

Crop water management using machine learning-based evapotranspiration estimation

R. Meenal^{1*}, Prakash Kumar Jala², R. Samundeswari³, E. Rajasekaran⁴

 Department of Electrical and Electronics Engineering, V.S.B. Engineering College, Karur, Tamil Nadu, India. Department of Electrical and Electronics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India. Department of Agriculture, Karunya Institute of Technology and Sciences, Coimbatore, India. Department of Science and Humanities, V.S.B. Engineering College, Karur, Tamil Nadu, India.

1. INTRODUCTION

Due to climate change and the country's expanding population, the agricultural division of India must look for more precise and effective strategies to provide a sufficient and consistent supply of food for the community while using the least amount of water possible. Measuring evapotranspiration (ET) is the most significant factor in the irrigation schedule [\[1](#page-5-0)[,2\].](#page-5-1)

Accurate estimation of ET is the first step in evaluating the water requirements of field crops [[3\]](#page-5-2). Several artificial intelligence-based models are used for irrigation planning [\[4\].](#page-5-3) The valuation of water for crops is most important in water distributions $[5]$. ET_0 refers to the loss of water to the surface of plants and soil $[6]$ $[6]$. Evaporation parameters are used to study water budgets, water resource management, and irrigation system design as well as to estimate plant growth and height [[7\]](#page-5-6). ET plays an important role in different fields of hydrology and agriculture [[2\]](#page-5-1). Precise estimation of ET_{0} is very important for irrigation planning, scheduling, design, and crop water management. ET is measured by various methods such as (i) Lysimetric, (ii) field experiment, (iii) water balance, and (iv) soil moisture depletion study. Lysimeters are tough and expensive to build, their operation and maintenance require particular care, and their use is restricted to specific research purpose. ET changes with climate change and as the climate has many geographical

weather data and so no robust models. As a consequence of cost and difficulties in direct measurement techniques with a pyranometer and lysimeter, solar radiation and ET_{0} were predicted using suitable models [[8\]](#page-5-7). Different empirical models have been developed for ET_0 estimation rendering to various climatic conditions $[9,10]$ $[9,10]$. Many models such as empirical, artificial neural network, machine learning (ML), and deep learning exist in the literature to compute the global solar radiation (GSR) and ET_0 [\[11](#page-5-10)[,12\]](#page-5-11). However, the standard method recommended by the Food and Agriculture Organization (FAO), namely, the Penman-Monteith (PM-FAO56) equation requires an extensive range of data support for ET estimation. In this work, ML models with a limited number of input parameters are utilized to estimate the solar radiation and ET in chosen locations of Tamil Nādu. To estimate the ET_{ρ} , the solar radiation and temperature values are used. Empirical correlations are also utilized to estimate the $ET₀$ for comparison with ML methods. Based on the performance metrics, ML-based ET_0 estimations are more accurate than empirical-based estimations. To develop the ML model, SVM and random forest algorithms are employed with a reduced number of meteorological parameters.

variations, the pre-developed systems have not used all available

2. STUDY AREA AND DATA SOURCES

The study site, Coimbatore has a semi-arid tropical climate. The tomato is one of the horticultural products produced in the study location. The water requirement for tomato plant in the study location is calculated by the relation of crop coefficient and ET. The geographical parameters of the study location are given in [Table](#page-1-0) 1 and [Figure](#page-2-0) 1 shows the monthly

^{}Corresponding Author:*

R.Meenal, Department of Electrical and Electronics Engineering, V.S.B. Engineering College, Karur, Tamil Nadu, India. E-mail: meenasekar5 @ gmail.com

^{© 2024} R.Meenal, *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License -NonCommercial-ShareAlike Unported License (http://creativecommons.org/licenses/by-nc-sa/3.0/).

average GSR and temperature values. Meteorological parameters associated with this study were obtained from India meteorological department (IMD), India and AQUASTAT tool. The measured experimental data for the estimation of GSR and ET, namely, the monthly average maximum and minimum temperature, bright sunshine duration, and daily solar radiation in kW/m²/day for 12 locations in India were obtained from IMD. Table 2 presents the measured GSR data obtained from IMD. The whole dataset is partitioned into two sets, namely, training sub-set (70%) and testing sub-set (30%). It is further processed with the assistance of empirical and ML models. The coefficients of the solar radiation empirical models are estimated with the help of a training data set and testing datasets are used to validate the models.

