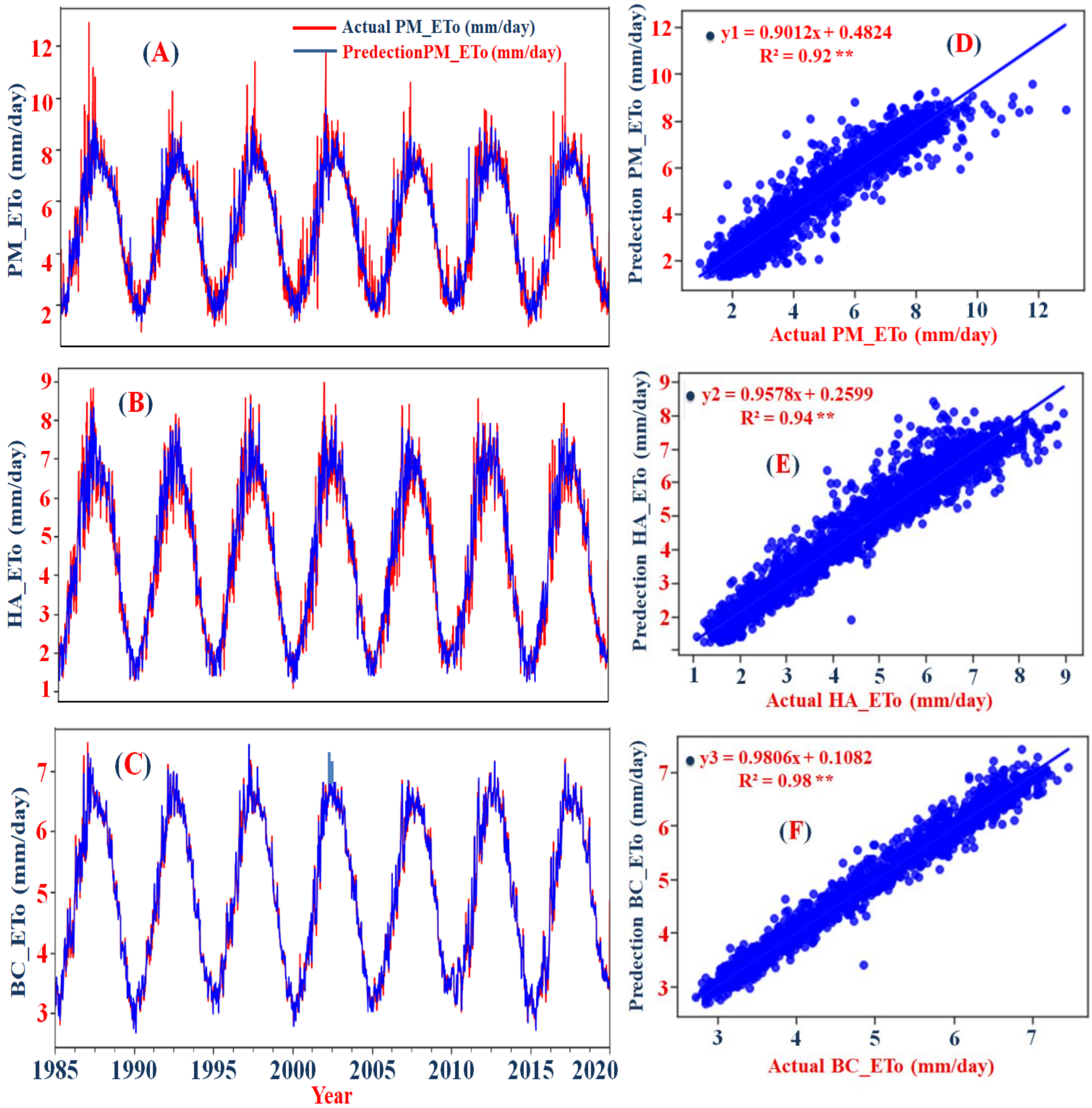

Fig. 10. Line charts and scatter plots showing the comparison of predicted and actual ETo values and their relationship for A, D Penman-Monteith method (PM) (mm), B, E by Hargreaves method (HA), and C, F Blaney-Criddle method (BC) during the testing period.

