Shahrood University of Technology

Journal of Hydraulic and Water Engineering (JHWE)

Journal homepage: https://jhwe.shahroodut.ac.ir

Iranian Hydraulic Association (IHA)

Comparative Evaluation of Machine Learning Algorithms for Evaporation Estimation in Shahrood Region

Ali Ebrahimi[1*], Abbas Pourdeilami[2], Sina Khoshnevisan[3], Mohammadreza Asli Charandabi[4]

[1*] Ph.D., Civil Engineering, Shahrood University of Technology, Shahrood, Iran
[2] Assistant Professor, School of Engineering, Damghan University, Damghan, Iran
[2] Assistant Professor, School of Engineering, Damghan University, Damghan, Iran
[3] MSc Student, Civil Engineering, Shahrood University of Technology, Shahrood, Iran
[4] Ph.D. Student, Civil Engineering, Shahrood University of Technology, Shahrood, Iran

Article Info

Article history:
Received 11 June 2025
Received in revised form 2 July 2025
Accepted 07 July 2025
Published online 14 July 2025

DOI:10.224/JHWE.2025.16384.1070

Keywords
Evaporation estimation, Machine learning algorithms, Artificial neural network, Shahrood, Meteorological Data

Abstract

Accurate prediction of evaporation is critical for effective water resource management, particularly in arid and semi-arid regions. This research evaluates the performance of five machine learning algorithms — Decision Tree, K-Nearest Neighbors, Support Vector Regression, Random Forest, and Artificial Neural Network — in estimating monthly evaporation rates using meteorological data collected at the Shahrood Synoptic Station from 1992 to April 2025. The dataset includes key climatic parameters, including average temperature, wind speed, precipitation, and relative humidity. Model performance was assessed through four metrics: Mean Absolute Error, Coefficient of Determination, Kling-Gupta Efficiency, and Average Absolute Relative Deviation. Results indicate that the Random Forest model outperformed all others, achieving the lowest MAE of 19.94 mm, the highest KGE of 0.973, and the lowest AARD of 0.521, reflecting superior accuracy and stability. The Artificial Neural Network model also demonstrated strong predictive capability, closely followed by Support Vector Regression. In contrast, simpler models such as Decision Trees and K-Nearest Neighbors performed comparatively poorly because they could not capture complex evaporation dynamics. Temporal analysis revealed that all models effectively captured seasonal evaporation patterns, with Random Forest and Artificial Neural Network most accurately tracing peak and trough fluctuations. The results demonstrate that machine learning models achieve strong predictive accuracy in evaporation estimation and provide a reliable approach for assessing evaporation and water loss.

1. Introduction

Water is a fundamental resource for life, playing a crucial role in sustaining ecosystems, supporting human health, and driving economic activities. Agriculture, in particular, relies heavily on water for irrigation to ensure food production and global food security. Additionally, water holds substantial cultural and social significance, often playing a central role in community traditions and rituals (Wu et al., 2020). However, challenges such as climate change, water pollution, and poor water management practices threaten water availability and quality, emphasizing the need for sustainable water resource management to protect ecosystems and support human well-being. Accurate evaporation estimation is critical for effective water resource management, agricultural planning, and hydrological modeling, especially in arid and semi-arid regions where water scarcity is prevalent (Gelete & Yaseen, 2024).

[*] *Corresponding author: ali.ebrahimi660231@gmail.com, Tel: +989124317506*

Evaporation represents a key component of the hydrologic cycle and has a substantial influence on the design and operation of irrigation systems, reservoir management, and climate studies (Shabani et al., 2020). Traditional empirical formulas often fall short of capturing the nonlinear and complex interactions among meteorological variables influencing evaporation, thereby necessitating robust, intelligent computational models (Deo et al., 2016).

