MedCrave
Step into the World of Research

International Journal of Hydrology

Research Article

Open Access

CrossMark

A review on developing machine learning techniques for estimation of reference evapotranspiration using meteorological data

Abstract

Evapotranspiration (ET0) is a combined loss of water from surface of the earth and vegetative surface to atmosphere under the influence of weather parameters. The weather parameters that are influencing ET0 are air temperature, solar radiation, relative humidity and wind speed. Evapotranspiration (ET0) is the major loss of water received by the crop through rainfall and irrigation and also it represents major portion of total water budget of crop. The estimation of accurate value of ET0 helps in irrigation scheduling at different critical stages of crop for maximum crop production and also for hydrological and water resource planning based on available weather data. This review discusses various methods for estimating ET, including traditional empirical approaches and machine learning techniques. Traditional methods, which are computationally simple and data demanding, may fail to accurately estimate ET. Machine learning methods, SVM, ANN and RF increase the precision of ET estimation by extracting features from all weather particularly in data-rich regions. However, model complexity can lead to difficulties in interpretation the performance of these methods may be limited in data-scarce areas. Collectively, future research should aim to improve data quality, optimize model generalizability, and explore methods that integrate physical processes with data-driven models. When selecting an ET estimation method, considerations should include data availability, model adaptability, estimation accuracy requirements, and technical operational complexity to meet the needs of specific research areas and applications.

Keywords: evapotranspiration, penman-monteith (FAO-56), machine learning techniques

Volume 9 Issue 2 - 2025

Bhavya TR, Ananta Vashisth, P Krishnan, Monika Kundu, Shivprasad

ICAR-Indian Agricultural Research Institute, Pusa, New Delhi 110 012, India

Correspondence: Ananta Vashisth, ICAR-Indian Agricultural Research Institute, Pusa, New Delhi 110 012, India

Received: July 29, 2025 | **Published:** August 25, 2025

Introduction

Evapotranspiration (ET) is the combined loss of water from a vegetative surface and a soil surface via transpiration. This mechanism is responsible for the water stored in the atmosphere and is affected by a variety of climatic variables.[1] The most essential factor in irrigation scheduling is the measurement of evapotranspiration. Timely irrigation schedule boosts agricultural output while also boosting farmer income through water conservation. As a result, conserving water resources would improve soil and groundwater quality. Temperature, solar radiation, relative humidity, and wind speed are the main meteorological elements that influence evapotranspiration.

Accurate monitoring and modelling of evapotranspiration processes are essential for advancing irrigation management and land reclamation practices. Crop growth, development, and yield are governed by a complex interplay of factors, including meteorological variables (e.g., temperature, solar radiation, precipitation, and extreme weather events), agronomic management, improved crop genotypes, optimal fertilizer application, tillage operations, and the strategic use of irrigation. Additionally, the influence of climate change introduces further variability, underscoring the importance of adaptive and data-driven approaches to crop production.[2–4] Irrigation scheduling should be synchronized with the crop's dynamic water requirements. These requirements vary substantially throughout the growth stages due to fluctuations in canopy structure and climatic conditions.

On a global scale, nearly 80% of the water designated for agricultural irrigation is lost through the process of evapotranspiration (ET).[5,6] The escalating demand for water is driven by both climate change and rapid population growth. Reports from the United Nations indicate that, over the past century, water usage has increased at twice the rate of global population growth. If current trends persist, it is projected that by 2025, approximately 1.8 billion people will be living in regions facing severe water scarcity (National Geographic Society, 2016). In many areas, the decline in surface water availability and recurring droughts have led to over-extraction of groundwater, with around 10% of the world's food now produced using unsustainable groundwater pumping practices. Accurate estimation of ET is critical for effective water resource management, including budgeting, allocation, and irrigation planning. For growers to schedule irrigation effectively, they must understand the atmospheric demand for water. Several techniques are available for the direct measurement of ET. One widely used method is the lysimeter, which estimates ET by tracking soil moisture fluctuations within a known soil volume planted with vegetation.[6] Although accurate, lysimeters are often costly and labor-intensive to install and maintain. Another approach involves the use of Evaporation pans to assess water loss through evaporation; however, since they do not account for plant transpiration, the values must be corrected using crop-specific coefficients to approximate total evapotranspiration (ET).[7–9]

Alternatively, ET can be measured by quantifying the moisture flux from the plant surface to the atmosphere using highly sensitive instruments that detect variations in meteorological parameters between the crop surface and a reference height above it. While these techniques provide highly accurate ET estimates, they are generally confined to research settings due to their high cost and complexity.[7,10] To make ET estimation more accessible, various predictive models have been developed that estimate ET throughout the crop growth period based on meteorological data. However, a significant challenge in using these models lies in their dependence on detailed weather information, which is not always readily available. This limitation often necessitates the use of simpler models with fewer data

requirements, even though more advanced models may offer greater accuracy.

Reference evapotranspiration (ET_0) represents the amount of water lost through evaporation and transpiration from a standardized crop surface—typically a well-watered, short, clipped grass—under optimal conditions. This metric reflects the maximum atmospheric demand for water on a given day. When adjusted using crop-specific coefficients, ET_0 becomes a measure of the actual water requirement, or consumptive use, for a particular crop. The portion of this requirement not fulfilled by precipitation or soil moisture storage is termed the water deficit, which must be compensated through irrigation to ensure optimal crop performance.[9] Accurate irrigation is essential, as both excess and insufficient watering can negatively affect crop health and yield. The economic significance of precise irrigation is also substantial; for example, a 1 mm ET loss over 1 hectare equals approximately 10 cubic meters (or 2,680 gallons) of water.[7] An overestimation of ET by just 1 mm could lead to unnecessary use of 2,680 gallons of water, incurring additional costs and contributing to groundwater depletion. Efficient irrigation scheduling based on accurate ET estimation can therefore lead to considerable savings in both water and expenses.[7,11]