3. MATERIALS AND METHODS

3.1. Estimation of Reference ET

The lysimetric method is used for the direct measurement of *in situ* $ET₀$ values which are considered as the only best method to get accurate estimates of ET_{0} . However, this lysimetric method has many drawbacks associated with the high cost and difficulties in the maintenance of

Table 1: Geographical parameters of the study site.

| Latitude ^o (n) | Longitude \circ (E) | Climate | Annual Solar Radiation $MJ/m^2/day$ | Average Evapotranspiration in mm/day |
|------------------------------|--------------------------|----------------|---|--|
| 11.02 | 76.95 | Tropical | 19.5 | 4.5 |

complex instrumentation. PM-FAO56 equation has been considered as the standard model for the estimation of ET_o for crop water requirements under different climate conditions and various time scales.

 ET_0 . Reference ET is given by the following equation [\[10\]](#page-5-9).

$$
FAO - ET_0 = \frac{0.408\Delta (R_n - G) + \gamma \frac{900}{T_{\text{avg}+273}} u_{2(e_s - e_a)}}{\Delta + \gamma (1 + 0.34 u_2)}
$$
(1)

Where R_n is GSR in MJ m⁻² d⁻¹, G is the soil heat flux density MJ m⁻² d⁻¹ T_{avg} is the average air temperature (°C) e_s is the saturation vapor pressure (kPa) e_a is the actual vapor pressure (kPa) ∆ Is the slope of the vapor pressure (kPa°C−1) γ is the psychrometric constant (kPa°C−1) u_2 is the wind speed in m/s.

The above equation (1) requires a huge exhaustive meteorological data set. Simpler empirical equations which require only the maximum and minimum air temperature and solar radiation data are used in this study for the estimation of ET.

3.2. Estimation of GSR Using Empirical Correlations

Empirical models are math and coefficient based. This classical statistical approach relies on the linear relationship between inputs and

Table 2: Measured monthly mean GSR in kW/m2 /day for Indian Cities from IMD for training ML models.

| Months | Location/global solar radiation | | | | | | | |
|---------------|---------------------------------|------------------|---------------|-----------|---------|--------------------|------------------|-------------------|
| | Patna | New Delhi | Nagpur | Hyderabad | Chennai | Bhubaneswar | Mangalore | Trivandrum |
| January | 3.3 | 3.2 | 4.3 | 5.2 | 4.7 | 4.1 | 4.9 | 4.9 |
| February | 4.2 | 4.1 | 5.2 | 5.8 | 5.8 | 4.7 | 5.6 | 5.5 |
| March | 5.2 | 5.1 | 5.8 | 6.4 | 6.2 | 5.6 | 5.5 | 5.4 |
| April | 5.7 | 5.9 | 6.5 | 6.6 | 6.3 | 6.4 | 5.9 | 5.1 |
| May | 5.7 | 5.8 | 6.6 | 6.7 | 6.0 | 6.2 | 5.2 | 4.8 |
| June | 4.8 | 5.6 | 5.2 | 5.3 | 5.6 | 4.6 | 4.9 | 4.6 |
| July | 4.0 | 4.7 | 3.8 | 4.8 | 5.1 | 4.3 | 3.6 | 4.5 |
| August | 4.5 | 4.4 | 4.0 | 4.6 | 5.2 | 4.2 | 4.0 | 4.9 |
| September | 4.1 | 4.3 | 4.9 | 4.9 | 5.2 | 4.0 | 4.8 | 5.0 |
| October | 4.1 | 4.3 | 5.1 | 5.0 | 4.6 | 4.9 | 5.1 | 4.5 |
| November | 3.6 | 3.5 | 4.4 | 5.4 | 4.0 | 4.4 | 5.0 | 4.0 |
| December | 3.2 | 2.9 | 4.1 | 5.0 | 3.7 | 3.7 | 4.9 | 4.4 |

ML: Machine learning, GSR: Global solar radiation, IMD: India meteorological department.