Evaporation estimation has traditionally relied on empirical methods such as the Penman-Monteith, Hargreaves-Samani, and Priestley-Taylor equations. While effective under certain conditions, these models often fail to generalize across diverse climatic regions due to their sensitivity to input parameters and their linearity assumptions. (Dong et al., 2013). Consequently, there has been a shift toward data-driven approaches, particularly machine learning (ML), which can model the complex, nonlinear interactions among meteorological variables influencing evaporation. Support Vector Machines (SVMs) are among the earliest ML methods explored for evaporation modeling. Several studies confirmed the SVMs' superiority over traditional regression techniques due to their robust performance under high-dimensional, nonlinear input conditions. (Deswal & Pal, 2008); (Moghaddamnia et al., 2009); (Yang & Chui, 2021). Their variants, such as Least Squares SVM and ε-SVR, have also been evaluated favorably in multiple climatic contexts. (Tezel & Buyukyildiz, 2016).

Artificial Neural Networks (ANNs) have been widely adopted for their ability to capture nonlinear patterns using input features such as temperature, humidity, wind speed, and solar radiation. These models consistently outperform traditional methods, especially when trained with sufficient data. (Sudheer et al., 2002); (Ali & Saraf, 2015), and newer training algorithms like Bayesian Regularization and Scaled Conjugate Gradient have been proposed to enhance ANN reliability (Aghelpour et al., 2022); (Falkenmark, 1995). Recent advances have led to the integration of ensemble models and hybrid techniques. For instance, Random Forests (RF), Gradient Boosting Machines (GBM), and Quantile Random Forests (QRF) have been shown to deliver competitive or superior accuracy compared to traditional ANN and SVM methods. (Shabani et al., 2020); (Al Sudani & Salem, 2022); (Gelete & Yaseen, 2024). Hybrid and ensemble models, such as extreme learning machines (ELMs), optimized using metaheuristic algorithms, have shown improved predictive accuracy. (Wu et al., 2020); (Ehteram et al., 2024), and recent developments include the use of deep learning approaches such as LSTM and GRU, which are effective in capturing temporal dynamics (Kisi et al., 2022); (Latif, 2024); (Yang & Chui, 2021).

The soil dispersivity parameter (a), which is fundamental for modeling contaminant transport in porous media, is traditionally measured in situ through costly, time-consuming experiments. In this study, three soft computing methods the adaptive neuro-fuzzy inference system (ANFIS), artificial neural network (ANN), and gene expression programming (GEP) were employed to estimate a based on readily measurable physical soil and hydraulic variables: travel distance (L), mean grain size (D_{50}), soil bulk density (qb), and contaminant velocity (Vc). Model performance was evaluated using mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination (R^2). Results indicated that the ANN achieved the best performance with RMSE = 0.00050 m and $R^2 = 0.977$, while the ANFIS (RMSE = 0.00062 m, $R^2 = 0.956$) and GEP reached nearly comparable accuracy. All soft computing approaches significantly outperformed multiple linear regression (MLR), and sensitivity analysis revealed that travel distance (L) had the most significant and bulk density (qb) the least influence on soil dispersity (Emamgholizadeh et al., 2017).

Predicting sesame seed yield with high accuracy is vital for effective breeding

strategies, yet conventional linear models often struggle to capture the underlying nonlinear dynamics of plant traits. In this study, we compared the performance of an artificial neural network (ANN) with a multiple linear regression (MLR) model using readily measurable morpho-phenological variables: days to full flowering, plant height, number of capsules per plant, 1,000-seed weight, and seeds per capsule, based on field trial data. The ANN outperformed the MLR, achieving an RMSE of 0.339 t/ha and an R2 of 0.901, whereas the MLR showed an RMSE of 0.346 t/ha and an R2 of 0.779. Sensitivity analysis further indicated that capsule count per plant was the strongest predictor of yield, while flowering time had the least effect (Emamgholizadeh et al., 2015).

Estimating suspended sediment load in rivers is essential for hydraulic engineering, yet traditional sediment rating curves (SRCs) often exhibit low precision and high uncertainty. Leveraging daily discharge and sediment concentration records from the Kasilian and Telar stations over 1964–2014, this work assessed three AI-driven techniques—gene expression programming (GEP), artificial neural networks (ANN), and adaptive neuro-fuzzy inference system (ANFIS) against the SRC benchmark. The AI models consistently outperformed the SRC, delivering higher coefficients of determination (R^2) and reduced mean absolute errors (MAE), with GEP achieving the top predictive accuracy. These findings underscore the potential of AI, particularly GEP, to markedly enhance suspended sediment load estimation for improved water resources planning and management (Emamgholizadeh & Demneh, 2019).