Due to the complexity of directly measuring ET, it is essential to develop models that can provide reliable estimates for agricultural water management. While lysimeters are a common tool for measuring ET in field conditions, their readings are often limited to small-scale areas and may not accurately represent broader landscapes because of spatial variability and surface heterogeneity.[12]

In addition to the challenges associated with its complex installation and maintenance, the lysimeter method is also unsuitable for use in large areas with heterogeneous vegetation, as it cannot provide a representative measurement for such diverse conditions.[13] The Penman-Monteith FAO 56 method is widely recognized as the international standard for calculating reference evapotranspiration (ET_0). However, its implementation is often constrained by the extensive meteorological data it requires—data that are typically available only at a limited number of weather stations—thus restricting its applicability in many regions.[7]

To enhance the estimation of reference evapotranspiration (ET), researchers have increasingly turned to machine learning, a specialized branch of artificial intelligence. Machine learning algorithms are capable of predicting future outcomes by learning patterns from historical data and training on labeled datasets. Among these, artificial neural networks (ANNs) have been widely applied for evapotranspiration estimation.[14–19] Findings from these studies demonstrate that ANN models outperform traditional approaches, such as regression and empirical equations, in predicting evapotranspiration. Another promising technique is the support vector machine (SVM), a learning algorithm rooted in statistical learning theory and the principle of structural risk minimization. SVM is particularly well-suited for modeling nonlinear systems[20] and has been shown to offer more consistent and accurate performance than ANN under equivalent training conditions.[21]

Several studies have explored the potential of SVM for ET_0 modeling in various regions. For example,[22] applied SVM to estimate ET_0 in central California, USA, while[23] evaluated its performance in a semi-arid highland region of Iran. In another study, SVM was used to estimate daily pan evaporation and compared against ANN models, revealing the superior predictive capabilities of SVM.[24] These investigations consistently suggest that SVM provides more accurate and reliable ET_0 estimates than both ANN and empirical models. Additionally, the Random Forest algorithm has shown high accuracy in estimating root zone soil moisture, further highlighting the value of machine learning approaches in hydrological modeling.[25]

ET_0 is estimated mostly by empirical equation but the use of machine learning is lagging behind for ET_0 estimation. Estimation of ET_0 by machine learning techniques using different weather input combination has not been done in detail for some regions of India due to lack of availability of all weather parameters for evapotranspiration estimation..Among other machine learning models the promising capabilities of SVM, Random Forest (RF), and Artificial Neural Networks (ANN), the application of these machine learning techniques for ET_0 estimation remains relatively limited—especially in semi-arid regions where meteorological data are scarce.

Empirical methods for estimation of evapotranspiration

Reference evapotranspiration (ET_0) can be estimated using various methods, ranging from complex energy balance equations[7] to simpler empirical formulas that require limited meteorological inputs.[26,27] The estimation of crop evapotranspiration (ETc) involves both direct measurement techniques and indirect modelling approaches. Accurate estimation of ETc is critical for efficient water resource management in agriculture.[23] When lysimeters are unavailable, the field water balance method is commonly employed to estimate actual crop evapotranspiration (ETa).[28–30] However, direct measurement methods are labor-intensive, costly, and their results are often site-specific, limiting broader applicability.

ET_0 plays a central role in regulating numerous hydrological processes, and its accurate estimation is essential for water planning and management strategies.[31] Quantifying ET typically begins with determining reference evapotranspiration (ET_0).[32] Numerous empirical models have been developed for this purpose using meteorological data (Table 1). Among these, the FAO Penman-Monteith method has been endorsed by the Food and Agriculture Organization (FAO) of the United Nations as the global standard for ET_0 estimation. This method defines the reference crop as a hypothetical grass surface with a height of 0.12 m, a surface resistance of 70 s/m, and an albedo of 0.23—representing a well-watered, actively growing, uniform grass cover.[7] Compared to earlier models, such as the FAO-24 Penman method, the FAO-56 Penman-Monteith formulation provides results that are more consistent with observed crop water use worldwide.[33]

The scientific community widely accepts the Penman-Monteith FAO-56 method due to its reliability and accuracy across diverse climatic regions.[34,35] In contrast, calibration studies have shown that the Hargreaves method tends to significantly overestimate ET_0 values when compared to the Penman-Monteith method—by as much as 118–167%—making it less suitable for regional-scale applications.[36,37] Given the critical role of ET_0 in influencing climate change, global temperature trends, crop productivity, water availability, and runoff dynamics, its prediction has been the focus of extensive research in recent decades.[38–40]

Although the Penman–Monteith FAO-56 method is regarded as a physically based model and the most accurate available—being an approximate linear solution to equations governing energy balance, thermodynamics, and vertical transport of heat and water vapor.[38,40] The main limitation lies in its requirement for a comprehensive set of meteorological data, which may not always be available in all regions Table 1.[41,42]

Table 1 Empirical Models for estimation of reference evapotranspiration

S. No	Model type	Model	Equation	References
1	Combination model	Penman-Monteith FAO-56	$ET_0 = \frac{0.408\Delta(R_n - G) + \frac{900}{T}\gamma u2\delta e}{\Delta + \gamma(1 + 0.34u_2)}$	2
2	Temperature based	Hargreaves and Samani	ET0=0.0023*Ra*√(Tmax-Tmin)*(Tmean+17.8)/ λ	21
3		Ritchie	ET= α 1[3.87*10-3 *Rs(0.6Tmax+0.4Tmin+29)]	27

Machine learning techniques

Machine learning methods are often known as data-driven methods. It is a subset of computer science that is classified as an artificial intelligence technique. It interferes with the perfection of strategies that enable the computer to acquire. Simply put, algorithms evolve so that the computer may learn, complete jobs, and perform actions. Over time, numerous approaches for machine learning jobs have been created. It can be used in a wide range of applications, and this technique allows models to address challenges that are difficult to represent mathematically.

