Comparative Analysis of Shallow Regression Models for Evopotranspiration Estimation with Climate Variables

Gokhan Yuksek
Department of Electrical andElectronics Engineering
Batman University
Batman, Turkey
gokhan.yuksek@batman.edu.tr

Rıdvan Fırat Çınar
Department of Computer Engineering
Batman University
Batman, Turkey
ridvanfirat.cinar@batman.edu.tr

Timur Lale
Department of Electrical and Electronics Engineering
Batman University
Batman, Turkey
timur.lale@batman.edu.tr

Serdar Ekinci
Department of Computer Engineering
Batman University
Batman, Turkey
ekinciser@yahoo.com

Davut Izci
Department of Electrical and Electronics Engineering
Bursa Uludag University
Bursa, Turkey
davutizci@gmail.com

2025 Innovations in Intelligent Systems and Applications Conference (ASYU) | 979-8-3315-9727-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ASYU67174.2025.11208404

***Abstract*—Accurate estimation of reference evapotranspiration (ETo) is essential for sustainable agricultural water management, particularly in regions where field-based measurements are limited and remote sensing data are either unavailable or impractical to use. This study investigates the predictive capability of six shallow learning regression algorithms, extreme gradient boosting (XGBoost) , random forest, multi-layer perceptron regressor (MLP), linear regression (LR), support vector regression (SVR), and decision tree (DT), for daily ETo estimation using only five commonly recorded meteorological variables: air temperature, soil temperature, relative humidity, wind speed, and shortwave radiation. The dataset, derived from the Open-Meteo platform for Batman, Turkiye. The models were trained, evaluated and assessed based on R^2, MAE, MSE, and RMSE metrics. Among the tested models, XGBoost achieved the highest accuracy (R^2 = 0.9946,RMSE = 0.0208), followed closely by LR and MLP methods. In contrast, SVR and DT models exhibited lower precision and generalization capability. Correlation analysis and scatter plots confirmed that shortwave radiation, air temperature, and soil temperature were the most influential predictors. The results demonstrate that accurate and interpretable ETo prediction is feasible using a limited set of accessible meteorological inputs and computationally efficient shallow learning models.**

Keywords—Evapotranspiration, Shallow Learning, Penman-Monteith, Meteorological data, Regression models

I. Introduction

Reference evapotranspiration (ETo), which regulates the exchange of water between the earth's surface and the atmosphere, is a decisive process in terms of both the continuity of natural ecosystems and the planning of human activities [1]. Obtaining ETo information in a timely and accurate manner is crucial in various areas, including agricultural irrigation management, flood and drought prediction, climate modelling, and water resource planning. However, the direct measurement of ETo is highly complex, time-consuming, and often limited to local scales. This situation makes it challenging for decision-makers in large areas to access reliable ETo information [2].

To overcome measurement difficulties, physical models based on energy balance and aerodynamic principles have been developed. These models can be scaled to large areas with the support of satellite-based remote sensing data, enabling ETo estimates based on parameters such as surface temperature and vegetation cover density. However, these methods present application challenges due to their complex model structures and dependence on numerous input variables. Additionally, most of these models operate under the assumption of a homogeneous surface, which may call into question their reliability under variable topography and land cover conditions [3], [4].

Although approaches based on physical models provide high accuracy in theory, in practice they encounter difficulties such as obtaining parameters specific to terrain conditions, model calibration, and data compatibility. Variables such as heterogeneous land use, soil properties, and vegetation cover in large areas reduce the generalizability of models and increase uncertainty [5], [6]. Restrictions on access to high-resolution satellite data and difficulties in synchronizing model inputs complicate the integration of physical models into operational use. For all these reasons, simpler, data-driven, and widely applicable alternative methods have been developed for ETo estimation [7], [8].

In recent years, data-driven approaches have emerged as a means of overcoming these limitations, with statistical models and machine learning-based methods becoming widely used in ETo forecasting. Regression-based algorithms can make highly accurate predictions when sufficient data is available, as they eliminate much of the complexity encountered in physical modelling by learning directly from meteorological data. However, these approaches generally use models that require significant computing power, such as deep learning, making them less applicable in regions with limited hardware infrastructure [9],[10],[11]. Additionally, the low interpretability of these models' internal structures makes it difficult for decision-makers in application areas such as agriculture and the environment to understand and confidently use model outputs [12].