Extreme Learning Machines (ELMs) have emerged as a promising method for their fast learning speed and simplicity, especially when integrated with optimization algorithms like Flower Pollination Algorithm (FPA) and Whale Optimization Algorithm (WOA), significantly improving predictive accuracy (Wu et al., 2020). Further optimization has been achieved using hybrid metaheuristic strategies like Gaussian Mutation-Alpine Skiing Optimization, which enhance feature selection and temporal-spatial pattern extraction (Ehteram et al., 2024). In parallel, deep learning approaches, particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, have been shown to outperform traditional models due to their ability to capture long-term temporal dependencies in climatic data (Kisi et al., 2022), (Latif, 2024); (Yang & Chui, 2021); (Ercin & Hoekstra, 2014). These models are especially advantageous in regions with strong seasonal patterns or limited availability of high-quality data. Additionally, comparative studies remain crucial for evaluating the relative performance of these diverse methods. Several researchers have benchmarked multiple ML algorithms on identical datasets to identify optimal modeling strategies under different climatic and geographic conditions (Yang & Chui, 2021); (Amer & Farah, 2025); (Hashemi et al., 2018)

One of the main challenges in hydraulic engineering is accurately estimating river suspended sediment load, and the traditional sediment rating curve (SRC) method is limited by low accuracy and high uncertainty. This study compares three artificial intelligence models ,gene expression programming (GEP), artificial neural network (ANN), and adaptive neuro-fuzzy inference system (ANFIS) with the SRC method for estimating daily suspended sediment load at two hydrometric stations in the Casilan (342.9 km^2) and Talar (1,768.6 km^2) watersheds in northern Iran over the 1964–2014 period. The results show that all three AI models outperform the SRC method, achieving higher coefficients of determination (R^2) and lower mean absolute errors (MAE), with GEP yielding the highest R^2 and lowest MAE and therefore the best predictive performance. Thus, the application of AI techniques ,especially GEP ,can be an effective tool for improving the accuracy of suspended sediment load estimation in water resources

planning and management (Emamgholizadeh & Demneh, 2019).

Predicting local scour depth around bridge piers is notoriously tricky due to the combined effects of pier geometry (length Lp, width Wp, attack angle θ), flow conditions (velocity V, depth y), and sediment characteristics (D50, D84). In this work, a multilayer perceptron (MLP) neural network trained on both dimensional and dimensionless datasets via Buckingham's π-theorem was evaluated against multiple linear regression, nonlinear regression, and the Colorado State University empirical formula. The optimal MLP architecture, consisting of a single hidden layer with a hyperbolic tangent activation function, delivered outstanding performance: in the dimensional analysis, it achieved R2=0.99, RMSE = 0.01 m, MAE = 0.01 m; in the dimensionless form, R2=0.81, RMSE = 0.32 m, MAE = 0.32 m. By comparison, linear and nonlinear regressions produced R2≈0.58R^2\approx0.58–0.60 and RMSE ≈ 0.20–0.42 m, while the CSU equation yielded R2=0.84 and RMSE=0.52 m. Overall, the MLP reduced prediction errors by over 70% relative to linear regression, 85.5% versus nonlinear regression, and 87.7% compared to the CSU model, demonstrating its clear advantage for accurate scour-depth estimation (Emamgholizadeh & Rahimi, 2022).

These studies collectively highlight that no single ML algorithm is universally optimal; model performance depends heavily on input features, local climatic variability, and the quantity and quality of training data. Therefore, a systematic, comparative evaluation using uniform performance metrics is essential to identify context-specific best practices for evaporation estimation (Hoekstra & Chapagain, 2008).

This research focuses on the arid region of Shahrood to examine and evaluate the performance of machine learning models in estimating evaporation. For this purpose, it utilizes meteorological and climatological data collected by the Shahrood synoptic station. These data include parameters such as average temperature, average relative humidity, average wind speed, and total monthly Precipitation, which help estimate evaporation. Additionally, other parameters, such as average maximum and minimum temperatures, minimum and maximum relative humidity, are considered to improve model accuracy. To evaluate the models' performance, consistent and valid metrics such as mean squared error, coefficient of determination, and mean absolute error are used. The ultimate goal of this study is to introduce and select the most appropriate machine learning model for accurately estimating evaporation under the specific climatic conditions of the Shahrood region. This selection will be based on the models' performance in predicting evaporation and their alignment with the region's local and climatic conditions. The results of this study can assist water resource managers in better planning for water resources in arid areas and in preventing excessive evaporation.