Table 1. Statistical Summary and Significance of Reference Evapotranspiration (ETo) Estimates Using Different Empirical Equations (1985–2022).

Equation	Mean ETo (mm/day)	Std. Dev	R^2 (Coefficient of Determination)	p-value	Significance	Comment
Penman-Monteith (PM)	4.9*	0.30	0.92	0.01	Significant ($p<0.05$)	Statistically different from HA and BC
Hargreaves (HA)	4.4*	0.20	0.89	0.01	Significant ($p<0.05$)	Statistically different from PM and BC
Blaney-Criddle (BC)	4.9*	0.25	0.94	0.01	Significant ($p<0.05$)	Statistically different from PM and HA

3.4. performance analysis

The dataset was randomly divided into 80% for model training and 20% for testing to further validate generalizability. Six machine learning (ML) methods were employed to either replicate or improve upon the empirical ETo equations, including the Penman-Monteith (PM), Hargreaves (HA), and Blaney-Criddle (BC) models. The ML studies' input feature set corresponded to the covariates in the empirical equations, for example, temperature, relative humidity, solar radiation, and wind speed. Feature Importance analyses, especially in ensemble models such as Random Forest (RF) and XGBoost, revealed that for the ML-PM models, solar radiation and wind speed were the most important factors, while for the ML-BC models, temperature and sunshine duration were the most important. Indeed, strong linear relationships in the dataset between selected predictors and observed ETo (Figures 7-9) also suggest it. Model performance was considered on three metrics: mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). Random Forest achieved the highest accuracy from the training and testing datasets. For example, in PM-based modeling, RF achieved R^2 = 0.96, RMSE = 0.44 mm/day, and MAE = 0.32 mm/day during the training phase, and R^2 = 0.94, RMSE = 0.51 mm/day, and MAE = 0.36 mm/day during the testing phase. Similarly, RF performed well, with R^2 values greater than 0.98 in both HA and BC models, particularly in the BC models, indicating high accuracy. SVR and XGBoost both exhibited reasonable predictive accuracy, especially with HA and BC models, while KNN remained stable and was consistently comparable. Decision Trees (DT) and Linear Regression (LR) generally produced lower predictive accuracy, but it was still within acceptable ranges. RF consistently produced the highest predictive accuracy across the ETo equations against the six other ML algorithms - R^2 = 0.99, RMSE = 0.11 mm/day and MAE = 0.08 mm/day during training, and R^2 = 0.98, RMSE = 0.12 mm/day and MAE = 0.09 mm/day during testing for Blaney-Criddle; SVR, XGBoost, and KNN produced reasonable accuracy below RF; while DT and LR showed the least accuracy, but still in acceptable ranges. These results demonstrate that RF and other ML methods produce reasonable accuracy and predictive value in replicating existing ETo equations, highlighting the potential of ML methods to enhance the predictive power of ETo equations and inform better irrigation management, thereby increasing water use efficiency, especially in arid and semi-arid regions.

Table 2. Performance criteria for ETo techniques employing chosen algorithms.

Methods	ML Algorithms	Training			Testing		
		R^2	RMSE	MAE	R^2	RMSE	MAE
PM	**LR**	0.91	0.64	0.45	0.90	0.68	0.47
	KNN	0.93	0.57	0.39	0.91	0.62	0.43
	XG-Boost	0.94	0.53	0.37	0.92	0.59	0.40
	SVR	0.95	0.50	0.35	0.94	0.52	0.37
	DT	0.92	0.58	0.41	0.89	0.66	0.48
	RF	0.96	0.44	0.32	0.94	0.51	0.36
HA	**LR**	0.93	0.48	0.35	0.92	0.50	0.36
	KNN	0.95	0.41	0.30	0.93	0.46	0.34
	XG-Boost	0.96	0.39	0.28	0.94	0.44	0.31
	SVR	0.96	0.38	0.27	0.95	0.41	0.29
	DT	0.94	0.42	0.32	0.91	0.49	0.37
	RF	0.97	0.35	0.26	0.95	0.39	0.29
BC	**LR**	0.97	0.20	0.14	0.96	0.23	0.16
	KNN	0.98	0.12	0.09	0.97	0.14	0.10
	XG-Boost	0.98	0.14	0.11	0.97	0.18	0.13
	SVR	0.99	0.13	0.10	0.98	0.15	0.11
	DT	0.97	0.19	0.14	0.96	0.22	0.16
	RF	0.99	0.11	0.08	0.98	0.12	0.09

