Predictive Modeling and Comparative Analysis of Reference Evapotranspiration with Machine Learning Algorithms

Abstract

Accurate estimation of reference evapotranspiration (ET_0) is crucial for a multitude of applications, encompassing drought detection, irrigation scheduling, water resource management, and disaster risk reduction. This investigation utilized the FAO-PM equation for ET_0 estimation and subsequently incorporated meteorological variables as input variables with machine learning (ML) models to enhance ET_0 predictions. The dataset was bifurcated into training and testing data segments. Four distinct machine learning models were deployed in this study, namely Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), and Linear Regression (LR). The performance of these models was evaluated using various statistical indices, including Mean Absolute Error (MAE), Mean Sum of Error (MSE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2), to pinpoint the most efficacious ML algorithm. After conducting a comprehensive analysis involving both training and testing data, the results unequivocally identify GBM with MAE values of 0.054 and 0.077, MSE values of 0.005 and 0.011, MAPE values of 0.014 and 0.022, RMSE values of 0.072 and 0.107, and an R2 value of 0.096 and 0.092 during training and testing, respectively. This model has been selected as the optimal choice for precise ET0 estimation within the study region. Subsequently, SVM, RF, and LR follow as alternatives in terms of performance, in descending order.

Keywords:

Evapotranspiration, Machine Learning, RandomForest, Support Vector Machine, Gradient Boosting Trees, Linear Regression

Introduction

The Earth's limited natural resources face an escalating risk of depletion, primarily due to the concurrent factors of rapid population growth, extensive industrial development, and the profound impacts of climate change. India is expected to become the world's most populous

nation by 2023, surpassing China, which will need to provide sustenance for approximately 1.66 billion people by 2050 (UN, 2022). Among these resources, water emerges as a critical factor, especially in its vital role as the lifeblood of agriculture through irrigation, a fundamental component of food security for countries that make significant contributions to Gross Domestic Product (GDP) and employment. According to the Ministry of Jal Shakti, Government of India, the average annual per capita water availability stood at 1816 cubic meters in 2001, 1545 cubic meters in 2011, and 1487 cubic meters in 2021. It is projected to further decline to 1367 cubic meters by 2031(PIB, 2020).The resource crisis intensified by climate change is exemplified by the disruption of precipitation patterns, prolonged droughts, and heightened evapotranspiration rates, all of which amplify the competition for finite water resources. In response to this impending water scarcity crisis, there is an urgent need for innovative strategies to ensure the sustainability and efficacy of irrigation methods, with the goal of safeguarding food production systems and economic stability. Given this pressing scenario, the necessity for effective and sustainable irrigation water management has never been more evident.In this context, emerging technologies are poised to play a central role in addressing these challenges. These technologies encompass a diverse range of approaches, including precision agriculture, remote sensing, data analytics, and advanced control systems, collectively providing solutions to optimize irrigation practices, enhance water use efficiency, and minimize wastage.

Evapotranspiration (ET), a pivotal component of the hydrological cycle, is a non-linear and intricate phenomenon influenced by a myriad of factors encompassing micrometeorological variables, soil attributes, crop characteristics, and agricultural management practices (Kumari et al., 2022). Reference evapotranspiration (ET_0) stands as the most widely utilized parameter for calculating crop water requirements, devising irrigation schedules, maintaining hydrological water balances, simulating crop yields, and designing irrigation systems. ET_0 signifies the volume of water that escapes from a continuous expanse of vegetation under specific climatic conditions when water availability is not a limiting factor (Gangopadhyay et al., 1966). Consequently, the precise estimation of ET_0 is of paramount importance, particularly in regions plagued by water scarcity. Several methods can be employed to compute ET_0, including micro-weather techniques based on energy balance and vapor/mass flux transfer, empirical methodologies, lysimetric measurements, and soil moisture balance approaches.

Nevertheless, recent advancements in data-driven techniques, such as machine learning (ML), have proven to be more accurate when compared to empirical models, especially under varying environmental conditions. ML models excel in simulating the intricate and non-linear nature of ET_0. The utilization of ML models for ET forecasting has gained significant traction in recent years (Feng et al., 2017; Chia et al., 2020b, 2021; Duhan et al., 2021; Rajput et al., 2022). Various ML algorithms are employed worldwide for ET_0 prediction (Rajput et al., 2023), including adaptive neuro-fuzzy neural networks (Keshtegar et al., 2018), least square-support vector machines (LS-SVM) (Reis et al., 2019), fuzzy logic (Malik & Kumar, 2015), multiple-layer perceptron neural networks (Seifi & Riahi, 2020), relevance vector machines (Bachour et al., 2016), multivariate regression splines (Mehdizadeh et al., 2017), and Least Square-Support Vector Regression (LS-SVM) (Dimple et al., 2023). Multiple studies have demonstrated that ML-based models provide more accurate ET_0 estimates compared to empirical methods like the Hargreaves-Samani method, Blaney-Criddle method, Thornthwaite method, Makkink method, and Penman method across various regions globally(Rajput et al., 2023).

Numerous field studies have demonstrated that labor-intensive and time-consuming field measurements can be effectively replaced by machine learning models possessing strong predictive capabilities, offering significant time and cost savings. These models excel in handling the intricate complexities inherent in Evapotranspiration (ET_0) calculations, including non-linear relationships and adaptability to changing environmental conditions. Nevertheless, despite the availability of various machine learning models, challenges persist in achieving accurate ET_0 estimations under specific climatic conditions. In such cases, local calibration and accuracy assessment become imperative. As a result, this investigation was carried out with the primary aim of evaluating and selecting the optimal machine learning model for estimating ET_0 in Thrissur district, Kerala.