2. Materials and methods

2.1. Study area

Shahrood is one of the key cities in Semnan Province, located in the northern part of the province near Iran's central desert. Situated approximately 350 kilometers northeast of Tehran, it lies at the border of North Khorasan and Yazd provinces. Shahrood's geographic position provides convenient connectivity to other regions of the country via major highways and railway networks. The climate of Shahrood is predominantly arid to semi-arid, characterized by hot, dry summers and cold, dry winters. These climatic conditions significantly influence the area's natural resources, agriculture, and economic activities. Annual precipitation is limited, averaging less than 200 millimeters, which imposes considerable constraints on agriculture and water availability. The region also experiences seasonal monsoon winds and dust storms during specific periods, further shaping its atmospheric dynamics. Due to these environmental factors, agricultural practices in Shahrood are primarily focused on drought-

resistant crops such as wheat, barley, and cotton. However, in some surrounding mountainous areas, milder temperatures allow for the cultivation of a wider variety of horticultural and field crops. Despite the challenging climate, Shahrood has sustained economic and industrial growth, supported by its rich mineral resources and well-developed transportation infrastructure. The study area is illustrated in Figure 1.

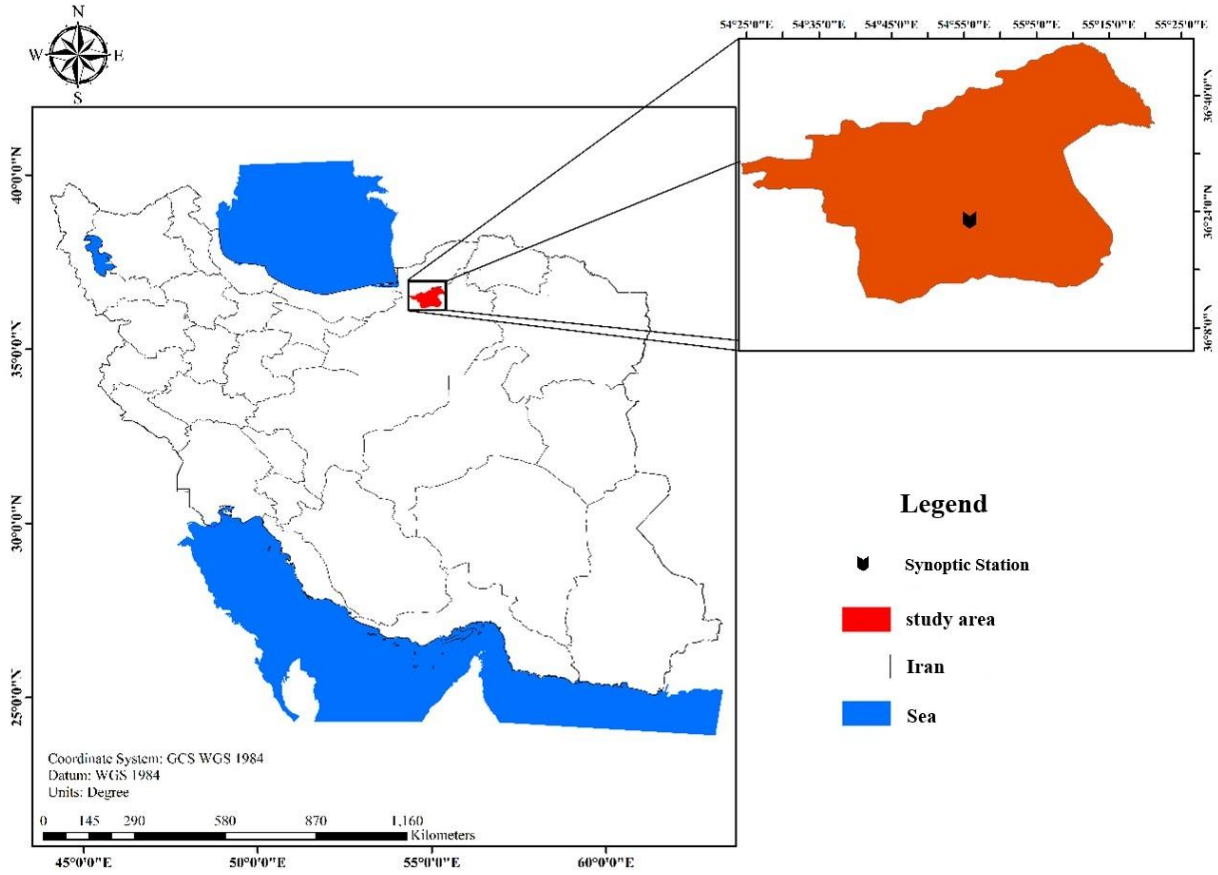

Figure 1. Study area: Geographic location of Shahrood Synoptic Station.

2.2. Data Collection

The dataset used in this study comprises meteorological data collected at the Shahrood Synoptic Station from 1992 to April 2025. The data are recorded monthly, providing comprehensive coverage over more than three decades. The climatic variables incorporated in this dataset are shown in Table 1.

2.3. Networks

2.3.1. Decision Tree

The Decision Tree algorithm is a machine learning approach commonly employed for regression tasks. It operates by recursively partitioning the input feature space, progressively dividing the data into smaller subsets to form a hierarchical tree-like structure. At each node of the tree, a specific feature is selected as the decision criterion, and the data is split into branches based on a threshold for that feature. The choice of the optimal feature