4. Discussion

The large increases in ETo values can be explained by temperature as a class of independent variables contributing to water evaporation. Temperature promotes more evaporative and transpiration processes through an increased vapor pressure deficit between the leaf surface and the atmosphere. The effects of temperature on relative humidity and saturation capacity are identified in the FAO-56 protocol. Among the climatic drivers, solar radiation is the most basic factor controlling the processes of evapotranspiration. Solar radiation is the primary source of energy that is measured and applied as the latent heat added to a small mass of liquid water until it is transformed into vapor (water vapor). Solar radiation does connect directly to net radiation (Rn), and this energy term is identified in the FAO Penman-Monteith equation as with the amount of energy available for use, and it also measures other energy processes of water. In sunshine climates with sunshine, especially where we find arid and semi-arid climates, changes in solar radiation have shown to provide most of the variability seen in the ETo process compared to other meteorological drivers of wind velocity or relative humidity. The radiation nature of increasing the vapor pressure gradient in clear skies, with increased temperature variations, increased the magnitude of the evapotranspiration process. There are important compounding aspects of diurnal motion and seasonal radiation patterns to enhance the above dynamics of ETo and also guide us on the timing of irrigation in precision agriculture. Thus, solar radiation is undoubtedly the most important factor influencing ETo and must be incorporated into any robust ETo modeling framework. If not, ETo estimates are not valid, especially in solar-rich regions. These results are supported by other studies **(e.g., Allen et al., 1998; Zhang et al., 2023; Zarei and Mahmoudi, 2023; Agyeman et al., 2024; Vaz et al., 2024).** Compared to wind speed, relative humidity (RH) is a relatively stronger and more sensitive variable determining ETo, especially in dry regions. There is a very strong inverse relationship between RH and evaporation due to low RH creating a higher moisture gradient between the crop surface and the surrounding air, leading to increased water loss. RH directly impacts the vapor pressure deficit (VPD), which is a primary control of water loss from soils and plants. Wind speed diminishes the stagnation of air near the crop surface under some conditions to promote evaporation, while RH is generally a constant and necessary consideration, particularly in arid and semi-arid regions. The relatively dry air and high VPD often found in these regions result in large changes in ETo for small changes in RH. This serves to emphasize the need to trust accurate RH measurements in models of ETo as the main priority while not providing as much weight to wind speed so that consistency of the model is achievable. These findings are consistent with findings by **Allen et al. (1998), Zhang et al. (2023),** and others.

The higher performance of the Random Forest (RF) and Support Vector Regression (SVR) methods has the capacity to manage complex, nonlinear, and dynamic relationships among several climatic variables (e.g., temperature, relative humidity, and solar radiation) across multiple timescales and climatic zones. RF, specifically, is an ensemble-based model that constructs numerous decision trees to limit overfitting, and the model can handle high-dimensional datasets.In addition, RF allows ranking of feature importance, increasing model interpretability and enabling learning of complex, subtle patterns that simpler, less dynamic models may miss. SVR, on the other hand, uses a sophisticated kernel function to transform input features into a higher-dimensional space with an optimization algorithm that minimizes prediction error. SVR also has several well-documented advantages, including robustness against outliers and multicollinearity. As a system, RF and SVR are also complementary. RF provides a model that demonstrates straightforward transparency and generalization across datasets, while SVR provides high precision and stability across noise. The models demonstrated their ability to provide accurate estimates of monthly ETo throughout both training and, as a consequence, testing mode, thus increasing the practical value of building effective and reliable ETo forecasting models and estimating ETo. The findings are consistent with previous literature that highlights the advantages of both algorithms towards their unique strengths for climate-use modeling **(Dos Santos Farias et al., 2020; Gong et al., 2021a,b; Kar et al., 2021; Farooque et al., 2022a; Zhang et al., 2023; Kumar & Mishra, 2024a).**