2. Materials and Methods

2.1 Study Area

Thrissur district, located between latitudes 10°10’22’’ and 10°46’54’’ North, and longitudes 75°57’20’’ and 76°54’23’’ East, experiences a wet climate with distinct seasons. These include a hot summer (March to May), southwest monsoon (June to September), northeast monsoon (October to December), and a cooler, pleasant period (January and February). The

district receives an average annual rainfall of 3198.133 mm, with the highest precipitation during the southwest monsoon (71.24%). The maximum temperature ranges from 29.3 to 36.20°C, while the minimum varies from 22.1 to 24.90°C (Figure 1). December to April is the warmest period, and November to February is the coldest. Sunshine hours are abundant from December to April, resulting in higher evaporation rates (up to 7.4 mm/day), whereas the monsoon months (June to October) see lower evaporation (minimum of 2.9 mm/day)(figure 2). The prevalent soil type is lateritic.

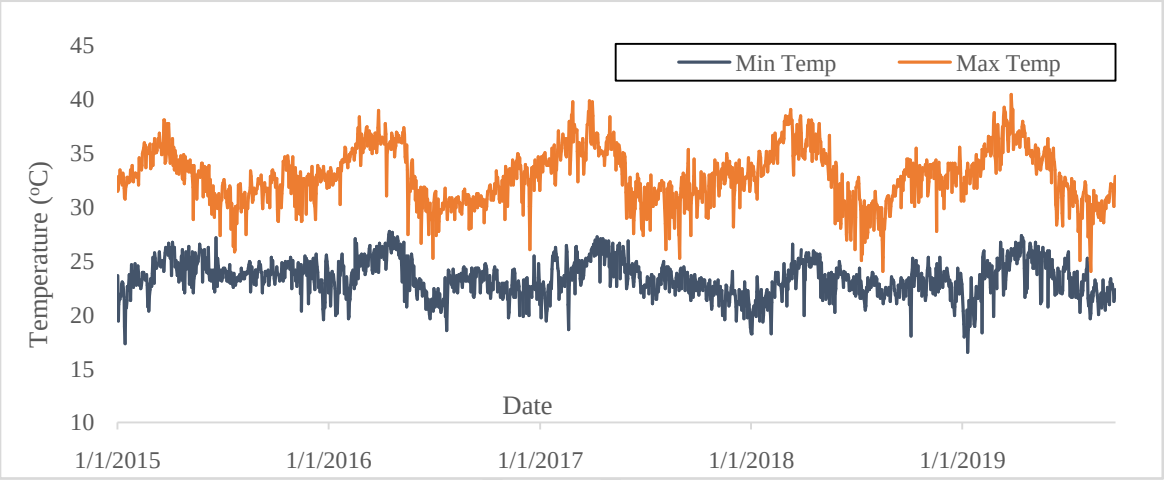

Fig 1. Graphical representation showing temperature changes scenario

2.2 Reference Evapotranspiration (ET_0)

The ET_0 calculations for the past four years (2015–2019) relied on daily micrometeorological data, encompassing parameters such as maximum air temperature (Tmax, °C), minimum air temperature (Tmin, °C), maximum relative humidity (RHmax, %), minimum relative humidity (RHmin, %), wind speed (m/s), and sunshine hours. The estimation of ET_0 was accomplished through the FAO-PM method-based ET calculator, utilizing all five variables as inputs. Equation 1 was applied for the ET_0 computation, and comprehensive guidelines for this estimation method were delineated by Allen et al. (1998). This method serves as the benchmark for comparing the performance of various machine learning approaches.

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_{a)}}{\Delta + \gamma(1 + 0.3u_2)} \quad \text{...1}$$

Where, Rn is net solar radiation (MJ m^{-2} day^{-1}), Rs is the solar radiation (MJ m^{-2} day^{-1}), λ is the latent heat of evaporation (MJ kg^{-1}),T is the daily mean temperature (°C), U_2 is the mean daily wind speed at 2meter height (m/s), e_s *and* e_a- Saturation and actual vapor pressure (kPa), G - soil heat flux (MJ m^{-2} day^{-1}), Δ-the slope of saturated water vapor pressure curve (kPa/c)

2.3Description of Machine learning model

ML models, viz.Random Forest, SVM, gradient bosting method, liner regression was developed for predicting ET_0 usingTmax, Tmin, RHmax, RHmin, wind speed and number of sunshine hours as predictors and PM-ET as a prediction in calibratedand validated models.

A) Random forest

The Random Forest (RF) algorithm, introduced by Breiman in 2001, is based on a model known as Classification and Regression Trees (CART). This algorithm encompasses both regression (RFR) and classification (RFC) methods and finds applications in a wide range of tasks, including regression, classification, and unsupervised learning.The fundamental idea behind the RF algorithm is rooted in statistical theory. It involves the repeated and random selection of K samples from the original training dataset of size N, creating a new set of training samples through a method called bootstrap resampling. Subsequently, K decision trees are generated, and a random forest is assembled based on these bootstrap sample sets (Mishra et al., 2020).For classification tasks, the predictions for new data are determined by tallying the votes obtained from the classification trees. In regression scenarios, the final prediction is derived by averaging the predictive values from all the decision trees. The operation of the Random Forest algorithm can be summarized in the following steps: (i)Multiple resampling of the original training data. (ii)The random selection of a subset of features for each resampling step. (iii)The estimation of a decision tree based on a resample and the selected features. (iv)Accumulation of multiple decision trees to create a Random Forest model.These steps collectively form the foundation of the RF algorithm.