Table 3: Empirical correlations used for the estimation of solar radiation.

outputs and exists as mathematical equations. There are two sets of empirical models. The first set of models estimate solar radiation using bright sunshine hours and the second set of temperature-based models estimate the solar radiation using the minimum, maximum, and mean temperature data which is suitable for locations where the sunshine data is unavailable.

Some of the sunshine and temperature-based empirical models are shown in Table 3. [Figure](#page-3-0) 2 presents the linear relationship between the solar radiation and the input weather parameters by estimating

Figure 1: Monthly average solar radiation in MJ/m²/day and temperature of the study site-Coimbatore.

Table 4: Performance metrics of empirical models.

| Performance <i>indicators</i> | Linear model | Quadratic model | Cubic model |
|----------------------------------|-----------------|---------------------------|-----------------------|
| Sunshine models | | | |
| R | 0.9784 | 0.9783 | 0.9779 |
| RMSE | 0.6953 | 0.6963 | 0.6953 |
| MAPE | 0.6450 | 0.6400 | 0.6127 |
| Temperature models | | | |
| R | 0.9191 | 0.9198 | 0.9217 |
| RMSE | 1.2428 | 1.2434 | 1.2318 |
| MAPE | 1.0419 | 1.0576 | 1.0559 |

RMSE: Root mean squared error, MAPE: Mean absolute percentage error.

Table 5: An overview of AI/ML-based Climate-smart agriculture approaches.

the "R" value $[13]$. From this, it is observed that sunshine duration and temperature received the first two ranks and are more closely correlated with GSR. Hence, sunshine duration and temperature-based empirical and ML models are selected for the estimation of GSR. Table 3 summarizes the sunshine and temperature-based empirical correlations employed for the estimation of the solar radiation. Under the empirical category, although the model performances are close to each other, the simple linear model is recommended as the optimum model due to lesser computational effort when compared with quadratic and cubic models.

Temperature and sunshine-based empirical correlations and ML techniques such as SVM and tree-based random forest ML algorithms are used for the estimation of GSR. Spyder an open-source crossplatform integrated development environment written in python language for scientific programming is used to develop computer codes for ML models.

Various statistical indexes, namely, correlation coefficient (R), root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean bias error (MBE) are used to assess the performance of the models. For better modeling accuracy RMSE, MAPE, and MBE indices should be nearer to zero, but R-value should be nearer to 1.

The MAPE shows the average absolute percentage deviation between the calculated and the actual observed GSR data and is determined by:

$$
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{H_{im} - H_{ic}}{H_m} \right| \times 100\%
$$
 (2)

Information regarding the long-term performance is provided by the mean bias error. This is an indicator of the average deviation of the forecast values from the actual observed GSR values. Mean bias error is determined by:

$$
MBE = \frac{1}{n} \sum_{i=1}^{n} (H_{im} - H_{ic}) \times 100
$$
\n(3)

The RMSE determines the model's accuracy by comparing the deviation between the predicted and actual GSR data. The RMSE always has a non-negative value and is computed by:

ML: Machine learning, RMSE: Root mean squared error, MAPE: Mean absolute percentage error, GSR: Global solar radiation.

Figure 2: Linear relationship between solar radiation and various weather parameters.