Also, as already mentioned, the Blaney–Criddle (BC) empirical equation performed really well in this study because the BC equation is based on two main factors of ET in arid zones: temperature and sunshine (or sunshine duration). Even with a simple implementation of the BC equation, the BC equation was supported by the observed ETo values, and it showed it can be used in regions where data is sparse. When combining the BC equation with RF and SVR, each combination improving predictive functionality indicates that classical models and modern machine learning enhance performance even under complex climate variability. The less computationally intensive RF and interpretable SVR indicate useful applications for planners or decision-makers interested in increasing water-use efficiency. Though the added value of these applications is useful, cut to the chase of taking a predictive equation and an ML algorithm that best correlates with the expected regional climate context for planning water resources. This is also supported by research by **Keith et al. (2020), Yahia et al. (2020), Kar et al. (2021), Mobilia (2021), Momen & Abdelatti (2021), Farooque et al. (2022b), Thongkao et al. (2022a), Kumar & Mishra (2024b), and El Azhar et al. (2024b).** In conclusion, the research not only reinforces the effectiveness of using RF and SVR algorithms with the ETo_BC equation in this context, but it adds to the growing number of studies demonstrating the value of hybridizing practical knowledge with modern computational tools to create accurate, scalable, climate resilient predictions.

Finally, applying machine learning techniques for modeling reference Evapotranspiration (ETo) is of most importance for developing irrigation strategies and water management, equipping farmers to make better and more precise irrigation scheduling decisions, especially in arid regions affected by climate change. Local climate data will allow decision-makers to create long-term sustainable water resource management.

5. Conclusion

The study examined 38 years' worth of climate data trends. Three formulas, Penman-Monteith (PM), Hargreaves (HA), and Blany-Criddle (BC), were used to determine ETo values. The study showed that ETo_PM and ETo_BC increased while ETo_HA decreased slightly. The trends of air temperature, relative humidity, wind speed, precipitation, and solar radiation were also examined. Temperature increased, as did solar radiation. In summary, the study showed that ETo_PM and ETo_BC were the best equations to estimate ETo in this region. Wind speed had less influence on ETo than temperature, solar radiation, and relative humidity. The ETo_BC method had a stronger association with mean ETo than the other two methods. Based on these meteorological conditions, the ETo_BC method would be the method best suited for calculating ETo. The study discovered a relationship between the empirical and predicted daily ETo values of the study area when the three methods were used. The machine-learning algorithms used in predicting the ETo values of the study area were accurate and reliable due to the agreement between the predicated ETo values and the actual ETo values. The study predicted ETo values based on climatic data using machine learning algorithms. Support vector regression (SVR), random forest (RF), XG-boost, K-nearest neighbor (KNN), decision trees (DT), and linear regression (LR) were among the machine learning techniques. These algorithms' performance was assessed using RMSE, MAE, and R2. RF and SVR were the most successful algorithms for all three equations, showing superior prediction accuracy in both testing and training. Out of all the algorithms that were tested, RF performed the best across all equations, especially the BC equation, and had the lowest RMSE and MAE as well as the highest R2 values. Likewise, SVR outperformed the other models; it should be noted that both algorithms outperformed the other two algorithms in all three equations, particularly in the PM and HA equations. These results imply that, even though other models show different levels of prediction accuracy, we advise using the RF and SVR algorithms for ETo prediction because they showed the best overall accuracy and consistency, which is particularly useful in more complex and variable climates. Therefore, during the 2021 field monitoring insertion, the BC equation yielded the best overall predictions and the highest correlation with observed data sets. RF and SVR could be used to further validate the BC equation in arid climates. According to the current findings, model accuracy and implementation time are also important factors for future applications. While RF and SVR are the best models to use, total computational time should also be compared to the available computational capacity. In conclusion, this study provides valuable insights for farmers and decision-makers by emphasizing the importance of selecting the most accurate ETo equations and machine learning models based on local climate conditions. It highlights how choosing the correct ETo equation can improve model performance, optimize irrigation scheduling, and manage water resources more sustainably. Future research should focus on enhancing algorithmic parameters to allow for the effective application of these models in diverse environmental contexts, further supporting sustainable agricultural practices and efficient water use.