B) Support Vector Machine (SVM)

It is a robust machine learning model utilized in both classification and regression tasks. The SVM algorithm, originally introduced by Vapnik in 1995, is a supervised machine learning model commonly used for pattern recognition and data analysis. It has found extensive

application in regression and forecasting within various domains, including agriculture, hydrology, meteorology, and environmental studies.The SVM model conducts regression estimation through a series of kernel functions, which effectively transform the original input data from a lower-dimensional space to a higher-dimensional feature space. SVM operates by converting the input vector into this feature space and establishing the connection with the output vector. Its primary objective is to identify a hyperplane that optimally separates the data into distinct categories while maximizing the margin.Initially SVR approach used for rainfall-runoffmodeling in hydrology.

C) Gradient Boosting Trees

GBM, also known as Gradient Boosting Decision Tree (GBDT) or Multiple Additive Regression Trees (MART), was introduced by Geigy et al. in 1975 and later by Jensen et al. in 2012. It serves as a robust machine learning algorithm that excels in practical applications. As defined by Blaney in 1892 and 1952, GBR is composed of three key components: a loss function, a weak learner, and an additive model. Whenever a decision tree performs as a weak learner then the resulting algorithm is called gradient-boosted trees.These elements work in harmony to optimize, make predictions, and progressively incorporate weak learners to minimize the loss function.The term 'boosting' refers to the iterative process and uses a gradient descent for the optimization. It is used to improve the accuracy of trees (Naganna et al.,2020). Gradient Boosting Machine (GBM) is built by many decision trees that reduce the residual errors from the last iteration (Bhagat et al.,2020). GBM is an ensemble-based method for regression, classification purposes, and applying a weak classifier on the data to generate the set of decision trees (Bhagat et al.,2021).

D) Linear Regression

Linear regression is a type of supervised machine learning algorithm that computes the linear relationship between a dependent variable and one or more independent features. The goal of the algorithm is to find the best linear equation that can predict the value of the dependent variable based on the independent variables. The ET_0 calculation method was tested by focusing on the most basic linear regression analysis algorithm among machine learning algorithms. MLR and polynomial regression (PR) algorithms were applied based on the composition of independent variables. Linear regression learns relatively quickly, has a high explanatory power,

and has no significant difference in performance compared to other algorithms. The MLR model is applied when one dependent variable and two or more independent variables are used as input data. This method is significant when each independent variable has a linear relationship with the dependent variable;

2.4 Development and validation of models

For the training and testing of the models, the data set was randomly divided into a training set (with 70% of the data) and a test set (with 30% of the data). The training set was used to calibrate ET_0 equations and to model ET_0 with heuristic models. The prediction of the test set was used to evaluate the performance of the equations and models**.**

2.5 Model performance metrics

The efficacy of ML models in the estimation of ET_0 was assessed through a comparative analysis against the established FAO-PM approach, recognized as the standard. Initially, the FAO-PM equation was applied as a reference method, and ET_0 values derived from other ML models were juxtaposed with this standard to ascertain accuracy.To gauge the performance and precision of the developed ML models, various statistical metrics were employed. These included the coefficient of determination (R^2, Nagelkerke 1991), Mean Absolute percentage Error (MAPE), Root Mean Square Error (RMSE, Huffman 1997), Mean Absolute Error (MAE), and Mean Square Error (MSE). The model that yielded the highest R^2 value while simultaneously producing the lowest RMSE, MBE, and MAE values was identified as the optimal model. Detailed mathematical expressions for these statistical indicators can be found in equation 2-6 for reference.

$$R^2 = \left\{ \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{\left[n\sum {x_i}^2 - (\sum x_i)^2\right]\left[n\sum {y_i}^2 - (\sum y_i)^2\right]}} \right\}^2 \quad \ldots 2$$

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{N} \quad \ldots 3$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{N}} \quad \ldots 4$$

$$MSE = \frac{1}{n}\sum_{i=1}^{N}(y_i - x_i)^2 \quad \ldots 5$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - x_i}{y_i}\right| \quad \ldots 6$$

Where, y_i is observed i^{th} value, x_i is predicted i^{th} value, $\overline{y_i}$ is average of observed values, N is the total number of observations.

3. Result and Discussion

The characteristics of daily meteorological variables are shown in Table 1. The daily minimum temperature varied from 16.5 to 27.7 °C, while the maximum temperature varied from 24 to 40.4 °C. The daily wind speed and number of sunshine hours in the study area ranged from 0 to 3.8 m/s and 0 to 10.8 h/day, respectively. The ET_0 varied between 1.76 and 8.41 mm/day.**Figure 2 and 3** shows the monthly variation of climatic variables during 2015–2019 in the study area.