Despite the high predictive power offered by deep learning-based methods, the limitations of these models, such as data dependency, hardware requirements, and low explainability, pose a significant obstacle, especially in low-data-density and rural areas. In addition, many studies require a large number of input variables, which reduces the model's applicability not only in theoretical environments but also in practice. However, agricultural lands, rural hydrological regions, and weather stations in developing countries can only regularly provide a limited number of meteorological parameters. In this context, there is a need for methods that can work with fewer but reliable and field-obtainable climatic variables, have low computational costs, and offer high explainability [13], [14].

In this study, shallow learning-based models were developed using five basic meteorological variables (air temperature, soil temperature, humidity, wind speed, and shortwave radiation) that can be measured regularly in the field for ETo estimation. These methods eliminate the data requirements of complex physical models and prevent the high computing power requirements of deep learning methods. Thus, they offer a low-cost and widely applicable estimation approach. The developed models provide a balanced structure in terms of both accuracy and simplicity, demonstrating that ETo can be reliably estimated under different regional conditions.

II. Methods

In this study, five different shallow learning regression algorithms were employed to estimate daily reference evapotranspiration (ETo) based on a minimal set of meteorological inputs. These algorithms included random forest regressor (RF), extreme gradient boosting (XGBoost), support vector regression (SVR), linear regression (LR), and multi-layer perceptron (MLP). All models were trained and tested using meteorological data retrieved from the Open-Meteo platform, corresponding to Batman City, Türkiye, for a continuous one-month period (May 2025) [15]. The data, recorded initially at hourly intervals, were aggregated into daily averages for model development.

The feature set consisted of five fundamental meteorological variables: air temperature (°C), soil temperature (°C), relative humidity (%), wind speed (m/s), and shortwave radiation (W/m²). These variables were selected for their direct physical relevance to the ETo process, representing thermal input, moisture availability, and atmospheric transport capacity. All features were standardized prior to modelling, and the dataset was partitioned into training (80%) and testing (20%) subsets.

The ETo values used as the target variable were obtained directly from the Open-Meteo API, which calculates ETo based on a standardized implementation of the Penman-Monteith equation. This method estimates the ETo rate from a hypothetical reference crop surface (uniform grass of 0.12 m height, thoroughly watered, and actively growing) under standard meteorological conditions [16]. The equation is given as:

$$ETo = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

where ETo is in the daily frequency (mm/day), R_n is the net radiation at the crop surface (MJ/m²/day), G is the soil heat flux (MJ/m²/day), T is the mean air temperature at 2 meters (°C), u_2 is the wind speed (m/s), e_s and e_a are the saturation and actual vapour pressure (kPa), Δ is the slope of the vapour pressure curve (kPa/°C), and γ is the psychrometric constant (kPa/°C). Since the Open-Meteo API internally integrates all necessary physical parameters and interpolated radiation data, manual computation of intermediate terms was not required in this study.

Model performance was evaluated using four standard regression metrics: root mean square error (RMSE), the coefficient of determination (R²), root squared error (RSE), and mean absolute error (MAE) [17]. These metrics quantify different aspects of prediction accuracy, including overall deviation, average absolute deviation, and the model's predictive reliability. The formulas for each metric are given in (2) – (5).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{true,i} - y_{pred,i}\right)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{true,i} - y_{pred,i})^2}{\sum_{i=1}^{n}(y_{true,i} - \bar{y}_{true,i})^2} \quad (3)$$

$$RSE = 1 - R^2 \quad (4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{true,i} - y_{pred,i}\right| \quad (5)$$

A. *Random Forest Regressor*

RF is a learning method that consists of multiple decision trees and makes predictions by combining the results of these trees. Each tree is trained with randomly selected data and feature subsets. This structure reduces the variance of the model and prevents overfitting. RF stands out as a suitable choice for ETo prediction because it can effectively model non-linear relationships between variables [18].