Computer software for machine learning

Machine learning methods can be performed in MATLAB, PYTHON, R software, but R is one of the most preferred. It makes statistical computing and graphical representations very easy in R. Advance statistical and machine learning packages are provided in R software along with various other packages and in built functions which greatly simplify statistical analysis. It supports efficient data handling and storage, even for large datasets, and provides versatile plotting capabilities tailored to various analytical needs. R is particularly valuable for tasks such as predictive analytics, data preprocessing, statistical modeling, data visualization, and model deployment. The steps followed is represented in Figure 1.[43]

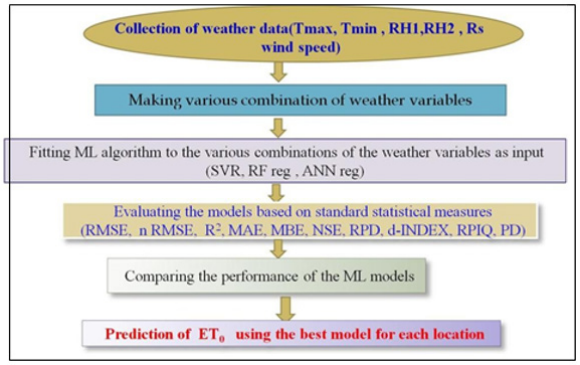

Figure 1 Flowchart for developing ET_0 estimation model.

Support vector machine

The Support Vector Machine (SVM) was first introduced in 1992 and gained widespread recognition through the work of.[44] Support Vector Machine are supervised learning techniques used in classification and regression. They fall within the category of generalised linear classifiers, or, to use more modern terminology, regression prediction and classification methods that apply machine learning theory to take advantage of analytical precision while naturally avoiding over-fitting to the data Figure 2. Regression and grouping problems are equally amenable to the SVM.[45] Support Vector Machine (SVM), a modern learning algorithm based on statistical learning theory and the principle of structural risk minimization, can be effectively used for modeling nonlinear systems.[46]

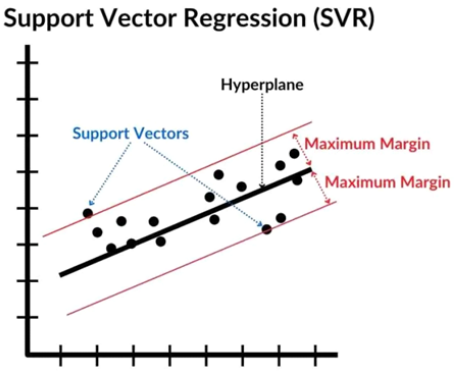

Figure 2 Schematic representation of support vector machine.

Artificial neural network

Artificial neural networks (ANN) are being employed extensively these days because to their ease of usage in resolving complicated and challenging interactions. ANN is used in numerous scientific domains. In order to get good results, this method is also employed in hydrology and hydraulics, among other scientific disciplines. Researchers have calculated the application of AI techniques to forecast hydrological phenomena like ET0 or evaporation in recent years.[47–49] An Artificial Neural Network (ANN) is composed of interconnected layers—input, hidden, and output—each consisting of an array of artificial neurons as shown in Figure 3. A completely connected neural network is one in which every neurone in a layer is connected to every other neurone in the layer above or below it. A mathematical model with parts that are similar to those of an actual neurone is called an artificial neurone. Over the past decade, artificial neural networks (ANNs) have attracted significant attention for their application in estimating evapotranspiration.

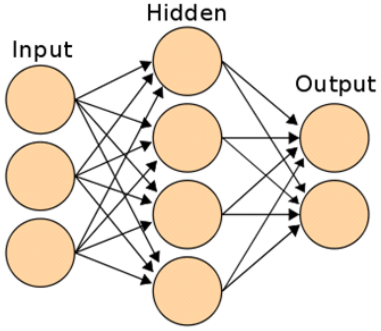

Figure 3 Diagram showing input, hidden and output layers of an ANN.

Citation: Bhavya TR, Vashisth A, Krishnan P, et al. A review on developing machine learning techniques for estimation of reference evapotranspiration using meteorological data. *Int J Hydro.* 2025;9(2):85–93. DOI: 10.15406/ijh.2025.09.00407

Random forest

RF, also referred to as bagging, is an ensemble technique that can perform both regression and classification. It is considered the most advantageous algorithms for forecasting. The core concept behind the Random Forest (RF) algorithm is to combine the outputs of multiple decision trees, rather than relying on a single tree for prediction Figure 4. During training, multiple decision trees are constructed, and the final output is determined by aggregating their results—using the mean prediction for regression tasks or the mode of the predicted classes for classification .This ensemble approach mitigates the tendency of individual decision trees to overfit the training data. In the case of Extremely Randomized Trees (Extra Trees), each decision tree is built from the initial training sample, and at each split node, the tree selects the optimal feature from a randomly chosen subset of k features based on a mathematical criterion, typically the Gini Index.

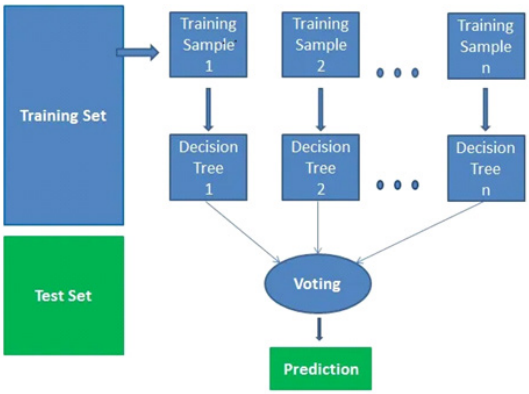

Figure 4 Schematic representation of Random Forest.

The model can be mathematically expressed as:

$$f(x) = f_0(x) + f_1(x) + f_2(x) + ...,$$

where the final model f(x) is the sum of several simpler base models $f_i(x)$, each representing an individual decision tree regressor.