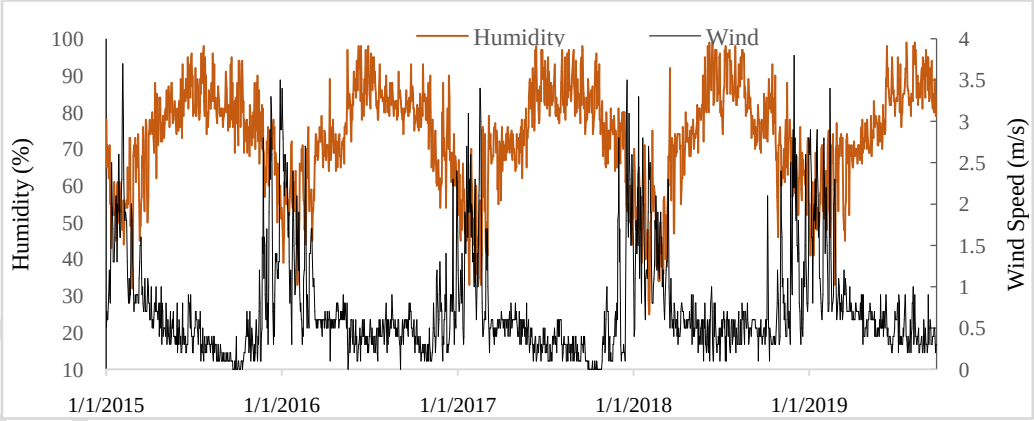

Figure 2. Annual fluctuation in humidity and wind speed within the study area

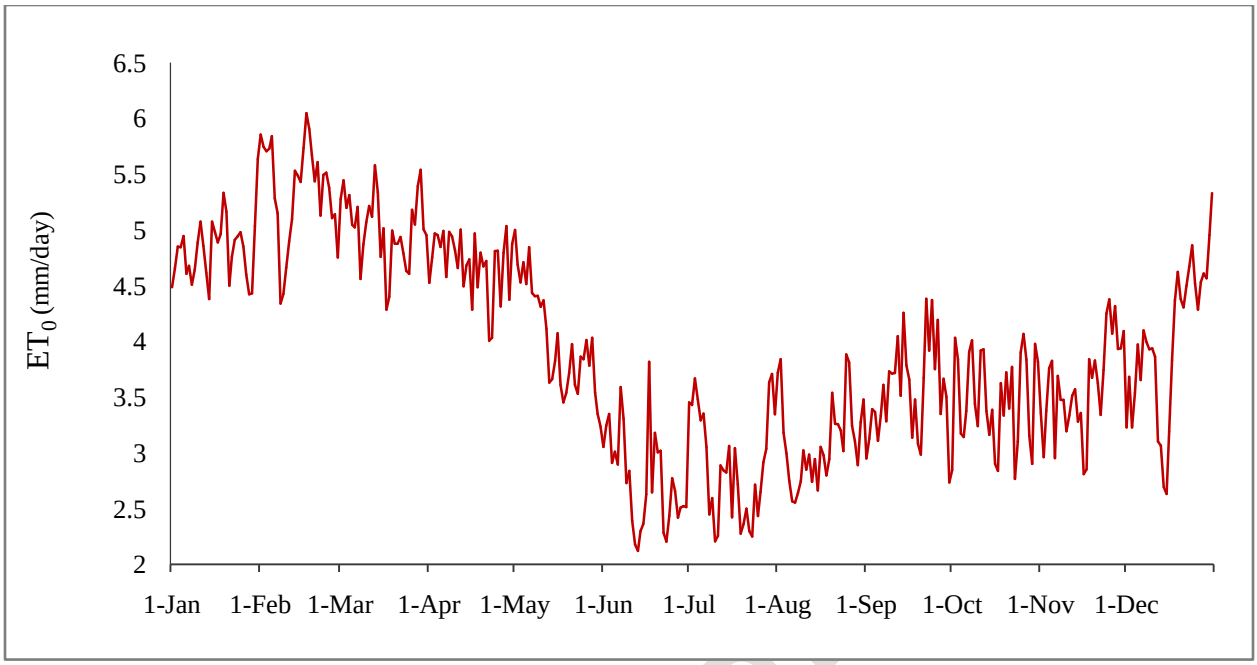

Figure 3. Annual fluctuation in reference Evapotranspiration within the study area

The results of the performance of various Machine learning models are presented in Table **1**for training and testing periods. The evaluation was conducted on both training and testing datasets to assess the models' ability to generalize and predict outcomes accurately.

Table 1: Statistical indices of ML models for performance analysis

	MSE		**RMSE**		**MAE**		**MAPE**		**R^2**	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
RF	0.007	0.016	0.085	0.127	0.049	0.08	0.012	0.021	0.995	0.989
SVM	0.007	0.011	0.085	0.105	0.062	0.072	0.018	0.024	0.995	0.992
LR	0.019	0.015	0.137	0.122	0.094	0.089	0.024	0.026	0.987	0.99
GBM	0.005	0.011	0.072	0.107	0.054	0.077	0.014	0.022	0.996	0.992

In the present study, all four models (Random Forest, Support Vector Machine, Logistic Regression, and Gradient Boosting Machine) exhibited relatively low Mean Squared Error (MSE) values for both the training and testing datasets.Among these models, SVM and GBM consistently demonstrated the lowest MSE values for both training and testing datasets, with SVM recording 0.007 and 0.011 for training and testing, respectively, and GBM recording 0.005