B. *Extreme Gradient Boosting*

XGBoost is a shallow learning technique that focuses on minimizing errors by sequentially constructing decision trees. While offering a structure that prioritizes speed and accuracy, it reduces the risk of overfitting through regularization terms. The model can demonstrate exceptional performance on high-dimensional and complex data sets [19].

C. *Support Vector Regressor*

SVR is a shallow learning method that can model nonlinear regression relationships by converting data into a high-dimensional feature space. The model attempts to minimize errors outside of predictions that remain within a certain error tolerance. This structure is considered an effective method for making general predictions with limited input variables [20].

D. *Linear Regression*

Linear regression is a classic method that models the linear relationship between the dependent variable and one or more independent variables. It has been evaluated as a basic comparison model in order to determine the direct relationships between ETo and climatic variables and to interpret the performance of other methods [21].

E. Multi-Layer Perceptron Regression

MLP is a feedforward artificial neural network structure with at least one hidden layer. It can learn complex and nonlinear relationships between input variables and ETo output. During model training, the backpropagation algorithm is employed, and the ReLU function is preferred as the activation function. This structure can be generalized with a sufficient number of examples and provides high-accuracy predictions [22].

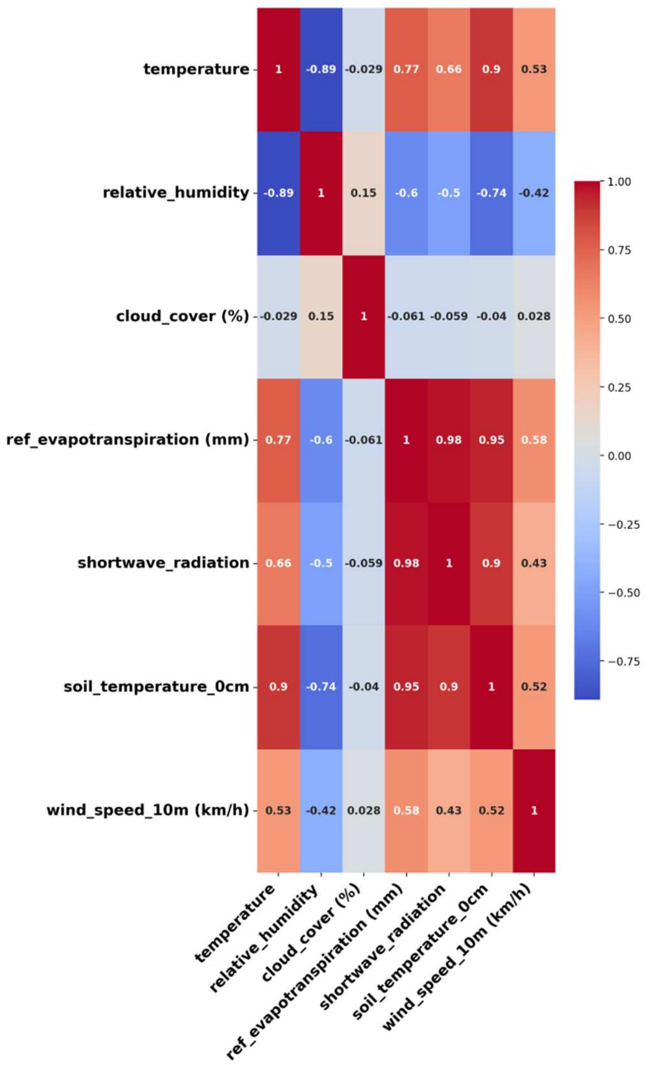

Fig. 1 Correlation analysis of the evapotranspiration dataset

F. Decision tree regressor

The Decision Tree (DT) Regressor is a non-parametric supervised learning algorithm that splits the input space into distinct regions based on feature thresholds in a recursive, tree-like structure. At each node, the algorithm selects a feature and split point that minimizes a given error metric, typically the mean squared error, resulting in a hierarchical partitioning of the data. While DTs are capable of capturing nonlinear relationships and interactions between variables, they are prone to overfitting, especially when the tree depth is not constrained. In this study, the DT is included as a baseline model due to its simplicity, interpretability, and fast training time, although its standalone performance was found to be less robust compared to ensemble or neural network approaches [23].