Case study

This study utilizes daily meteorological data from three automated weather stations—Windsor (38°31′35″N, 122°49′42″W), Oakville (38°26′02″N, 122°24′35″W), and Santa Rosa (38°24′04″N, 122°47′56″W)—operated by the California Irrigation Management Information System (CIMIS). The Santa Rosa and Windsor stations are located in Sonoma County, within the North Coast Valleys Region of California, USA, while the Oakville station is situated in Napa County, also within the same region.

The CIMIS program was established in 1982 through a collaboration between the University of California at Davis and the California Department of Water Resources, with the objective of supporting irrigators in managing water use efficiently. Today, CIMIS oversees a network of over 120 automated weather stations across the state.

The meteorological data used in this study were collected from CIMIS stations and include measurements of key variables. Solar radiation is measured using pyranometers installed at a height of 2.0 meters above ground level. Air temperature is recorded using thermistors located at 1.5 meters, and relative humidity sensors are co-located in the same enclosure at the same height. Wind speed is measured by three-cup anemometers positioned at 2.0 meters above the ground anemometers are employed to measure wind speed. The daily meteorological data and corresponding ET_0 values, calculated using the CIMIS Penman method, are obtained from the CIMIS website (http://www.ipm.ucdavis.edu/WEATHER/wxretrieve.html). Seasonal variations in the surrounding environments of the Windsor and Santa Rosa stations may occur, potentially influencing the recorded data. This study utilizes ten years of meteorological data (1998–2007), including daily observations of wind speed (U_2), relative humidity (RH), air temperature (T), and solar radiation (R_s).

For model development, Support Vector Regression (SVR) models are calibrated using data from the first seven years (1998–2004), while the remaining three years (2005–2007) are used for validation. In Table 3, the statistical parameters $\bar{x}$ (mean), S_x (standard deviation), C_v (coefficient of variation), Cs_x (skewness), x_{min} (minimum), and x_{max} (maximum) are presented. The wind speed data for each station exhibit a skewed distribution, as indicated by the Cs_x values. Correlation coefficients in Table 3 reveal that solar radiation (R_s) has the strongest relationship with ET_0 across all stations, followed by air temperature (T), which also shows a significant correlation with reference evapotranspiration.

Table 2 Abbreviation used

Notation	Description	Unit
ET0	Reference evapotranspiration	mm day-1
Rn	net radiation	MJ m-2 day-1
G	soil heat flux density	MJ m-2 day-1
T	mean daily air temperature	°C
u2	wind speed	m $^{s-1}$
es	saturation vapour pressure	kPa
ea	actual vapour pressure	kPa
es-ea	saturation vapour pressure deficit	kPa
Δ	slope vapour pressure curve	kPa °C-1
γ	psychrometric constant	kPa °C-1
λ	Latent heat of vaporization=2.45	MJ/kg
Ra	extraterrestrial radiation	mm/d
Tmax	daily maximum temperature	°C
Tmin	daily minimum temperature	°C

Table 3 Daily statistical parameters of each data set

Station	Variable	xmean	Sx	Cv (Sx/xmean)	Csx	xmin	xmax	Correlation with ET0
	Rs	400	203	0.51	–0.05	0	917	0.958
	T	13.6	4.7	0.34	0.01	0.31	29.6	0.788
Windsor	RH	72.9	11.3	0.15	–0.14	14.5	100	–0.757
	U2	1.5	0.53	0.35	1.45	0.45	5.5	0.259
	ET0	3.31	2.13	0.64	0.2	–0.36	11.2	0
Oakville	Rs	417	200	0.49	-0.11	0	920	0.955

Table 3 Continued....

	T	14.8	4.94	0.33	0.06	1.11	31.6	0.814
	RH	71	12.2	0.17	-0.43	15.5	100	-0.715
	U2	1.61	0.5	0.3	1.1	0.54	4.83	0.35
	ET0	3.6	2.19	0.61	0.1	-0.33	10.5	1
	Rs	12.7	205	0.5	-0.04	0	1155	0.945
	T	410	4.48	0.35	-0.06	-2.7	28.6	0.755
Santa Rosa	RH	77	10.3	0.13	-0.28	19.5	100	-0.711
	U2	1.69	0.68	0.4	6.94	0.45	19.4	0.207
	ET0	3.18	2.02	0.63	0.19	-0.34	9.67	1

Results

Initially, reference evapotranspiration (ET_0) values for the Windsor, Oakville, and Santa Rosa stations are calculated using the Penman-Monteith FAO-56 method.[7] These ET_0 values, along with the corresponding meteorological inputs—air temperature (T), solar radiation (R_s), relative humidity (RH), and wind speed at 2 meters (U_2)—are used to calibrate the Support Vector Regression (SVR) models.

The performance of the SVR models is evaluated using three statistical metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), as presented in Table 4. The R^2 value indicates the strength of the linear relationship between observed and predicted variables. RMSE and MAE, on the other hand, provide complementary insights into the predictive accuracy of the model. Specifically, RMSE emphasizes the model's performance in capturing higher ET_0 values, while MAE offers a more general assessment of the model's overall fitting accuracy across the full range of ET_0 values.[50]

Table 4 Performance statistics of the models in the test period

Models	Model Inputs	RMSE (mm/d)	MAE (mm/d)	R^2
Windsor Station				
SVR1 (30, 0.01, 0.3)	Rs ,T, RH and U2	0.138	0.091	0.996
CIMIS Penman	Rs ,T, RH and U2	0.43	0.29	0.968
SVR2 (50, 0.1, 0.3)	Rs ,T	0.378	0.27	0.971
Hargreaves (1985)	Rs ,T	0.426	0.325	0.969
Ritchie (1990)	Rs ,T	0.418	0.321	0.969
Oakville Station				
SVR1 (60, 0.01, 0.3)	Rs ,T, RH and U2	0.114	0.081	0.997
CIMIS Penman	Rs ,T, RH and U2	0.367	0.274	0.975
SVR2 (60, 0.01, 0.08)	Rs ,T	0.445	0.321	0.953
Hargreaves (1985)	Rs ,T	0.488	0.358	0.948
Ritchie (1990)	Rs ,T	0.485	0.351	0.948
Santa Rosa Station				
SVR1 (15, 0.003, 0.05)	Rs ,T, RH and U2	0.155	0.099	0.994
CIMIS Penman	Rs ,T, RH and U2	0.402	0.272	0.958
SVR2 (20, 0.01, 0.07)	Rs ,T	0.493	0.332	0.948
Hargreaves (1985)	Rs ,T	0.466	0.347	0.947
Ritchie (1990)	Rs ,T	0.446	0.341	0.952