and its corresponding threshold is typically guided by metrics such as variance reduction or minimization of mean squared error. This recursive process continues until a predefined stopping condition is met, such as reaching a minimum number of samples in a node or achieving a satisfactory level of variance reduction. For prediction, the output value for a new instance is estimated by averaging the target values of training samples within the corresponding leaf node. Notably, decision trees are invariant to feature scaling, eliminating the need for standardizing input data. Owing to their high interpretability and ability to capture nonlinear relationships

among features, decision trees are widely utilized in various regression problems.

Table 1. Description of Climatic Parameters Used in the research

Parameter Name	Description	Unit
Average Temperature	Monthly average air temperature	°C
Average Wind Speed	Monthly average wind speed	m/s
Average Maximum Temperature	Monthly average of daily maxima	°C
Average Minimum Temperature	Monthly average of daily minima	°C
Total Monthly Precipitation	Total precipitation in the month	mm
Minimum Relative Humidity	Minimum relative humidity in month	%
Maximum Relative Humidity	Maximum relative humidity in month	%
Average Relative Humidity	Monthly average relative humidity	%
Total Monthly Evaporation	Total evaporation in the month	mm

2.3.2. KNN

The K-Nearest Neighbors (KNN) algorithm is an instance-based learning method used for regression tasks that relies on the similarity between data points. In this approach, the parameter k—representing the number of neighbors considered during prediction—is first specified. The algorithm then computes the distance between the new sample and each instance in the training set, typically using the Euclidean distance metric. Once the k nearest neighbors are identified, the predicted value for the new input is estimated by averaging the target values of these neighbors. Due to its sensitivity to the scale of input features, KNN requires data standardization to ensure all features contribute equally to distance calculations and to avoid biases caused by varying feature magnitudes. Owing to its non-parametric nature and reliance on local instance-based estimation, KNN is particularly effective for analyzing datasets with complex or unknown distributions (El Bilali et al., 2022).