Figure 3: Comparision between predicted evapotranspiration data using empiricaland machine learning method and satelite data for the study site: 11.0168° N latitude, 76.9558° E longitude.

$$
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (H_{im} - H_{ic})^2}
$$
 (4)

Where $H_{i,c}$ is the *i*th computed value $H_{i,m}$ is the *i*th observed value of GSR and *n* is the total number of observations. The error statistics of the temperature and sunshine models are available in [Table](#page-2-0) 4. From this, it is proved that the sunshine empirical models are more precise than temperature-based empirical models. Temperature models can be used in locations where sunshine records are not available.

3.3. Estimation of GSR and ET Using ML Models

ML is the division of study that gives the ability of computers to learn from huge amounts of data [[14\].](#page-5-18) These ML and deep learning are contemporary technologies that can be used efficiently to predict the water requirements for agricultural crops. Accurate estimation of ET is the first step in evaluating the water requirements of field crops. Data on solar radiation are crucial for estimating ET and managing crop water. SVM and Random Forest are the two most widely utilized ML techniques for estimating solar radiation and ET. In the literature, there are also deep learning models [[15\]](#page-5-19) and hybrid ML models for predicting ET [\[16](#page-5-20)[,17\]](#page-5-21). [Table](#page-2-0) 5 provides a summary of ML-based methods for climate-smart agriculture approach.

Random forest ML algorithm is more suited to classification problems when compared with regression problems [[18\].](#page-5-22) It creates decision trees using a variety of samples, taking into account the average in regression problems and the majority votes for classification problems. The random forest algorithm's ability to handle both continuous and categorical variables is one of its most important characteristics. Both classification and regression analysis can be performed using SVM, which are robust supervised learning models [\[19](#page-5-23),[20\].](#page-5-24) Finding a hyperplane with the best decision boundary that clearly classifies the data points is the goal of the SVM algorithm. [Table](#page-4-0) 6 presents the estimated solar radiation values using the ML models, namely, the random forest algorithm and SVM.

4. CROP WATER REQUIREMENT

For all major crops, the irrigation water requirement is determined as follows [\[21\]](#page-5-25):

Table 6: Estimation of solar radiation in MJ/m²/day using machine learning models for the location-Coimbatore.

| Months | Random forest | SVM |
|---------------|----------------------|------------|
| January | 20.98 | 23.13 |
| February | 22.39 | 24.99 |
| March | 22.53 | 25.39 |
| April | 21.77 | 22.07 |
| May | 19.99 | 19.77 |
| June | 15.95 | 16.11 |
| July | 14.87 | 15.03 |
| August | 15.03 | 16.08 |
| September | 17.44 | 17.77 |
| October | 17.44 | 17.11 |
| November | 17.45 | 17.88 |
| December | 18.07 | 18.96 |

ML: Machine learning, ET₀: Evapotranspiration.

- Step 1: Estimate the reference crop ET: ET_0
- Step 2: Get the crop factors: Kc
- Step 3: Compute the crop water need: $ETc = ET_0 \times Kc$

The ET_0 is estimated by solar radiation-based empirical models and random forest algorithm. To estimate the ET, the solar radiation and temperature values are used. The water requirement for the crops (ETc) in millimeter/day is determined by multiplying the estimated reference crop ET and crop coefficient (k_c) value.

The estimated solar radiation and the measured temperature values are used as input variables for empirical and random forest ML model to predict the ET_0 value for the particular crop per day to understand how much water is required for the crop. The water requirement for the tomato crop in millimeter/day is calculated by multiplying the estimated ET and crop coefficient (k_c) value. The estimated ET using empirical and random forest ML models is compared. From the results, Table 7 and [Figure](#page-3-0) 3, it is observed that ET_0 predictions using ML model are more precise than empirical models. Hence, the predicted $ET₀$ value is used for agro-meteorological applications. Figure 4 shows the water requirement for the tomato crop. From the results, it was observed that the annual average solar potential of Coimbatore is

Figure 4: Crop water need evapotranspiration in mm/day for the tomato crop.

around 19.69 MJ/m²/day. The reference crop ET ranges from 3.7 to 5.7 mm/day.