III. Results and Discussions

The correlation heatmap presented in Fig. 1 reveals the strength and direction of linear relationships between ETo and the selected meteorological variables. It has a dominant role in driving the energy balance and surface vapour fluxes. Soil temperature at 0 cm depth and air temperature also show strong positive correlations $(r = 0.95\, and\, r = 0.77, respectively)$, indicating the importance of thermal dynamics in ETo processes. In contrast, relative humidity displays a notable negative correlation with ETo $(r = -0.60)$, as increased atmospheric moisture reduces the vapour pressure gradient and thereby limits ETo.

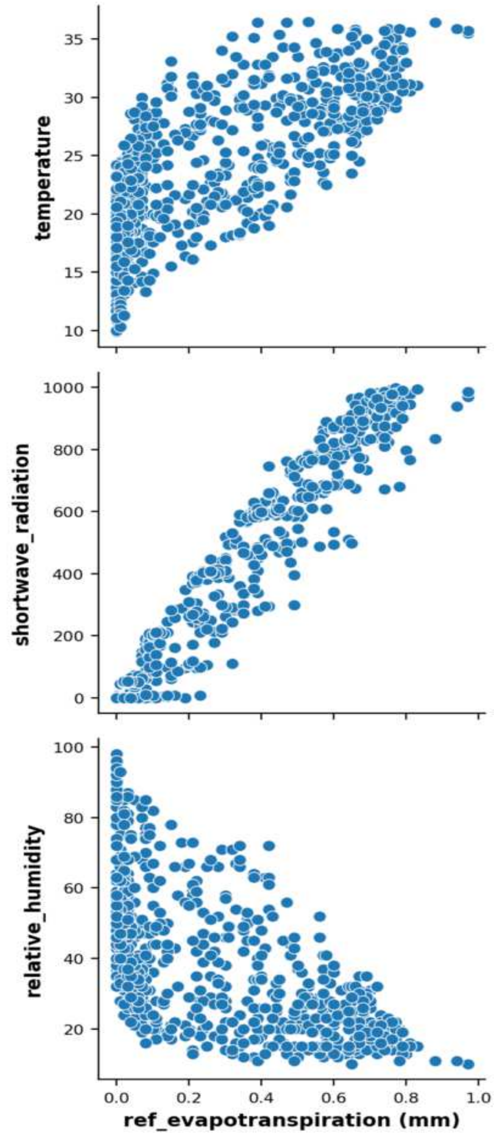

Fig. 2 Pair plot of the most relevant parameters with evapotranspiration

Moderate positive correlations are also observed with wind speed at 10 m $(r = 0.58)$, supporting the role of aerodynamic transport in moisture removal. Cloud cover, on the other hand, has a very weak and slightly negative correlation $(r = -0.061)$, suggesting a relatively minor influence in the context of the short observation period. These results validate the physical relevance of the selected predictors and justify their use as input features in the applied machine learning models.

The scatter plots in Fig. 2 illustrate the pairwise relationships between ETo and three significant meteorological variables: air temperature, shortwave radiation, and relative humidity. A strong and approximately linear positive association is observed between shortwave radiation and ETo, supporting the fundamental role of radiative energy input in driving surface water loss through evaporation and transpiration. Similarly, air temperature exhibits a clear positive relationship with ETo, with higher temperatures corresponding to increased ETo rates, as warmer air holds more energy and has a higher vapour pressure potential.

In contrast, relative humidity demonstrates a distinctly negative nonlinear relationship with ETo, indicating that drier air conditions facilitate higher rates of moisture transfer from the land surface to the atmosphere. These visual trends are consistent with the results of the correlation matrix and provide empirical support for selecting these variables as primary inputs in the proposed predictive models.