The RMSE and MAE are calculated as follows:

RMSE=

Where, Pi is the estimated value, Oi is the observed value and N is the number of observations

MAE=1\n

where yi = observed value, y^i = estimated value, n = sample size

Two different Support Vector Regression (SVR) models were developed in this study. To enable a valid comparison with the two-variable empirical Hargreaves and Ritchie models, an SVR model using only temperature (T) and solar radiation (R_s) as inputs was constructed. Table 5 presents the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) values for each model during the test period. R^2 indicates the strength of the linear relationship between predicted and observed values. RMSE is more sensitive to larger errors, particularly at high ET_0 values, while MAE provides a balanced measure of overall prediction accuracy.[50]

SVR training aims to find a nonlinear function that minimizes a regularized risk function. This involves selecting appropriate values for the penalty parameter (C), the ε-insensitive loss function, and the kernel function parameter (σ), typically by minimizing RMSE. A Fortran 90 program was used for model implementation and to compute the Lagrange multipliers in Equation (7). The optimal SVR model parameters—C, ε, and σ—are listed in Table 2. For instance, SVR1(30, 0.01, 0.3) denotes a model with C = 30, ε = 0.01, and σ = 0.3.The SVR models are evaluated against the CIMIS Penman,[51] Hargreaves,[26] and Ritchie (Jones & Ritchie, 1990) methods. The CIMIS Penman approach, based on a modified Penman equation,[33] incorporates a wind function developed at the University of California, Davis. It uses hourly average meteorological data to estimate hourly ET_0, which is then summed to calculate daily totals. This method aligns with the FAO-56 guidelines.[7]

Table 5 Total estimated evapotranspiration in test period

Models	Total Evapotranspiration(mm)			Relative error (%)		
	Windsor	Oakville	Santa Rosa	Windsor	Oakville	Santa Rosa
Penman-Monteith FAO	2494	2459	2232	-	-	-
SVR1	2517	2433	2250	0.9	-1.1	0.8
CIMIS Penman	2381	2337	2186	-4.5	-5	-2.1
SVR2	2418	2410	2079	-3.1	-2	-6.9
Hargreaves(1985)	2412	2371	2182	-3.3	-3.6	-2.3
Ritchie(1990)	2455	2391	2239	-1.5	-2.8	0.3

Model performance was compared for the Windsor, Oakville, and Santa Rosa stations using RMSE, MAE, and R^2 (Table 5). SVR1 includes four inputs—R_s, T, RH, and U_2—while SVR2 uses only T and R_s, like the Hargreaves and Ritchie models. SVR1 outperforms all other models across all evaluation metrics. The CIMIS Penman method performs better than the simpler empirical models, and among the two-parameter models, SVR2 shows superior accuracy except at the Santa Rosa station. At Santa Rosa, the Ritchie model outperforms SVR2 in terms of RMSE and R^2, and the Hargreaves model shows a lower RMSE than SVR2. The accuracies of Ritchie and Hargreaves are relatively similar. Scatterplot analyses show that SVR1 predictions are closer to FAO-56 ET_0 values than those from other models. In regression plots ($y = a_0x + a_1$), the slope (a_0) for SVR1 is closer to 1 and the R^2 values are higher, indicating better predictive alignment.

The total estimated ET_0, relevant for irrigation planning, also supports SVR1's superiority. For Windsor, SVR1 estimates 2517 mm, just 0.9% above the FAO-56 reference (2494 mm). CIMIS Penman, SVR2, Hargreaves, and Ritchie methods predict 2381, 2418, 2412, and 2455 mm, underestimating by 4.5%, 3.1%, 3.3%, and 1.5%, respectively. Thus, SVR1 is the most accurate, followed by Ritchie.

At Oakville, SVR1 estimates 2433 mm, close to the reference 2459 mm (1.1% underestimation), while CIMIS Penman, SVR2, Hargreaves, and Ritchie estimate 2337, 2410, 2371, and 2390 mm, underestimating by 5%, 2%, 3.6%, and 2.8%, respectively. Again, SVR1 performs best, followed by SVR2.

At Santa Rosa, total ET_0 values predicted by SVR1 and Ritchie are 0.8% and 0.3% below the reference value (2232 mm), whereas CIMIS Penman, SVR2, and Hargreaves overestimate by 2.1%, 6.9%, and 53.5%, respectively. In this case, the Ritchie model slightly outperforms SVR1.

Figure 5 To provide a comparative analysis, a feed-forward Artificial Neural Network (ANN) model was also used for the daily estimation of reference evapotranspiration (ET_0). The conjugate gradient method, known for its higher efficiency and faster convergence than the traditional gradient descent method, was adopted to optimize the ANN weights.[52] Sigmoid activation functions were applied to both hidden and output layers. Since there is no established rule for determining the optimal number of hidden nodes, different network configurations were tested to identify the most effective structure. Training was halted after 250 epochs, as suggested by Kisi O et al.[52]