for training and 0.011 for testing. On the other hand, LR and RF displayed relatively higher MSE values, with LR having 0.019 for training and 0.015 for testing, and RF with 0.007 for training and 0.016 for testing.Additionally, all models displayed low Root Mean Squared Error (RMSE) values, indicating their accuracy in estimating the target variable. SVM (with values of 0.085 for training and 0.105 for testing) and GBM (0.072 for training and 0.107 for testing) consistently outperformed RF (0.016 for training and 0.085 for testing) and LR (0.015 for training and 0.137 for testing) in terms of RMSE, underscoring their ability to provide more precise predictions with lower error. Furthermore, all models yielded low Mean Absolute Error (MAE) values for both training and testing datasets, signifying their proficiency in predicting the target variable with minimal absolute errors. RF and GBM had the lowest MAE values, with RF recording 0.049 for training and 0.08 for testing, and GBM with 0.054 for training and 0.077 for testing. SVM and LR also performed well in this aspect, though they had slightly higher MAE values, with SVM recording 0.062 for training and 0.072 for testing, and LR with 0.094 for training and 0.089 for testing.Additionally, all models maintained low Mean Absolute Percentage Error (MAPE) values, highlighting their ability to provide reliable predictions. RF and GBM exhibited the lowest MAPE values, with RF recording 0.012 for training and 0.021 for testing, and GBM with 0.014 for training and 0.022 for testing. SVM and LR also maintained low MAPE values, with SVM recording 0.018 for training and 0.024 for testing, and LR with 0.024 for training and 0.026 for testing, underscoring their practicality for various applications.The results also indicated that all machine learning models demonstrated high R^2 values for both training and testing datasets (Figure 4 and 5), suggesting their capability to explain a substantial portion of the variance in the data. GBM and SVM consistently achieved the highest R^2 values, implying their superiority in explaining the variation in the target variable, while RF and LR also delivered respectable R^2 values, indicating their strong predictive performance.