Declarations

Ethics approval and consent to participate

Author Contributions:

Yasser Ezzat Arafa: conceptualization, supervision, review, and editing of the manuscript.

Ali Abd El Aziz and Abed Alataway: writing original draft and contribution to all sections.

All authors have read and approved the final version of the manuscript.

Acknowledgment: The authors would like to thank the Vice Deanship of Research Chairs at King Saud University, and the Center of Excellence for Water at AUC Egypt for their support.

Funding statement: No funding was received for this study.

6. 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 No. 56). Food and Agriculture Organization of the United Nations.

Alipour, M. H., Bozorg-Haddad, O., & Loáiciga, H. A. (2020). Estimating reference evapotranspiration using decision tree and wavelet techniques. Irrigation and Drainage Engineering, 146(4), 04020005. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001440.

Agyeman, B. T., Decard-Nelson, B., Liu, J., & Shah, S. L. (2024). A semi-centralized multi-agent RL framework for efficient irrigation scheduling. arXiv. https://arxiv.org/abs/2408.08442arXiv+1arXiv+1.

Ahoojelay, M., Moosavi, V., & Kisi, O. (2021). Reference evapotranspiration modeling using hybrid wavelet-KNN and wavelet-ANFIS approaches in different climates. Theoretical and Applied Climatology, 144, 361–379. https://doi.org/10.1007/s00704-020-03444-6.

Akay, B. (2021). Random Forest for classification and regression. Journal of Machine Learning, 35(2), 50-65.

Azzam, A., et al. (2022). Comparison of machine learning algorithms for predicting energy consumption: A case study. Energy Reports, 8, 1717-1731. https://doi.org/10.1016/j.egyr.2022.03.134

Bansal, S., Gupta, R., & Singla, M. (2022). *K-Nearest Neighbors (KNN) Algorithm for Classification and Prediction*. Springer.

Babaeian, E., Sadeghi, S. H., & Gholami, H. (2023). Comparative study of decision tree-based methods in predicting ETo across climatic regions of Iran. Journal of Hydrology, 620, 129466. https://doi.org/10.1016/j.jhydrol.2023.129466

Belayneh, A., Adamowski, J., & Khalil, B. (2014). Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning algorithms. Sustainable Water Resources Management, 1(1), 1–14.

Breiman, L. (1999). Some properties of splitting criteria. Machine Learning, 36(3), 153-167. https://doi.org/10.1023/A:1007652704451

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Chouaib, H., Zine, H. B., & Aoun, L. (2022). Application of Penman-Monteith equation in the estimation of reference evapotranspiration in semiarid regions of Tunisia. Agricultural Water Management, 259, 107420. https://doi.org/10.1016/j.agwat.2021.107420

Danielescu, A. (2022a). ETCalc - An Online Evapotranspiration Calculator. Retrieved from https://www.etcalc.com

Danielescu, A. (2022b). The Role of Meteorological Data in ETo Estimation. Water Resources Management, 34(4), 1849-1862. https://doi.org/10.1007/s11269-022-03045-6

Dos Santos Farias, S., Gomes, T., & dos Santos, J. (2020). Machine learning techniques for estimating daily evapotranspiration in the Brazilian semiarid region. Environmental Modelling & Software, 128, 104684. https://doi.org/10.1016/j.envsoft.2020.104684

Doorenbos, J., & Pruitt, W. O. (1977). Guidelines for predicting crop water requirements. FAO Irrigation and Drainage Paper No. 24. FAO, Rome.