5. CONCLUSION

Although weather prediction systems have undergone significant advancements recently, their direct applications in efficient crop water management have not been fully investigated. Here, we made an attempt to apply the ML techniques to estimate the ET through which the water requirement is calculated for the field crops in the theme of the application of weather in agriculture. Random forest ML algorithm and empirical correlations are used to predict the ET_0 value for the tomato crop to understand how much water is required for the crop. The annual average solar potential of the study site is around 19.69 $MJ/m^2/day$ and ET_{0} value ranges from 3.7 to 5.7 mm/day. It is found that using the random forest method with reduced input variables, a better performance metrics is achieved with a less mean square error of 0.03. From the obtained results, it is observed that the solar radiation and ET_{0} predictions using the random forest ML model is more precise than the empirical models. Thus, ML algorithms can be attempted to estimate ET_{0} and the actual water requirement for the crops to increase the agricultural yield and to lower the water consumption. The findings of this study can be applied as a successful planning, design, and management strategy for irrigation as well as a solution to the existing challenges in agrometeorological applications.

6. FUTURE SCOPE

The future scope of the research is to investigate microclimate utilizing a mobile weather station attached to a drone for efficient agricultural water management and deep learning-based leaf disease prediction to assist farmers grow their crops with greater efficiency.

7. ACKNOWLEDGMENT

We would like to thank IMD, Pune for meteorological data support.

8. FUNDING

There is no funding to report.

9. CONFLICTS OF INTEREST

The authors report no financial or any other conflicts of interest in this work.

10. ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

11. DATA AVAILABILITY

The data are available on requesting corresponding author.

12. PUBLISHER'S NOTE

This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

REFERENCES

- 1. Wang W, Zhuo L, Li M, Liu Y, Wu P. The effect of development in water-saving irrigation techniques on spatial-temporal variations in crop water footprint and benchmarking. J Hydrol 2019;577:123916.
- 2. Cemek B, Tasan S, Canturk A, Tasan M, Simsek H. Machine learning techniques in estimation of eggplant crop evapotranspiration. Appl Water Sci 2023;13:136.
- 3. Han X, Wei Z, Zhang B, Li Y, Du T, Chen H. Crop evapotranspiration prediction by considering dynamic change of crop coefficient and the precipitation effect in back-propagation neural network model. J Hydrol 2021;596:126104.
- 4. Wang Y, Yin H, Guo X, Zhang W, Li Q. Distributed ANN-bi level two-stage stochastic fuzzy possibilistic programming with Bayesian model for irrigation scheduling management. J Hydrol 2022;606:127435.
- 5. Zhuo L, Li M, Zhang G, Mekonnen MM, Hoekstra AY, Wada Y, *et al*. Volume versus value of crop-related water footprints and virtual water flows: A case study for the Yellow River Basin. J Hydrol 2022;608:127674.
- 6. Zhang C, Li X, Guo P, Huo Z, Huang G. Enhancing irrigation water productivity and controlling salinity under uncertainty: A full fuzzy dependent linear fractional programming approach. J Hydrol 2022;606:127428.
- 7. Xiang Z, Bailey RT, Kisekka I. Using DSSAT-MODFLOW to determine the controls of groundwater storage and crop yield in groundwater-based irrigated regions. J Hydrol 2022;612:128161.
- 8. Han X, Zhou Q, Zhang B, Che Z, Wei Z, Qiu R, *et al*. Real-time methods for short and medium-term evapotranspiration forecasting using dynamic crop coefficient and historical threshold. J Hydrol 2022;606:27414.
- 9. Meenal R, Boazina PG, Selvakumar AI. Temperature based radiation models for the estimation of global solar radiation at horizontal surface in India. Indian J Sci Technol 2016;9:1-6.
- Luo J, Dou X, Ma M. Evaluation of empirical and machine learning approaches for estimating monthly reference evapotranspiration with limited meteorological data in the Jialing River Basin, China. Int J Environ Res Public Health 2022;19:13127.
- 11. Meenal R,·Binu D, Ramya KC, Michael PA, Kumar KV, Rajasekaran E, *et al*. Weather forecasting for renewable energy system: A review. Arch Comput Methods Eng 2022;29:2875-91.
- 12. Chia MY, Huang YF, Koo CH. ANN-Based Reference Evapotranspiration Estimation: Effects of Data Normalization and Parameters Selection.