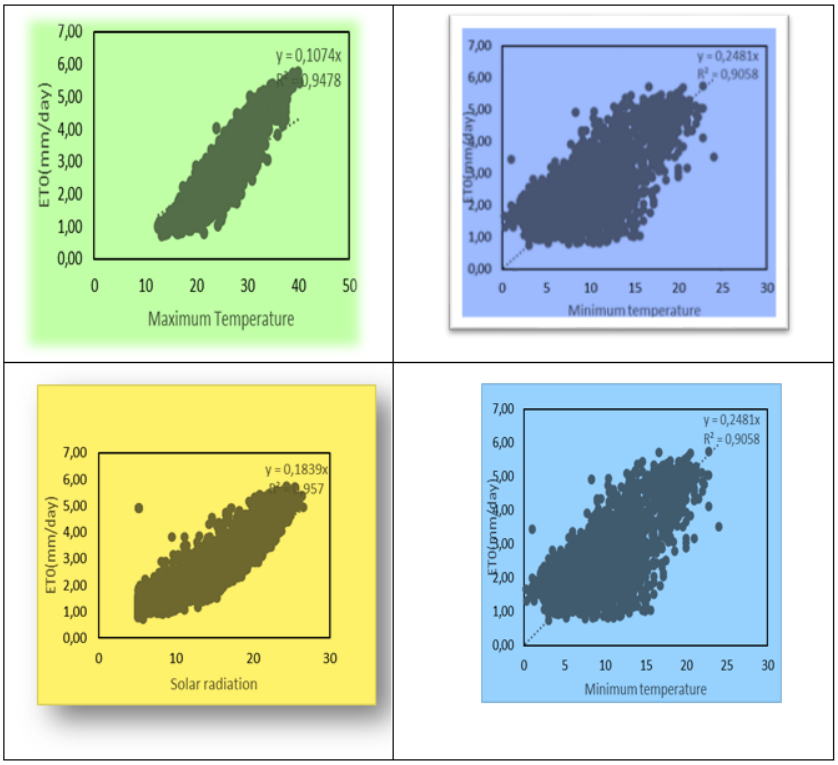

Figure 5 Relationship between evapotranspiration (ET_0) and important weather parameters.

After several trials, the most suitable ANN configurations for the Windsor, Oakville, and Santa Rosa stations were identified and are summarized in Table 4. For example, the structure ANN (4,8,1) refers to a model with four input nodes, eight hidden nodes, and one output node. When comparing the ANN and SVR models using RMSE, MAE, and R^2 metrics, the SVR models generally showed better performance (Table 6). This suggests that SVR is more effective in modeling the relationship between ET_0 and meteorological parameters compared to traditional models like CIMIS Penman, Hargreaves, and Ritchie equation. The CIMIS Penman method combines two terms: a radiative term representing the energy available for evaporation, and an aerodynamic term accounting for the transport of water vapor. Although widely used, the method's primary drawback is its reliance on extensive meteorological data. Despite this, the SVR2 model demonstrated that reliable ET_0 estimation is achievable with fewer and simpler inputs. The Hargreaves and Ritchie equations, derived from empirical relationships based on local meteorological data, can be seen as simplified alternatives to the Penman approach. These models focus mainly on the radiative term, which has a greater influence on ET_0 than the aerodynamic component.[27]

Table 6 Performance statistics of the SVR and ANN models in the test period

Models	Model Inputs	RMSE (mm/d)	MAE (mm/d)	R^2
Windsor Station				
SVR1 (30, 0.01, 0.3)	Rs ,T, RH and U2	0.138	0.091	0.996
ANN(4,8,1)	Rs ,T, RH and U2	0.142	0.103	0.996
SVR2 (50, 0.1, 0.3)	Rs ,T	0.378	0.27	0.971
ANN(2,4,1)	Rs ,T	0.379	0.284	0.969
Oakville Station				
SVR1 (60, 0.01, 0.3)	Rs ,T, RH and U2	0.114	0.081	0.997
ANN(4,4,1)	Rs ,T, RH and U2	0.166	0.122	0.994
SVR2 (60, 0.01, 0.08)	Rs ,T	0.445	0.321	0.953
ANN(2,7,1)	Rs ,T	0.45	0.347	0.952
Santa Rosa Station				
SVR1 (15, 0.003, 0.05)	Rs ,T, RH and U2	0.155	0.099	0.994
ANN(4,6,1)	Rs ,T, RH and U2	0.202	0.155	0.99
SVR2 (20, 0.01, 0.07)	Rs ,T	0.493	0.332	0.948
ANN(2,2,1)	Rs ,T	0.536	0.42	0.931

SVR's main strengths are its ability to model nonlinear relationships and its flexibility. However, it also has limitations—most notably, it is a data-driven "black-box" model without a physical basis and requires sufficient training data for effective implementation. In contrast, empirical models like Hargreaves and Ritchie are easier to interpret and apply. Nonetheless, SVR models generally provide more accurate ET_0 predictions than empirical approaches.

In hydrological modeling studies where direct estimates of evapotranspiration are unavailable, the results of this study can be highly valuable. SVR models can be integrated as a component within traditional hydrological analysis frameworks to enhance their predictive capabilities.

Summary and conclusion

Accurate estimation of reference evapotranspiration (ET_0) is essential for efficient irrigation scheduling and effective water resource management. In this study, ET_0 was calculated using three empirical methods, and the results showed that the Penman–Monteith FAO-56 method outperformed both the Ritchie and Hargreaves–Samani approaches. When compared with lysimeter-based ET measurements, the Penman–Monteith FAO-56 produced results that were more closely aligned.[7] This method is widely acknowledged as the most accurate for estimating ET_0 in various climatic regions around the world.[52,13,34]

Numerous researchers have validated the reliability of the Penman–Monteith (PM) method for ET_0 estimation.[53,52,31] A comparative analysis of 20 different models against lysimeter data from 11 stations across different climatic zones confirmed that the Penman–Monteith FAO-56 consistently provided the most accurate estimates.[54] Despite its strong performance and widespread use,[37] one of the main limitations of the PM FAO-56 method is its requirement for a comprehensive set of meteorological inputs.[40]

To address this issue, machine learning techniques have been explored as alternatives. For example, a study using SVM, M5P, and RF models to estimate monthly ET_0 at two locations (Nagina and Pantnagar) based on weather parameters (temperature, wind speed, solar radiation, and relative humidity) found SVM to offer better performance than empirical models.[55] Similarly, SVM has been shown to outperform ANN under similar training conditions,[21] and in the estimation of daily pan evaporation, SVM models also exceeded ANN models in accuracy.[24]