2.3.3. SVR

Support Vector Regression (SVR) is a machine learning technique rooted in Support Vector Machine (SVM) theory, designed to model complex nonlinear relationships between variables in regression problems. Rather than minimizing the absolute or squared error as in traditional regression approaches, SVR aims to identify an optimal hyperplane that predicts target values within a specified margin of tolerance, denoted by epsilon (ε). The objective is to construct a function that fits the data as accurately as possible while allowing for a predefined error margin, thereby reducing sensitivity to outliers. To capture nonlinear patterns, SVR maps the input data into a higher-dimensional feature space using kernel functions—such as linear, polynomial, or radial basis function (RBF) kernels. Within this transformed space, the algorithm defines two parallel hyperplanes that enclose the majority of the training data. Only the data points that fall outside this epsilon-insensitive region — known as support vectors — directly influence the final model. The optimization process in SVR involves minimizing a cost function that penalizes deviations beyond the epsilon-margin, controlled by the regularization parameter C. This parameter balances the trade-off between model complexity and the tolerance for errors. The proper tuning of C and ε is crucial for achieving a model that generalizes well, avoiding both overfitting and underfitting.

2.3.4. Random Forest

Random Forest is an ensemble-based machine learning algorithm that constructs a collection of decision trees to perform predictions. Each tree is trained independently using a randomly selected subset of the data and features, and the final output is determined by aggregating the predictions of all trees—typically through averaging for regression tasks or majority voting for classification. This method is particularly effective in reducing the risk of overfitting and enhancing predictive accuracy, especially in complex or noisy datasets. One of

the core strengths of Random Forest lies in its stochastic nature: both the sampling of training instances (via bootstrapping) and the selection of features at each split are randomized. This diversity among individual trees contributes to a more robust and generalizable model. Furthermore, Random Forest offers advanced capabilities, such as feature importance estimation, which provides valuable insights into each variable's relative contribution to the predictive task. The model's performance is influenced by key hyperparameters, notably the number of trees (n_estimators) and the maximum depth of each tree (max_depth), which can be tuned to balance model complexity and accuracy. Owing to its flexibility and resilience to outliers and high-dimensional data, Random Forest is widely applied in both regression and classification problems across various domains.

2.3.5. ANN (Artificial Neural Network)

Artificial Neural Networks (ANNs) are among the widely used methods in machine learning. These models consist of interconnected layers of computational units, where each unit receives a set of weighted inputs, processes them using an activation function, and passes the output to the subsequent layer. ANNs can learn complex patterns and nonlinear relationships in data by adjusting connection weights through optimization algorithms such as gradient descent and backpropagation. This learning process minimizes the prediction error on training data and enables the model to generalize to unseen samples. Key advantages of ANNs include their structural flexibility, adaptability to high-dimensional and large-scale datasets, and strong predictive power. They have been successfully applied in various tasks such as regression, classification, pattern recognition, and time series forecasting.

2.4. Evaluation Metrics

2.4.1. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a metric that quantifies the average prediction error of a model by calculating the mean of the absolute differences between the predicted and actual values. This metric represents the average absolute distance between predictions and actual values, with units consistent with the target variable, making its interpretation straightforward and intuitive. The MAE value indicates, on average, how far the model's predictions deviate from the actual observations. It is computed using equation (1):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_{\hat{i}} - y | \quad (1)$$

where y denotes the actual values, $y_{\hat{i}}$ the predicted values, and n the number of samples.

2.4.2. The Coefficient of Determination (R2)

The Coefficient of Determination (R^2) is a metric that indicates how well a predictive model fits the data and explains the variability of the target variable. It represents the proportion of the variance in the observed data that is accounted for by the model's predictions. The value of R^2 ranges from 0 to 1, with values closer to 1 indicating a better fit and higher predictive accuracy. An R^2 equal to zero indicates that the model fails to explain any of the variation in the data. This metric is calculated using equation (2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y - y_{\hat{i}})^2}{\sum_{i=1}^{n} (y - \bar{y})^2} \quad (2)$$

where y denotes the actual values, $y_{\hat{i}}$ the predicted values, and $\bar{y}$ the mean of the actual values.

2.4.3. R-Square Value (R^2)

The Kling-Gupta Efficiency (KGE) is a widely used performance metric for evaluating predictive models, especially in hydrology and environmental engineering. This index provides a comprehensive assessment by integrating three key components: the correlation between observed and predicted values, the ratio of their means, and the ratio of their standard deviations. The KGE value ranges from negative infinity to 1, where 1 indicates perfect agreement between predictions and observations. Unlike traditional metrics such as R^2 and NSE, the main

advantage of KGE lies in its simultaneous consideration of correlation, bias, and variability, offering a more balanced and insightful evaluation of model performance. The KGE is calculated using equation (3):

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\mu_{\hat{y}}}{\mu_y} - 1\right)^2 + \left(\frac{\sigma_{\hat{y}}}{\sigma_y} - 1\right)^2} \quad (3)$$

where r is the Pearson correlation coefficient between the predicted values $\hat{y}$ and the observed values y, μ denotes the mean, and σ represents the standard deviation of the respective datasets.