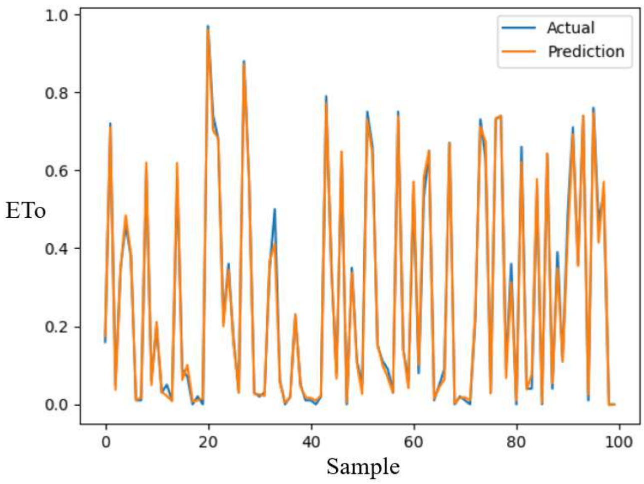

Fig. 3 Prediction plot of XGBoost model

Figs. 3 - 8 illustrate the prediction and regression performance of the six regression models in estimating daily ETo values. In the prediction plots, the comparison between observed ETo values and model-generated estimates is depicted across a randomly selected test segment. Among all models, GB demonstrates the most consistent alignment with the actual data, accurately following both sharp increases and decreases in ETo values, reflecting its ability to generalize well across varying temporal patterns.