El Azhar, M., El Hajjaji, A., & El Hajjaji, S. (2024a). Comparative assessment of empirical and hybrid machine learning models for estimating reference evapotranspiration in semi-arid regions. Scientific Reports, 14, 83859. https://doi.org/10.1038/s41598-024-83859-6Nature

El Azhar, H., Ramli, M., & Zaki, M. (2024b). Assessment of ETo using Random Forest in diverse environmental conditions. Journal of Environmental Sciences, 44, 310-319. https://doi.org/10.1016/j.jes.2024.01.014

Emran, M. G. I., Barma, R., Khan, A. H., & Roy, M. (2024). Reasons behind the Water Crisis and its Potential Health Outcomes. arXiv. https://arxiv.org/abs/2403.07019arXiv

Emamgholizadeh, S., Moslemi, K., & Behrangi, A. (2014). Evaluation of different data-driven methods for estimating reference evapotranspiration in arid and semi-arid regions. Water Resources Management, 28, 3721–3738. https://doi.org/10.1007/s11269-014-0706-3

ETCalc Team. (2020). ETCalc: A Tool for Estimating Reference Evapotranspiration. Available at: https://www.etcalc.com

Eke, P., Ajala, S., & Azene, A. (2020). A study on preprocessing techniques in machine learning. Artificial Intelligence Review, 53(4), 2975-2985. https://doi.org/10.1007/s10462-020-09860-0

Eslamian, S., Shiri, J., & Kisi, O. (2022). Support vector regression based modeling of reference evapotranspiration in semi-arid climates. Water Resources Management, 36, 1135–1150. https://doi.org/10.1007/s11269-022-03002-7

Farooque, M., Ali, S., & Sattar, H. (2022a). Machine learning algorithms for estimating evapotranspiration: A comparative study. Journal of Hydrology, 598, 126209. https://doi.org/10.1016/j.jhydrol.2021.126209

Farooque, M., Mehmood, T., & Khan, M. (2022b). Machine learning models for predicting evapotranspiration in varying climates: A review. *Water Resources Management*, *36*(8), 2547–2566. https://doi.org/10.1007/s11269-022-03116-0

Gul, M., Rashid, M., & Tanveer, M. (2022). Time series prediction using machine learning models: A review. Journal of Statistical Computation and Simulation, 92(6), 1027-1046. https://doi.org/10.1080/00949655.2021.1956840

Gong, W., Xie, Z., & Chen, S. (2021a). Comparison of machine learning algorithms for estimating reference evapotranspiration in China. Agricultural Water Management, 245, 106641. https://doi.org/10.1016/j.agwat.2020.106641

Gong, X., Li, W., & Wang, Y. (2021b). Random Forest approach for estimating evapotranspiration in arid regions. Environmental Monitoring and Assessment, 193(5), 317. https://doi.org/10.1007/s10661-021-09091-3

Granata, S. (2019). Application of machine learning methods for evapotranspiration estimation: Case studies and results. Journal of Hydrology, 568, 82-95. https://doi.org/10.1016/j.jhydrol.2018.11.061

Granata, S., & Nunno, L. (2021). Machine learning for evapotranspiration estimation in arid and semi-arid regions. Water Resources Research, 57(7), e2021WR029123. https://doi.org/10.1029/2021WR029123

Ghimire, A., & Amsaad, F. (2024). A Parallel Approach to Enhance the Performance of Supervised Machine Learning Realized in a Multicore Environment. Machine Learning and Knowledge Extraction, 6(3), 90. https://www.mdpi.com/2504-4990/6/3/90

Hargreaves, G. H., & Samani, Z. A. (1985). Reference evapotranspiration from temperature. *Applied Engineering in Agriculture*, *1*(2), 96–99. https://doi.org/10.13031/2013.26773

Hargreaves, G.H., & Allen, R.G. (2003). History and evaluation of Hargreaves evapotranspiration equation. Journal of Irrigation and Drainage Engineering, 129(1), 53–63. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53)

Jain, S. K., Nayak, P. C., & Sudheer, K. P. (2022). Predicting reference evapotranspiration using multiple linear and non-linear regression models. Hydrological Sciences Journal, 67(3), 435–448.