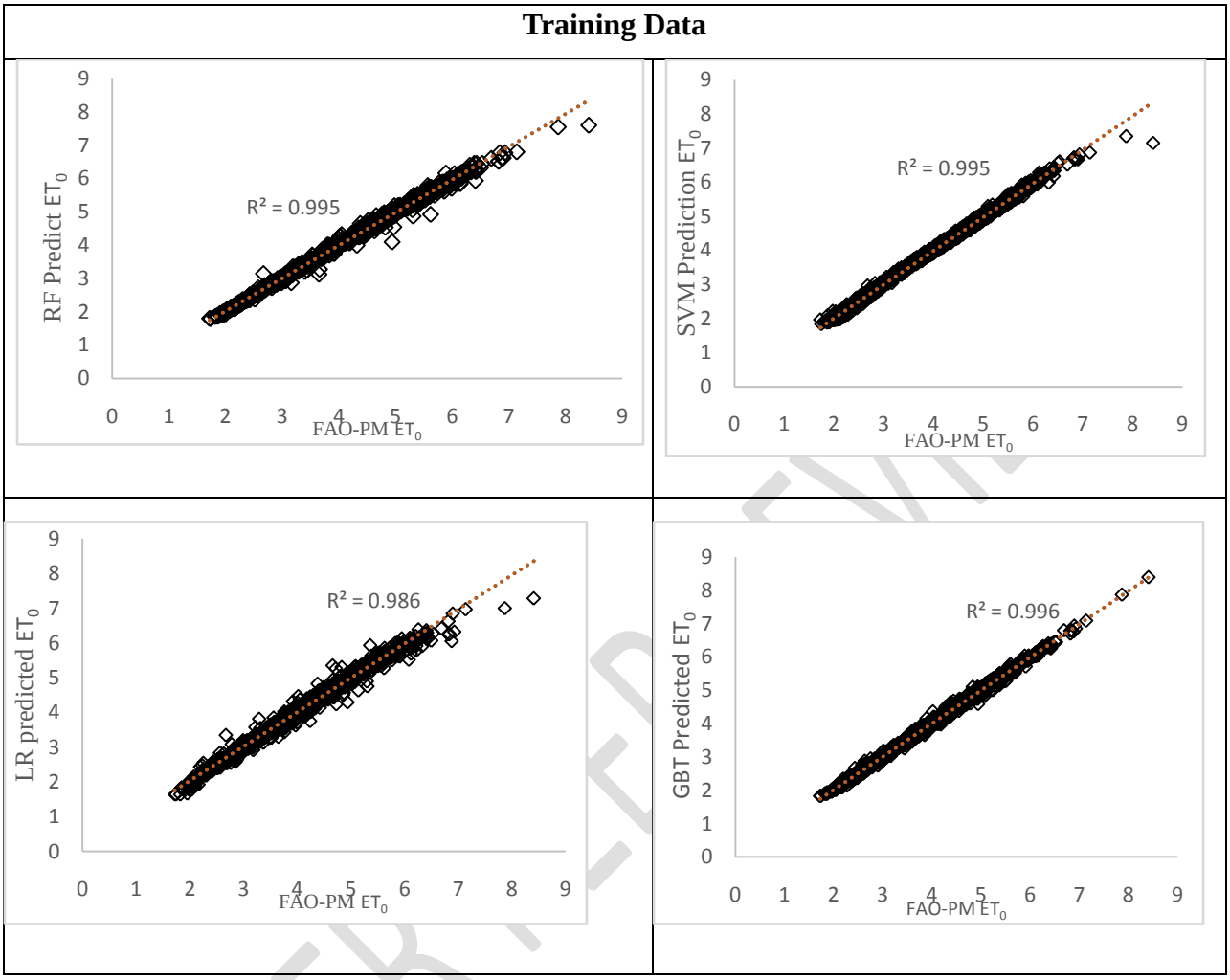

Figure 4. Coefficient of Determination for training set data

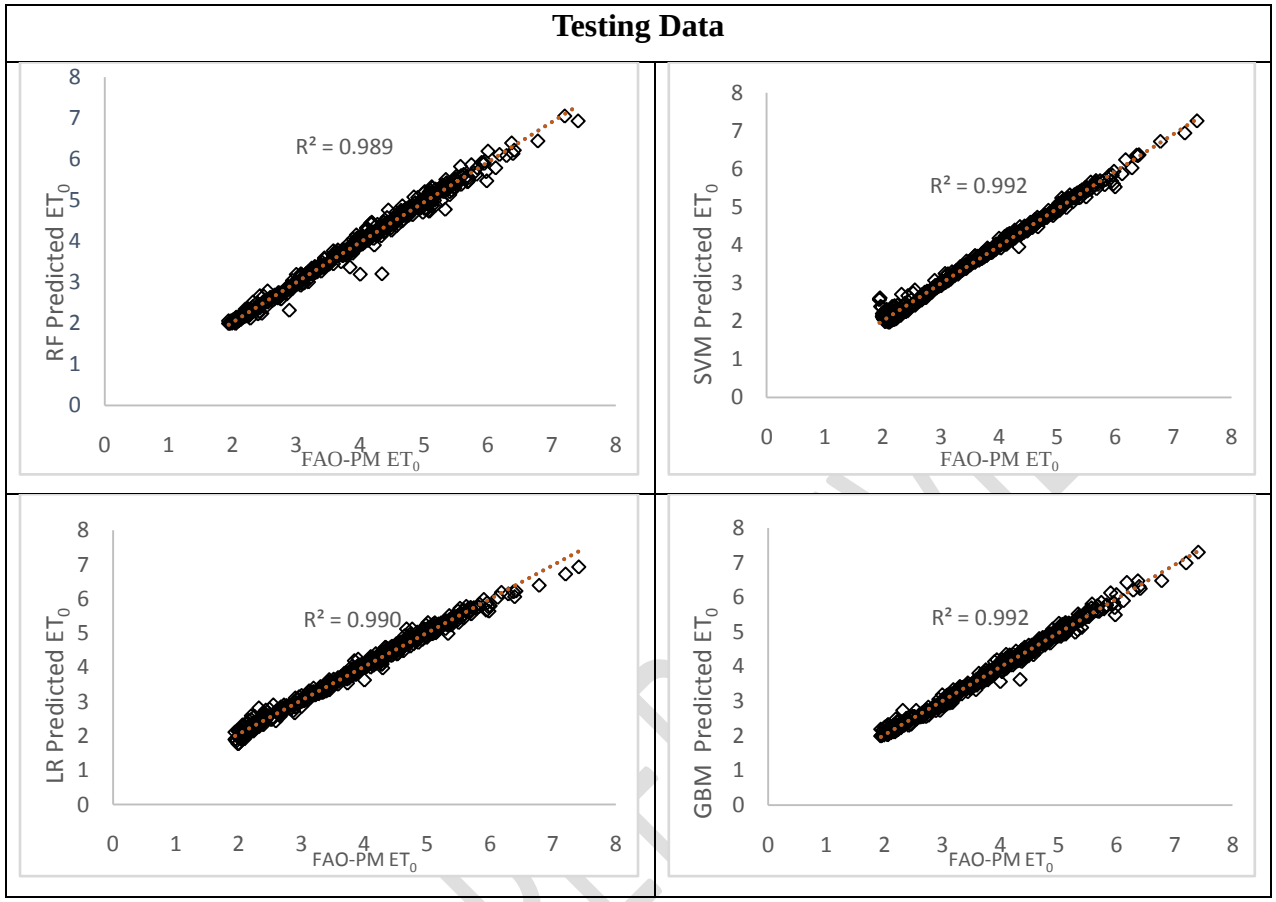

Figure 5. Coefficient of Determination for testing set data

In this study, we observed that SVM and GBM consistently outperformed RF and LR in terms of MSE, RMSE,and R^2.However, MAE and MAPE found that RF and GBM out performed SVM and LR ML models. This consistency suggests that SVM and GBM have a robust performance across various aspects of predictive modeling. The superiority of these ensemble-based methods is often attributed to their ability to capture complex relationships in the data by combining multiple learners.It's important to note that while SVM and GBM exhibited superior predictive accuracy, they may lack the interpretability that linear models like LR provide.SVM and GBM can be quite slow and need a lot of computer power because they have to make many different decisions. On the other hand, RF and LR are simpler and quicker to use, which makes them a better choice if you need fast results or have limited computer resources. The results emphasize the importance of considering multiple evaluation metrics to make informed decisions about which ML algorithm best suits a particular task. Moreover, it

highlights the potential of ensemble methods like SVM and GBM for achieving high predictive accuracy of ET_0 Assessment in the present study.SVM and GBM, which consistently outperformed RF and LR, could be preferred in scenarios where high prediction accuracy is paramount. Similar kind of study is also reported by the various researchers throughout the world. In a study carried out by Duhan et al. in 2023, an investigation was undertaken to evaluate the accuracy of estimating ET_0 using various machine learning models. The results revealed that machine learning models produced R^2 values ranging from 0.800 to 0.998, with the highest value (0.998) observed in the Least Square-SVM model. Similarly, in 2018, another study conducted by a different group evaluated the performance of different machine learning algorithms, including SVM, RF, and the extreme gradient boosting algorithm (XGboost), in estimating ET. The study found that SVM demonstrated the best performance with both accuracy and stability. Mokari et al.2022 conducted a study in New York to determine the most suitable machine learning model for a specific area. Their findings indicated that SVM, followed by the Random Forest model, and then genetic programming, were the most suitable options. These results align with previous research. However, Wu et al. in 2019 reported contrasting results. Their study, conducted in Jiangxi Province, China, found that the Random Forest model was the best choice for estimating ET_0, diverging from our findings and those of other researchers.