In: Al-Emran M, Al-Sharafi MA, Al-Kabi MN, Shaalan K, editors. Proceedings of International Conference on Emerging Technologies and Intelligent Systems. Lecture Notes in Networks and Systems. Cham: Springer; 2022. p. 322.

- 13. Meenal R, Selvakumar AI. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. Renewable Energy 2018;121:324-43.
- 14. Vanaja A, Yella VR. Evolution of machine learning in biosciences: Abibliometric network analysis. JAppl Biol Biotechnol 2022;10:45-51.
- 15. Akar F, Katipoğlu OM, Yeşilyurt SN, Taş MB. Evaluation of treebased machine learning and deep learning techniques in temperaturebased potential evapotranspiration prediction. Pol J Environ Stud 2023;32:1009-23.
- 16. Katipoğlu OM. Combining discrete wavelet decomposition with soft computing techniques to predict monthly evapotranspiration in semi-arid Hakkâri province, Türkiye. Environ Sci Pollut Res Int 2023;30:44043-66.
- 17. Sarıgöl M, Katipoğlu OM. Estimation of monthly evaporation values using gradient boosting machines and mode decomposition techniques in the Southeast Anatolia Project (GAP) area in Turkey. Acta Geophys 2023.
- 18. Meenal R, Michael PA, Pamela D, Rajasekaran E. Weather prediction using random forest machine learning model. Indones J Electr Eng Comput Sci 2021;22:1208-15.
- 19. Meenal R, Selvakumar AI. Assessment of Solar Energy Potential of Smart Cities of Tamil Nadu using Machine Learning with Big Data. Advances in Big Data and Cloud Computing. Singapore: Springer; 2019. p. 27-36.
- 20. Meenal E, Selvakumar AI, JeyaPrabha SB, Rajasekaran E. Solar mapping of India using support vector machine. J Phys Conf Ser 2018;1142:012010.
- 21. Hembram P, Suibudhi CR, Subudhi R. Water and irrigation requirement for tomato (*Solanum lycopersicum*) crop of North central plateau zone of Odisha. J Bio Innov 2020;9:701-10.
- 22. Dimitriadou S, Nikolakopoulos KG. Artificial neural networks for the prediction of the reference evapotranspiration of the Peloponnese Peninsula, Greece. Water 2022;14:2027.
- 23. Hu Z, Bashir RN, Rehman AU, Iqbal SI, Shahid MM, Xu T. Machine learning based prediction of reference evapotranspiration (ET0) using IoT. IEEE Access 2022;10:70526-40.
- 24. Zereg S, Belouz K. Modeling daily reference evapotranspiration using SVR machine learning algorithm with limited meteorological data in Dar-el-Beidha, Algeria. Acta Geophys 2023.
- 25. Meenal R, Selvakumar AI, Brighta K, Joice SC, Richerd CP. Solar Radiation Resource Assessment using WEKA. In: 2nd IEEE International Conference on Inventive Systems and Control (ICISC); 2018. p. 1038-42.
- 26. Poyen FB, Ghosh A, Kundu P, Hazra S, Sengupta N. Prototype model design of automatic irrigation controller. IEEE Trans Instrum Meas 2021;70:9502217.

How to cite this article:

R.Meenal, Jala PK, Samundeswari R, Rajasekaran E. Crop water management using machine learning-based evapotranspiration estimation. J App Biol Biotech. 2024;12(2):198-203. DOI: 10.7324/JABB.2024.155791