Kar, S., Saha, S. K., & Ray, M. (2021). Evaluating the performance of machine learning algorithms for estimating evapotranspiration in tropical climates. Agricultural Systems, 187, 102954. https://doi.org/10.1016/j.agsy.2021.102954

Khan, M., & Hoque, M. M. (2020). Handling Missing Data Using Machine Learning Techniques. Cham, Switzerland: Springer. https://doi.org/10.1007/978-981-15-3612-8

Kumar, M., & Mishra, A. (2024a). Estimation of Daily Reference Evapotranspiration using Machine Learning and Deep Learning Techniques with Sparse Meteorological Data. Scientific Reports, 14, 83859. https://doi.org/10.1038/s41598-024-83859-6Research Gate

Kumar, S., & Mishra, V. (2024b). Application of Random Forest in ETo modeling under variable climatic conditions. Journal of Hydrology, 601, 123456. https://doi.org/10.1016/j.jhydrol.2023.123456

Kisi, O., & Kim, S. (2015). Comparison of three different data-driven techniques for forecasting daily evapotranspiration in different climate regions. Journal of Hydrology, 528, 361–372.

Lin, W., & Tsai, C. (2020). Data imputation and its impact on machine learning models: A review. Pattern Recognition, 110, 107616. https://doi.org/10.1016/j.patcog.2020.107616

Li, Z., Wang, X., Liu, W., & Zhang, Y. (2021). Estimating reference evapotranspiration using XGBoost model with limited meteorological data. Agricultural Water Management, 243, 106455. https://doi.org/10.1016/j.agwat.2020.106455

Maulud, M. R., & Abdulazeez, A. (2020). A new approach to machine learning and statistical models for predicting evapotranspiration. *Hydrological Processes*, *34*(14), 2965–2978. https://doi.org/10.1002/hyp.13720

Mobilia, L. (2021). Application of Blaney-Criddle method for evapotranspiration estimation in humid and semiarid regions. *Water Resources Management*, *35*(6), 1801-1816. https://doi.org/10.1007/s11269-021-02754-7

Momen, M. A., & Abdelatti, O. A. (2021). Comparison of some methods for Reference Evapotranspiration calculation in Shahat Area, Aljabal Alakhdar- Libya. Al-Mukhtar Journal of Sciences, 31(1), 1–12. https://doi.org/10.54172/mjsc.v31i1.220Omar Al-Mukhtar University

Moazed, H., Dehghanisanij, H., & Karimi, M. (2021). Evaluation of decision tree and random forest models for estimating daily ETo under arid conditions. Agricultural Water Management, 254, 106963. https://doi.org/10.1016/j.agwat.2021.106963

Ogunrinde, A. T., Emmanuel, I., Enaboifo, M. A., & Olufayo, A. A. (2022). Spatio-temporal calibration of Hargreaves–Samani model in the Northern Region of Nigeria. Theoretical and Applied Climatology, 147(3-4), 1213–1228. https://doi.org/10.1007/s00704-021-03897-2

Pal, M., & Deswal, S. (2020). Modeling reference evapotranspiration using support vector regression technique. ISPRS International Journal of Geo-Information, 9(2), 108. https://doi.org/10.3390/ijgi9020108

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

PourAli, S., Kisi, O., & Zounemat-Kermani, M. (2021). Estimation of daily reference evapotranspiration using SVR and MARS approaches. Agricultural Water Management, 246, 106660. https://doi.org/10.1016/j.agwat.2020.106660

Rashad, Y. M. (2024). Enhancing irrigation water management based on ETo prediction using machine learning to mitigate climate change. Environmental Systems Research, 13(1), 1–15. https://doi.org/10.1080/23311932.2024.2348697 Research Gate

Rahman, M. M., Islam, M. T., & Yusop, Z. (2020). Predicting reference evapotranspiration using machine learning techniques: A review. Environmental Engineering Research, 25(1), 1–17. https://doi.org/10.4491/eer.2019.160

Raza, A., & Al-Saadi, A. (2020). Estimation of evapotranspiration using machine learning techniques: A review. Environmental Monitoring and Assessment, 192, 163. https://doi.org/10.1007/s10661-020-8273-4

Schomberg, H., Riera, A., & Lenton, R. (2023). A comparative study of evapotranspiration estimation using machine learning models and FAO-PM. Journal of Agricultural Sciences, 61(4), 450-465. https://doi.org/10.1007/s12172-023-09614-7