This study holds significant implications for the scientific community as it addresses the critical issue of accurately estimating ET_0, which has broad applications in drought detection, irrigation scheduling, water resource management. By combining the FAO-PM equation with machine learning models, including Random Forest, Support Vector Machine, Gradient Boosting Machine, and Linear Regression, the research enhances the precision, accuracy of ET_0 predictions. Besides that, ML large datasets efficiently, making them suitable for analyzing extensive historical weather data. These models are having ability to learn and retaining may enhance the predict capacity of the model. Additionally, ML automates the ETo estimation process, reducing the need for manual calculations and potentially costly errors associated with human input. The meticulous evaluation of these models using various statistical indices not only identifies the most effective ML algorithm—Gradient Boosting Machine—but also provides a valuable benchmark for future research and practical implementations. Furthermore, the study's regional relevance underscores the importance of tailoring ET_0 estimation methods to specific geographic contexts, offering insights and tools that cater to diverse regional needs

Moreover, the choice of the machine learning model should match the application's needs, considering factors like accuracy, efficiency, adaptability, and the balance between complexity and interpretability.The findings support the potential of machine learning in improving the precision and efficiency of ET_0 predictions, which are essential for sustainable agricultural practices and water resource management in a changing climate.Besides that, these data-driven algorithms have the ability to capture complex relationships and patterns in the data, which can lead to more accurate predictions of ET_0.

In recent years, there has been a frequent occurrence of natural disasters like drought, influenced by both natural forces and human activities. These disasters have displayed escalating intensity, along with peculiar and unpredictable patterns. Evapotranspiration serves as a crucial indicator for tracking drought conditions. Consequently, making accurate ET predictions holds significant importance for crafting precise irrigation strategies, monitoring dry conditions in croplands, and enhancing water usage efficiency. ML algorithms possess the capability to capture intricate relationships between input and output data, making them effective tools for solving nonlinear problems. As a result, ML techniques are widely employed in estimating ET. Therefore, our research involved a comparison of the most commonly used ML algorithms to identify the top-performing one for ET estimation on a regional scale.

Future Directions for Research

Future research can focus on fine-tuning the selected models or exploring new, more advanced machine learning techniques to achieve even higher accuracy in ET_0 estimation. Additionally, work may take to explore the synergy between ML models and remote sensing technologies to improve the accuracy and spatial resolution of ET_0 predictions. Furthermore, the accuracy of the estimated model in this study needs to be tested in other regions, and any similarities in accuracy among these regions should be assessed.

4. Conclusion

Precise estimation of cropland ET plays a pivotal role in drought detection, irrigation scheduling, water management and the implementation of effective mitigation measures to reduce disaster risk. This study initially employed the FAO-PM equation for ET_0 estimation and subsequently integrated meteorological variables as inputs into machine learning models for

enhanced ET_0 prediction. The model outputs were rigorously evaluated using various statistical indices, including MAE, MSE, MAPE, RMSE, and R^2 values, with the goal of identifying the best-performing model. Among the evaluated ML algorithms (RF, SVM, GBM, and LR), GBM emerged as the top-performing model, followed by SVM and RF, while LR exhibited the least accuracy when compared to GBM. Therefore, based on the findings, it is concluded that the GBM machine learning model is the most suitable choice for the study region. Future research may explore additional machine learning models to gain further insights and improve ET_0 prediction capabilities.

5. 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, FAO, Rome.

Arti Kumari, Ashutosh Upadhyaya, Pawan Jeet, Nadhir Al-Ansari * , Jitendra Rajput, Alban Kuriqi, P K Sundaram, Kirti Saurabh, Ved Prakash, AK Singh, Rohan Kumar Raman and Venkatesh Gaddikeri. 2022. Estimation of Actual Evapotranspiration and Crop Coefficient of Transplanted Puddled Rice using Modified Non-weighing paddy Lysimeter. *Agronomy*. 12(11), 2850; https://doi.org/10.3390/agronomy12112850

Bachour, R., Maslova, I., Ticlavilca, A. M., Walker,W. R. & McKee, M. 2016 Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. *Stochastic Environmental Research and Risk Assessment* 30 (1), 103–117. https://doi.org/10.1007/s00477-015-1039-z.

Bhagat, S. K., Tiyasha, T., Tung, T. M., Mostafa, R. R., & Yaseen, Z. M. (2020). Manganese (Mn) removal prediction using extreme gradient model. *Ecotoxicology and Environmental Safety*, 204, 111059.

Bhagat, S. K., Tung, T. M., & Yaseen, Z. M. (2021). Heavy metal contamination prediction using ensemble model: Case study of Bay sedimentation, Australia. Journal of Hazardous Materials, 403, 123492. https://doi.org/10.1016/j.jhazmat.2020.123492

Blaney, H. F. (1952). Determining water requirements in irrigated areas from climatological and irrigation data.

Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32

Chia, M. Y., Huang, Y. F. & Koo, C. H. 2021 Swarm-based optimization as stochastic training strategy for estimation of reference evapotranspiration using extreme learning machine. *Agriculture Water Management* 243, 106447. https://doi.org/10.1016/j.agwat.2020.