In the Kosice City region of Slovakia, daily ET_0 was estimated using MLP (multilayer perceptron), SVR (support vector regression), and MLR (multiple linear regression) with inputs including wind speed, relative humidity, temperature, and solar radiation. These machine learning models outperformed traditional empirical methods such as Hargreaves–Samani, Ritchie, and Turc.[56] Another study demonstrated that SVM models using Rs, T, RH, and wind speed provided more reliable ET_0 estimates compared to empirical equations like Penman, Hargreaves, and Priestley–Taylor.[22]

In a semi-arid region of Iran, various models including SVM, ANFIS (adaptive neuro-fuzzy inference system), MLR, and MNLR were evaluated against the Penman–Monteith FAO-56 method using input parameters such as Tmax, Tmin, RH, wind speed, and Rs. The findings showed that both SVM and ANFIS models, particularly those using four input variables (Tmean, RH, Rs, and wind speed), delivered higher accuracy.[23] Similarly, in China's Xinjiang region, Least Squares SVM models using temperature, radiation, humidity, and wind speed achieved strong accuracy for ET_0 estimation.[57] Another study conducted in the arid Ejina Basin of China found that SVM models using Tmax, Tmin, U2, and Rs outperformed both ANN and empirical methods in estimating daily ET_0.[6]

In conclusion, machine learning models—especially SVM—demonstrate significant potential for estimating ET_0 using fewer and simpler input variables than traditional empirical methods. These models can play a vital role in improving irrigation scheduling, managing water resources more efficiently, and determining crop water requirements based on local weather conditions.

In the future, machine learning techniques can enable the estimation of ET_0 using fewer input variables than traditional methods. This approach can support more efficient irrigation scheduling at critical crop growth stages, improve water resource management, and help determine the optimal water requirements for crops based on available weather data.[58]

Acknowledgments

None.

Conflicts of interest

The author declares there is no conflict of interest.

References

1. Traore S, Wang YM, Kerh T. Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. *Agricultural Water Management*. 2010;97(5):707–714.
2. Martínez-Fernández J, González-Zamora A, Sánchez N, et al. A soil water-based index as a suitable agricultural drought indicator. *J Hydrol.* 2015;522:265–273.
3. Nagai T, Makino A. Differences between rice and wheat in temperature responses of photosynthesis and plant growth. *Plant Cell Physiology*. 2009;50(4):744–755.
4. Osborne T, Rose G, Wheeler T. Variation in the global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. *Agricultural and Forest Meteorology*. 2013;170:183–194.
5. Liu X, Mei X, Li Y, et al. Calibration of the Ångström–Prescott coefficients (a, b) under different time scales and their impacts in estimating global solar radiation in the Yellow River basin. *Agric For Meteorol.* 2009;149(3-4):697–710.
6. Xu T, Bateni SM, Margulis SA, et al. Partitioning evapotranspiration into soil evaporation and canopy transpiration via a two source variational data assimilation system. *Journal of Hydrometeorology*. 2016;17(9):2353–2370.
7. Allen RG, Pereira LS, Raes D, et al. Crop evapotranspiration: Guidelines for computing crop water requirements. FAO irrigation and dainage paper 56. *Rome: FAO*. 1998:300(9):D05109.
8. Barnett N, Madramootoo CA, Perrone J. Performance of some evapotranspiration. *Can Agric Eng.* 1998;40(2):89.
9. Dingman SL. *Physical Hydrology*. Englewood Cliffs, NJ: Prentice Hall; 1994.
10. Geiger R, Aron RH, Todhunter P. *The Climate Near the Ground.* 6th ed. Maryland, USA: Rowman & Littlefield Publishers; 2003.
11. Hansen VE, Israelsen OW, Stringham GE. *Irrigation Principles and Practices.* 1980;10(4):18–35.
12. Verstraeten WW, Veroustraete F, Feyen J. Assessment of evapotranspiration and soil moisture content across different scales of observation. *Sensors*. 2008;8(1):70–117.
13. García-Navarro MC, Evans RY, Montserrat RS. Estimation of relative water use among ornamental landscape species. *Sci Hortic.* 2004;99(2):163–174.
14. Chauhan S, Shrivastava RK. Performance evaluation of reference evapotranspiration estimation using climate-based methods and artificial neural networks. *Water Resour Manag.* 2009;23(5):825–837.
15. Citakoglu H, Cobaner M, Haktanir T, et al. Estimation of monthly mean reference evapotranspiration in Turkey. *Water Resour Manag.* 2014;28(1):99–113.
16. El-Shafie A, Najah A, Alsulami HM, et al. Optimized neural network prediction model for potential evapotranspiration utilizing ensemble procedure. *Water Resour Manag.* 2014;28(4):947–967.
17. Kisi O, Cengiz TM. Fuzzy genetic approach for estimating reference evapotranspiration of Turkey: Mediterranean region. *Water Resour Manag.* 2013;27(10):3541–3553.
18. Laaboudi A, Mouhouche B, Draoui B. Neural network approach to reference evapotranspiration modeling from limited climatic data in arid regions. *Int J Biometeorol.* 2012;56(5):831–841.
19. Rahimikhoob A. Comparison between M5 model tree and neural networks for estimating reference evapotranspiration in an arid environment. *Water Resources Management*. 2014;28(3):657–669.
20. Vapnik VN. *The Nature of Statistical Learning Theory*. Springer, New York. 1995.
21. He Z, Wen X, Liu H, et al. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J Hydrol.* 2014;509:379–386.
22. Kisi O, Cimen M. Evapotranspiration modelling using support vector machines. *Hydrol Sci J.* 2009;54(5):918–928.
23. Tabari H, Kisi O, Ezani A, et al. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology*. 2012;444–445:78–89.
24. Lin GF, Lin HY, Wu MC. Development of a support-vector-machine-based model for daily pan evaporation estimation. *Hydrol Process.* 2013;27(22):3115–3127.
25. Carranza C, Nolet C, Pezij M, et al. Root zone soil moisture estimation with Random Forest. *J Hydrol.* 2021;593:125840.
26. Hargreaves GH, Samani ZA. Reference crop evapotranspiration from temperature. *Appl Eng Agric.* 1985;1(2):96–99.
27. Jensen ME, Burman RD, Allen RG. Evapotranspiration and irrigation water requirements. *ASCE Manual No. 70.* New York: ASCE; 1990:360.
28. Bandyopadhyay PK, Mallick S. Actual evapotranspiration and crop coefficients of wheat (Triticum aestivum) under varying moisture levels of humid tropical canal command area. *Agric Water Manag.* 2003;59(1):33–47.
29. Kar G, Kumar A, Martha M. Water use efficiency and crop coefficients of dry season oilseed crops. *Agric Water Manag.* 2007;87(1):73–82.
30. Prihar SS, Sandhu BS. *Irrigation of field crops: Principles and practices*. Indian Council of Agricultural Research, New Delhi. 1987;42 pp.
31. Souza FD, Yoder RE. *ET estimation in the Northeast of Brazil: Hargreaves or Penman-Monteith equation*. 1994.
32. Lopez-Urrea R, Martín de Santa Olalla F, Fabeiro C, et al. Testing evapotranspiration equations using lysimeter observations in a semiarid climate. *Agric Water Manag.* 2006;85:15–26.
33. Pruitt WO, Doorenbos J. *Empirical calibration, a requisite for evapotranspiration formula based on daily or longer mean climatic data*. In: Proc. Int. Round Table Conference on Evapotranspiration. International Commission on Irrigation and Drainage, Budapest, Hungary. 1977.
34. Gavilán P, Lorite IJ, Tornero S, et al. Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment. *Agric Water Manag.* 2006;81(3):257–281.
35. Khoob AR. Artificial neural network estimation of reference evapotranspiration from pan evaporation in a semi-arid environment. *Irrig Sci.* 2008;27(1):35–39.
36. Kisi O. Evapotranspiration estimation using feed-forward neural networks. *Hydrol Res.* 2006;37(3):247–260.