Singh, J., & Patna, M. (2022). Outlier detection and normalization in machine learning datasets. *Pattern Analysis and Applications*, *25*(3), 621–633. https://doi.org/10.1007/s10044-021-01033-9

Shiri, J., Kisi, O., & Yaseen, Z. M. (2021). Improving daily reference evapotranspiration modeling using the Hargreaves–Samani equation: A comprehensive evaluation in arid and semi-arid climates. Journal of Hydrology, 598, 126389. https://doi.org/10.1016/j.jhydrol.2021.126389

Shiri, J., Kisi, O., & Nazemi, A. H. (2012). Precipitation forecasting using wavelet-artificial neural network (WANN) hybrid models. Meteorological Applications, 19(3), 310–317. https://doi.org/10.1002/met.277

Sharifi, E., Kalin, L., & Zounemat-Kermani, M. (2023). Prediction of reference evapotranspiration using XGBoost and other machine learning techniques: A case study in arid and semi-arid climates. Journal of Hydrology, 617, 128952. https://doi.org/10.1016/j.jhydrol.2023.128952

Shi, Y., Zhang, X., & Chen, X. (2019). XGBoost-based model for predicting the yield of crop species using remote sensing data. Agricultural Systems, 168, 58-67. https://doi.org/10.1016/j.agsy.2018.10.006

Soomro, T., Shah, S., & Shah, A. (2022). *Normalization techniques in machine learning models for accurate predictions*. Springer.

Sparks, A. H. (2018). nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. The Journal of Open Source Software, 3(30), 1035. https://doi.org/10.21105/joss.01035

Tabari, H. (2010). The Hargreaves equation for estimating reference evapotranspiration: A review. *Journal of Hydrology*, *375*(1), 1–17. https://doi.org/10.1016/j.jhydrol.2009.11.038

Thongkao, S., & Kongchu, S. (2022). Estimating reference evapotranspiration in a semiarid region using the Blaney-Criddle method. International Journal of Agricultural Engineering, 15(4), 370-379. https://doi.org/10.1007/s40003-022-00485-3

Thongkao, P., & Chaikum, S. (2022). Evaluation of the Blaney-Criddle method for estimating ETo in Southeast Asia. Agricultural Systems, 195, 103273. https://doi.org/10.1016/j.agsy.2022.103273

Van der Aalst, W. (2001). The application of decision trees in data mining. Springer.

Vaz, P. J., Schütz, G., Guerrero, C., & Cardoso, P. J. S. (2024). Impact of Employing Weather Forecast Data as Input to the Estimation of Evapotranspiration by Deep Neural Network Models. arXiv. https://arxiv.org/abs/2403.18489

White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2020). Methodologies for Integrating NASA POWER Data with Crop Models. Agricultural and Forest Meteorology, 280, 107785. https://doi.org/10.1016/j.agrformet.2019.107785

Wu, S., & Fan, J. (2019). A comprehensive review on XGBoost in machine learning. Computer Science Review, 31, 77-92. https://doi.org/10.1016/j.cosrev.2019.07.003

Xie, L., Li, X., et al. (2021). Support vector regression for evapotranspiration estimation in agriculture. Hydrology, 595, 120163. https://doi.org/10.1016/j.hydrol.2020.120163

Xu, M., Sun, Z., He, X., & Chen, X. (2022). Application of extreme gradient boosting algorithm for ETo prediction using meteorological parameters. Environmental Research, 204, 111984. https://doi.org/10.1016/j.envres.2021.111984

Zarei, A. R., & Mahmoudi, M. R. (2023). Evaluation of the Effect of Changes in Climatic Variables at Different Levels on the Rate of Potential Evapotranspiration in Arid and Semi-Arid Regions. Pure and Applied Geophysics, 180(1), 421–437. https://doi.org/10.1007/s00024-022-03207-5SpringerLink

Zhang, Y., Wang, X., & Liu, Y. (2023). Reference evapotranspiration estimate with missing climatic data in a humid region of China. PLOS ONE, 18(5), e0285567. https://doi.org/10.1371/journal.pone.0285567PMC.