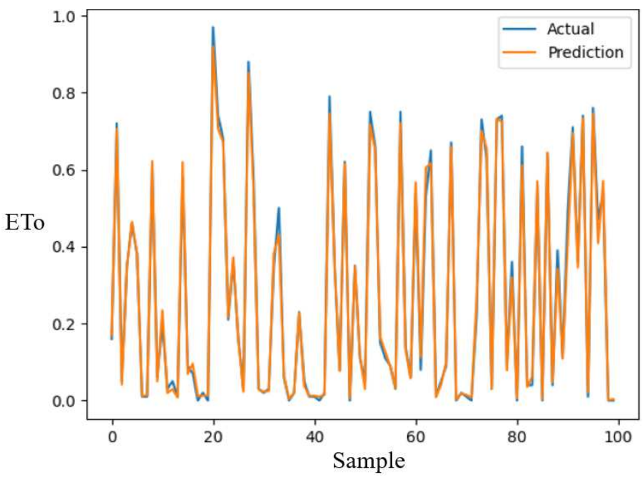

Fig. 4 Prediction plot of the RF model

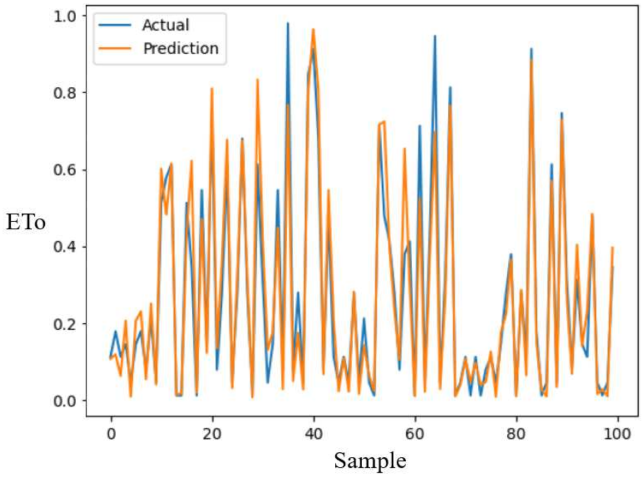

Fig. 5 Prediction plot of the MLP model

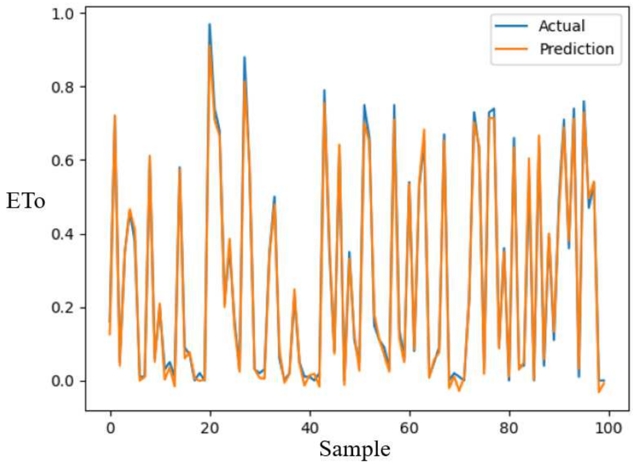

Fig. 6 Prediction plot of LR model

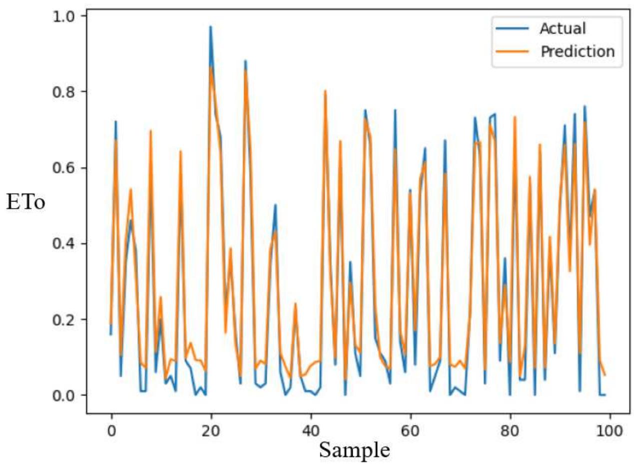

Fig. 7 Prediction plot of the SVM model

RF and MLP also track the actual signal closely, though RF exhibits slightly more fluctuation in regions with abrupt change, and MLP tends to underpredict at peaks slightly. The performance of LR remains reasonably stable but lacks the flexibility to capture nonlinear shifts, resulting in systematic underestimation or overestimation in certain intervals. The SVR model displays higher volatility, especially under low humidity and high radiation conditions, which likely reflects the influence of kernel sensitivity. Finally, DT shows the most significant discrepancies, particularly during rapid transitions, indicating overfitting and reduced robustness.

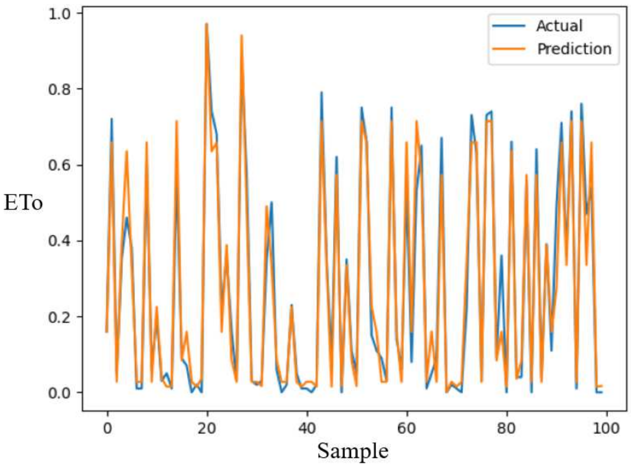

Fig. 8 Prediction plot of DT model

Table 1 summarizes the quantitative performance metrics of the six machine learning models used for ETo estimation. Among the evaluated models, XGBoost yielded the highest overall accuracy, achieving the highest coefficient of determination ($R^2 = 0.9946$) and the lowest error values ($MAE = 0.0145, RMSE = 0.0208$).

TABLE I. PERFORMANCE COMPARISON OF SHALLOW MODELS

Model	R²	MAE	RSE	RMSE
XGBoost	0.9946	0.0145	0.0054	0.0208
LR	0.9942	0.0168	0.0058	0.0215
MLP	0.9942	0.0167	0.0058	0.0216
RF	0.9931	0.0156	0.0069	0.0234
SVR	0.9554	0.0523	0.0446	0.0596
DT	0.9443	0.0447	0.0557	0.0666

LR and MLP Regressor followed closely with nearly identical R^2 values (0.9942), although their error metrics were marginally higher than those of GB. The RF also demonstrated strong predictive capability ($R^2 = 0.9931$), though its $RMSE$ value (0.0234) indicates slightly reduced precision under more variable conditions. On the other hand, SVR and DT models exhibited considerably lower performance. SVR resulted in higher residual errors ($MAE = 0.0523, RMSE = 0.0596$), likely due to its sensitivity to parameter tuning and nonlinear dynamics. DT, with the lowest R^2 value (0.9443) and the highest $RMSE$ (0.0666), exhibited a clear tendency toward overfitting and insufficient generalization, particularly in regions of high variability. These findings support the conclusion that ensemble-based and neural network models provide superior accuracy and robustness for ETo prediction compared to simpler tree-based or kernel-based regressors.