Chia, M. Y., Huang, Y. F., Koo, C. H. & Fung, K. F. 2020b Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques – a review. *Agronomy* 10 (1), 101. https://doi.org/10.3390/agronomy10010101

Dimple, Pradeep Kumar Singh, Jitendra Rajput, Dheeraj Kumar, Venkatesh Gaddikeri, Ahmed Elbeltagi. Combination of discretization regression with data-driven algorithms for modeling irrigation water quality indices. *Journal of Ecological Informatics*. https://doi.org/10.1016/j.ecoinf.2023.102093

Duhan, D., Singh, D. & Arya, S. 2021 Effect of projected climate change on potential evapotranspiration in the semiarid region of central India. *Journal of Water and Climate Change* 12 (5), 1854–1870.

Fan, J.; Yue, W.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol*. 2018, 263, 225–241.

Feng, Y., Peng, Y., Cui, N., Gong, D., & Zhang, K. (2017). Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Computers and Electronics in Agriculture*, 136, 71-78.

Gangopadhyay, M., Uryvaev, V. A., Oman, M. H., Nordenson, T. J. & Harbeck, G. E. 1966 Measurement and Estimation of Evaporation and Evapotranspiration. World Weather Organization, Geneva, Switzerland

Geigy, R., Jenni, L., Kauffmann, M., Onyango, R. J., & Weiss, N. (1975). Identification of T. brucei-subgroup strains isolated from game. *Acta tropica*, 32(3), 190-205.

generalized regression neural network only with temperature data. *Computers and Electronics in Agriculture* 136, 71–78.

Huffman, G. J. 1997 Estimates of root-mean-square random error for finite samples of estimated precipitation. *Journal of Applied Meteorology and Climatology* 36 (9), 1191–1201.

Jensen, M.E., Walter, I.A., Allen, R.G., Elliott, R., Itenfisu, D., Mecham, B., Howell, T.A., Snyder, R., Brown, P., Echings, S., Spofford, T., Hattendorf, M., Cuenca, R.H., Wright, J.L. & Martin, D., 2012.ASCE's Standardized Reference Evapotranspiration Equation 1-11.https://doi.org/10.1061/40499(2000)126

Jitendra Rajput, Man Singh, K. Lal, Manoj Khanna, A. Sarangi, J. Mukherjee,Shrawan Singh. 2023. Selection of Alternate Reference Evapotranspiration Models based on Multi-Criteria Decision Ranking for Semi-Arid Climate. *Journal of Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-023-03234-9

Jitendra Rajput, Man Singh, K. Lal, Manoj Khanna, A. Sarangi, J. Mukherjee, Shrawan Singh. 2023. Assessment of Data Intelligence Algorithms in Modeling Daily Reference Evapotranspiration under Input Data Limitation Scenarios in Semi-Arid Climatic Condition. *Journal of Water Science and Technology*. 87(10):2504–2528. doi: https://doi.org/10.2166/wst.2023.137

Jitendra Rajput, Man Singh, K. Lal, Manoj Khanna, A. Sarangi, J. Mukherjee, Shrawan Singh. 2022. Performance Evaluation of Soft Computing Techniques for Forecasting Daily Reference Evapotranspiration. *Journal of Water and Climate Change*. 14 (1): 350–368. doi: https://doi.org/10.2166/wcc.2022.385

Keshtegar, B., Kisi, O., Arab, H. G. & Zounemat-Kermani, M. 2018 Subset modeling basis ANFIS for prediction of the reference evapotranspiration. *Water Resources Management* 32 (3), 1101–1116.

Malik, A. & Kumar, A. 2015 Pan evaporation simulation based on daily meteorological data using soft computing techniques and multiple linear regression. *Water Resources Management* 29 (6), 1859–1872.

Mehdizadeh, S., Behmanesh, J. & Khalili, K. 2017 Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and Electronics in Agriculture* 139, 103–114. https://doi.org/10.1016/j.compag.2017.05.00.

Misra, S.; Li, H. Chapter 9—Noninvasive Fracture Characterization Based on the Classification of Sonic Wave Travel Times. In *Machine Learning for Subsurface Characterization*; Misra, S., Li, H., He, J., Eds.; Gulf Professional Publishing: Houston, TX, USA, 2020; pp. 243–287. ISBN 978-0-12-817736-5.

Naganna, S. R., Beyaztas, B. H., Bokde, N., & Armanuos, A. M. (2020). On the evaluation of the gradient tree boosting model for groundwater level forecasting. *Knowledge-Based Engineering and Sciences*, *1*(1), 48–57. https://doi.org/10.51526/kbes.2020.1.01.48-57

Nagelkerke, N. J. D. 1991 A note on a general definition of the coefficient of determination. *Biometrika* 78 (3), 691–692

Press Information Bureau (PIB). (2020). Per Capita Availability of Water. Available online https://pib.gov.in/PressReleasePage.aspx?PRID=1604871#:~:text=The%20average%20annual%20per%20capita,years%202021%20and%202031%20respectively.

Reis, M. M., daSilva, A. J., Zullo Junior, J., TuffiSantos, L. D., Azevedo, A. M. & Lopes, É. M. G. 2019 Empirical and learning machine approaches to estimating reference evapotranspiration based on temperature data. *Computers and Electronics in Agriculture* 165,104937.

Seifi, A. & Riahi, H. 2020 Estimating daily reference evapotranspiration using hybrid gamma test-least square support vector machine, gamma test-ANN, and gamma test-ANFIS models in an arid area of Iran. *Journal of Water and Climate Change*. 11 (1), 217–240

United Nations (UN), Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Highlights (ST/ESA/SER.A/423),

Vapnik, V. The Nature of Statistical Learning Theory; Springer: New York, NY, USA, 1995.

Wu, L., Peng, Y., Fan, J., & Wang, Y. (2019). Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. *Hydrology Research*, 50(6), 1730-1750.