2.4.4. Average Absolute Relative Deviation (AARD)

AARD (Average Absolute Relative Deviation) is a metric that quantifies a model's prediction error relative to actual values, expressed as a percentage. This indicator calculates the average absolute difference between predicted and observed values relative to the observed values, making it convenient for comparing the accuracy of different models or datasets. A lower AARD value indicates higher predictive accuracy. It is calculated using equations (4) and (5):

$$ARD_i = \frac{(y - y_i)}{y_i} \quad (4)$$

$$AARD = \frac{1}{N}\sum_{i=1}^{N} |ARD_i| \quad (5)$$

In the equation y is the observed value, y_i is the predicted value, and N is the number of samples.

3. Results and Discussion

In this research, five machine learning models Decision Tree, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Random Forest, and Artificial Neural Network (ANN) were evaluated for evaporation prediction using four performance metrics: Mean Absolute Error (MAE), coefficient of determination (R^2), Kling–Gupta Efficiency (KGE), and Average Absolute Relative Deviation (AARD). As shown in Table 2, the Decision Tree performed worst, with an MAE of 25.82, an R^2 of 0.945, a KGE of 0.963, and an AARD of 0.895. KNN also performed relatively poorly, likely due to its high sensitivity to data noise, yielding an MAE of 22.45, an R^2 of 0.959, a KGE of 0.936, and an AARD of 1.098. SVR offered a more balanced trade-off between accuracy and robustness, with an MAE of 22.09, an R^2 of 0.957, a KGE of 0.960, and an AARD of 1.042. The Artificial Neural Network (ANN) demonstrated solid predictive capability, achieving an R^2 of 0.961, an MAE of 22.60, and an AARD of 0.627. These results suggest that the ANN is well-suited to capturing the complex patterns of evaporation variability. However, it fell slightly behind the Random Forest in minimizing both absolute and relative error. Analysis of Table 1 shows that the Random Forest model achieved the best overall performance, with the lowest MAE (19.94), an R^2 of 0.957, the highest KGE of 0.973, and the lowest AARD of 0.521. This superiority reflects its ability to aggregate multiple decision trees, thereby reducing variance and minimizing error. Based on the results, Random Forest achieved the highest accuracy and stability for evaporation prediction, while ANN and SVR also performed well. In contrast, simpler models such as Decision Trees and KNN struggled to capture the underlying complexity of the evaporation data, leading to higher prediction errors.

Table 2. Comparative performance metrics of the machine learning models

Model	MAE (mm)	R2	KGE	AARD
Decision Tree	25.82	0.945	0.963	0.895
KNN	22.45	0.959	0.936	1.098
SVR	22.09	0.957	0.960	1.042
Random Forest	19.94	0.957	0.973	0.521
ANN	22.60	0.961	0.958	0.627

As illustrated in Figure 2, which presents the time series of evaporation predictions, all models successfully capture the seasonal pattern of evaporation with prominent peaks typically occurring in June and September, and troughs observed in spring and autumn. However, the precision with which each model tracks these peaks and troughs varies. The decision tree model frequently underestimates peak evaporation values, especially during high-evaporation periods such as the summers of 1997 and 2006. In contrast, the KNN model tends to overestimate peaks in certain years, for instance, in the summers of 1995 and 2002. The SVR model generally predicts peak values slightly lower than observed and shows notable errors during low-evaporation periods, such as the early spring of 2003 or the autumn of 2015. The random forest and artificial neural network (represented by dark blue and green lines, respectively) tend to follow the actual observations (light blue line) more closely, particularly in capturing intense evaporation peaks in 2005, 2010, and 2018. Nevertheless, RF occasionally exaggerates peak values ,as in summer 2009, while ANN sometimes slightly overestimates very low evaporation levels, as in autumn 2014. Overall, the most significant deviations occur during extreme peaks and troughs, where RF and ANN have shown superior performance in minimizing tracking errors. Both models effectively preserve the seasonal structure of evaporation dynamics and closely follow the timing and magnitude of fluctuations. Still, when compared directly, RF shows a slightly higher tendency to overpredict peaks, whereas ANN exhibits relatively more error in estimating sharp declines.