IV. CONCLUSIONS

In this study, the performance of six shallow learning regression models was evaluated to predict daily ETo using an accessible set of meteorological inputs such as air temperature, soil temperature, relative humidity, wind speed, and shortwave radiation. The models were trained and tested on a one-month dataset obtained from the Open-Meteo platform for the city of Batman. Among the tested models, XGBoost achieved the highest prediction accuracy, followed by the LR and MLP Regressor models. These models demonstrated strong agreement with observed ETo values, as evidenced by both statistical metrics and visual analyses. RF also yielded competitive results, albeit with slightly higher variance. In contrast, SVR and DT models performed weaker in regions with high variability. The results demonstrate that high-accuracy ETo predictions can be obtained using shallow learning techniques and a limited number of meteorological features, without relying on complex or data-intensive modeling frameworks. This approach offers a practical, interpretable, and computationally efficient solution for estimating evapotranspiration in regions with limited resources or sparse instrumentation.

REFERENCES

[1] Y. Yang, M. L. Roderick, H. Guo, D. Miralles, L. Zhang, S. Fatichi, X. Luo, Y. Zhang, T. R. McVicar, Z. Tu, T. F. Keenan, J. B. Fisher, R. Gan, X. Zhang, S. Piao, B. Zhang and D. Yang "Evapotranspiration on a greening Earth," *Nat Rev Earth Environ*, vol. 4, no. 9, pp. 626–641, Aug. 2023, doi: 10.1038/s43017-023-00464-3.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] S. S. Anapalli, D. K. Fisher, K. N. Reddy, N. Rajan, and S. R. Pinnamaneni, "Modeling evapotranspiration for irrigation water management in a humid climate," Agric Water Manag, vol. 225, p. 105731, Nov. 2019, doi: 10.1016/j.agwat.2019.105731.

[3] R. Singh and G. Senay, "Comparison of Four Different Energy Balance Models for Estimating Evapotranspiration in the Midwestern United States," Water (Basel), vol. 8, no. 1, p. 9, Dec. 2015, doi: 10.3390/w8010009.

[4] B. N. Aryalekshmi, R. C. Biradar, K. Chandrasekar, and J. Mohammed Ahamed, "Analysis of various surface energy balance models for evapotranspiration estimation using satellite data," The Egyptian Journal of Remote Sensing and Space Science, vol. 24, no. 3, pp. 1119–1126, Dec. 2021, doi: 10.1016/j.ejrs.2021.11.007.

[5] D. Long, L. Longuevergne, and B. R. Scanlon, "Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites," Water Resour Res, vol. 50, no. 2, pp. 1131–1151, Feb. 2014, doi: 10.1002/2013WR014581.

[6] Gloria I. Ezenne, Nkpouto U. Eyibio, Jane L. Tanner, Felix U. Asoiro, And Sunday E. Obalum, "An overview of uncertainties in evapotranspiration estimation techniques," Journal of Agrometeorology, vol. 25, no. 1, Feb. 2023, doi: 10.54386/jam.v25i1.2014.

[7] N. Guo, H. Chen, Q. Han, and T. Wang, "Evaluating data-driven and hybrid modeling of terrestrial actual evapotranspiration based on an automatic machine learning approach," J Hydrol (Amst), vol. 628, p. 130594, Jan. 2024, doi: 10.1016/j.jhydrol.2023.130594.

[8] X. Hu, L. Shi, G. Lin, and L. Lin, "Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation," J Hydrol (Amst), vol. 601, p. 126592, Oct. 2021, doi: 10.1016/j.jhydrol.2021.126592.

[9] J. Dong, Y. Zhu, X. Jia, M. Shao, X. Han, J. Qiao, C. Bai, X. Tang "Nation-scale reference evapotranspiration estimation by using deep learning and classical machine learning models in China," J Hydrol (Amst), vol. 604, p. 127207, Jan. 2022, doi: 10.1016/j.jhydrol.2021.127207.

[10] X. Dou and Y. Yang, "Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems," Comput Electron Agric, vol. 148, pp. 95–106, May 2018, doi: 10.1016/j.compag.2018.03.010.