37. Granata F. Evapotranspiration evaluation models based on machine learning algorithms: A comparative study. *Agric Water Manag.* 2019;217:303–315.

38. Jing W, Yaseen ZM, Shahid S. Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions. *Eng Appl Comput Fluid Mech.* 2019;13(1):811–823.

39. Milly PCD, Dunne KA. Potential evapotranspiration and continental drying. *Nat Clim Chang.* 2016;6(10):946–949.

40. Karimi S, Shiri J, Marti P. Supplanting missing climatic inputs in classical and random forest models for estimating reference evapotranspiration in humid coastal areas of Iran. *Comput Electron Agric.* 2020;176:105633.

41. Saggi MK, Jain S. Application of fuzzy-genetic and regularization random forest (FG-RRF): estimation of crop evapotranspiration (ET) for maize and wheat crops. *Agricultural Water Management*. 2020;229:105907.

42. Team RCR: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2013.

43. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: *Proc 5th Annu Workshop on Computational Learning Theory*. 1992:144–152.

44. Bray M, Han D. Identification of support vector machines for runoff modelling. *J Hydroinform.* 2004;6(4):265–280.

45. Aytek A, Guven A, Yuce MI, et al. An explicit neural network formulation for evapotranspiration. *Hydrol Sci J.* 2008;53(4):893–904.

46. Feng Y, Cui N, Zhao L, et al. Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southwest China. *J Hydrol.* 2016;536:376–383.

47. Partal T. Comparison of wavelet based hybrid models for daily evapotranspiration estimation using meteorological data. *KSCE Journal of Civil Engineering*. 2016;20(5):2050–2058.

48. Karunanithi N, Grenney WJ, Whitley D, et al. Neural networks for river flow prediction. *J Comput Civil Eng.* 1994;8(2):201–220.

49. Snyder R, Pruitt W. *Estimating reference evapotranspiration with hourly data*. Ch. VII in: California Irrigation Management Information System Final Report (ed. by R. Snyder et al.). Land, Air and Water Resources Paper #10013, University of California-Davis, USA. 1985.

50. Kisi O, Uncuoglu E. Comparison of three back propagation training algorithms for two case studies. *Indian J Eng Mater Sci.* 2005;12:443–450.

51. Chiew FHS, Kamaladasa NN, Malano HM, et al. Penman-Monteith FAO-24 reference crop evapotranspiration and class-a pan data in Australia. *Agric Water Manag.* 1995;28(1):9–21.

52. Allen RG, Jensen ME, Wright JL, et al. Operational estimates of reference evapotranspiration. *Agron J.* 1989;81(4):650–662.

53. Jensen ME, Wright JL, Pratt BJ. Estimating soil moisture depletion from climate, crop and soil data. *Trans ASAE.* 1971;14(5):954–959.

54. Seong C, Sridhar V, Billah MM. Implications of potential evapotranspiration methods for streamflow estimations under changing climatic conditions. *International Journal of Climatology*. 2018;38(2):896–914.

55. Kaya YZ, Zelenakova M, Üneş F, et al. Estimation of daily evapotranspiration in Košice City (Slovakia) using several soft computing techniques. *Theor Appl Climatol.* 2021;144(1):287–298.

56. Chen X, Ke S, Wang L, et al. Classification of rice appearance quality based on LS-SVM using machine vision. In: *Int Conf Information Computing and Applications*. 2012:104–109.

57. Tyagi NK, Sharma DK, Luthra SK. Evapotranspiration and crop coefficients of wheat and sorghum. *Journal of Irrigation and Drainage Engineering*. 2000;126(4):215–222.

58. Watson I, Burnett AD. *Hydrogeology. An environmental approach*: CRC Press, Boca Raton, Florida. 1995. p. 702.