[11] F. Granata, "Evapotranspiration evaluation models based on machine learning algorithms—A comparative study," Agric Water Manag, vol. 217, pp. 303–315, May 2019, doi: 10.1016/j.agwat.2019.03.015.

[12] H. Zhang, G. Wang, S. Li, and P. Cabral, "Understanding Evapotranspiration Driving Mechanisms in China With Explainable Machine Learning Algorithms," International Journal of Climatology, vol. 45, no. 6, May 2025, doi: 10.1002/joc.8774.

[13] A. R. Troncoso-García, I. S. Brito, A. Troncoso, and F. Martínez-Álvarez, "Explainable hybrid deep learning and Coronavirus Optimization Algorithm for improving evapotranspiration forecasting," Comput Electron Agric, vol. 215, p. 108387, Dec. 2023, doi: 10.1016/j.compag.2023.108387.

[14] S. Khan, H. Wang, U. Nauman, R. Dars, M. W. Boota, and Z. Wu, "Climate Impact on Evapotranspiration in the Yellow River Basin: Interpretable Forecasting with Advanced Time Series Models and

Explainable AI," Remote Sens (Basel), vol. 17, no. 1, p. 115, Jan. 2025, doi: 10.3390/rs17010115.

[15] Open-Meteo, "Open-Meteo: Free Weather API for Non-Commercial Use," 2025.

[16] W. Treder, K. Klamkowski, K. Wójcik, and A. Tryngiel-Gać, "Evapotranspiration Estimation Using Machine Learning Methods," Journal of Horticultural Research, vol. 31, no. 2, pp. 35–44, Dec. 2023, doi: 10.2478/johr-2023-0033.

[17] Y. Xie, C. Li, M. Li, F. Liu, and M. Taukenova, "An overview of deterministic and probabilistic forecasting methods of wind energy," iScience, vol. 26, no. 1, p. 105804, Jan. 2023, doi: 10.1016/j.isci.2022.105804.

[18] S. Shadkani, S. Hashemi, A. Pak, and A. B. Lahijan, "Random Forest and Multilayer Perceptron hybrid models integrated with the genetic algorithm for predicting pan evaporation of target site using a limited set of neighboring reference station data," Earth Sci Inform, vol. 17, no. 2, pp. 1261–1280, Apr. 2024, doi: 10.1007/s12145-024-01237-2.

[19] M. Achite, H. Nasiri, O. M. Katipoğlu, M. Abdallah, R. Moazenzadeh, and B. Mohammadi, "A coupled extreme gradient boosting-MPA approach for estimating daily reference evapotranspiration," Theor Appl Climatol, vol. 156, no. 2, p. 113, Feb. 2025, doi: 10.1007/s00704-024-05313-x.

[20] S. Zereg and K. Belouz, "Modeling daily reference evapotranspiration using SVR machine learning algorithm with limited meteorological data in Dar-el-Beidha, Algeria," Acta Geophysica, vol. 72, no. 3, pp. 2009–2025, May 2023, doi: 10.1007/s11600-023-01107-3.

[21] N. S. Hailegnaw, H. K. Bayabil, M. L. Berihun, F. T. Teshome, V. Shelia, and F. Getachew, "Integrating machine learning and empirical evapotranspiration modeling with DSSAT: Implications for agricultural water management," Science of The Total Environment, vol. 912, p. 169403, Feb. 2024, doi: 10.1016/j.scitotenv.2023.169403.

[22] H. Tabari and P. Hosseinzadeh Talaee, "Multilayer perceptron for reference evapotranspiration estimation in a semiarid region," Neural Comput Appl, vol. 23, no. 2, pp. 341–348, Aug. 2013, doi: 10.1007/s00521-012-0904-7.

[23] J. Rajput, M. Singh, K. Lal, M. Khanna, A. Sarangi, J. Mukherjee and S. Singh, "Data-driven reference evapotranspiration (ET0) estimation: a comparative study of regression and machine learning techniques," Environ Dev Sustain, vol. 26, no. 5, pp. 12679–12706, Oct. 2023, doi: 10.1007/s10668-023-03978-4.