Figure 3 illustrates the scatter plots of predicted versus observed evaporation values for each model, providing a visual assessment of their accuracy and error patterns. The dispersion and alignment of points relative to the 45-degree reference line (the ideal fit) vary significantly across models. The predictions from the Random Forest and Artificial Neural Network models exhibit the closest clustering around this line, indicating minimal absolute error and the highest correlation with observed data. Both models demonstrate superior accuracy not only at peak evaporation levels but also across lower and moderate ranges, with fewer instances of under- or overestimation than the others. In contrast, the Decision Tree model shows the most incredible spread, suggesting weaker predictive performance. The KNN and SVR models fall somewhere in between. At the same time, they maintain relatively strong correlations with actual values; however, their predictions exhibit more scatter, particularly across certain evaporation intervals, than those of RF and ANN. This visual analysis reinforces the earlier findings, confirming that Random Forests and Artificial Neural Networks have the greatest capacity to capture evaporation variability accurately. Simpler models, such as the Decision Tree, struggle to capture the data's complexity and consequently produce higher errors.

Figure 4 presents a bar chart comparing the predicted evaporation values of five machine learning models against actual observations across a range of sampled dates. The Random Forest and Artificial Neural Network models exhibit the least deviation from the observed values. Notably, they closely align with actual evaporation rates during both peak periods and sharp declines. In contrast, the Decision Tree and K-Nearest Neighbors models show greater variability in their predictions and, in many

cases, particularly during the middle of the sampling range and at very low evaporation levels, tend to overestimate or underestimate the values systematically. Although the SVR model demonstrates a reasonable correlation with observed data, it tends to underestimate peak evaporation events. This performance pattern underscores the superior ability of the Random Forest and ANN models to accurately capture evaporation fluctuations and maintain predictive stability across variable conditions. These findings suggest that future research and practical applications involving evaporation prediction should prioritize the use of these two models.

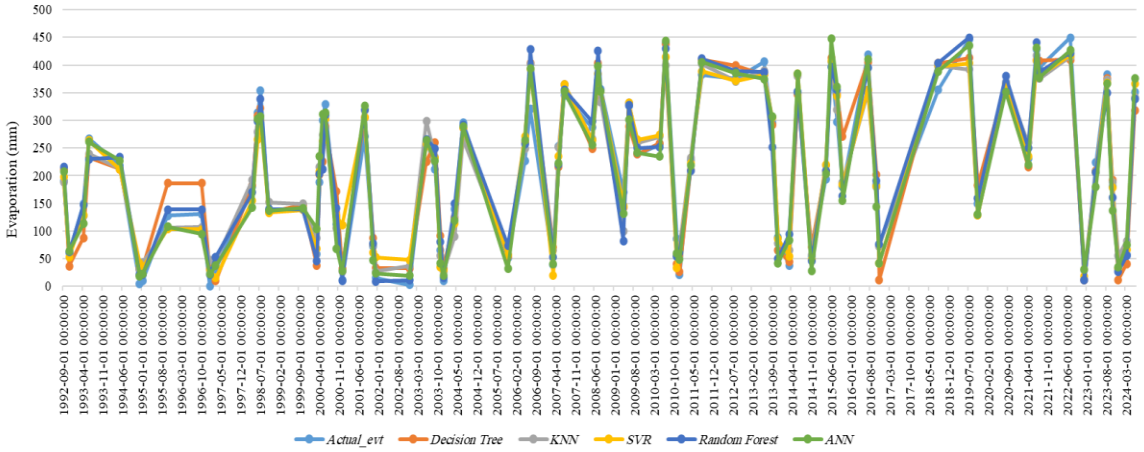

Figure 2. Comparison of predicted evaporation values by five machine learning models with actual observations over a range of sampled dates.

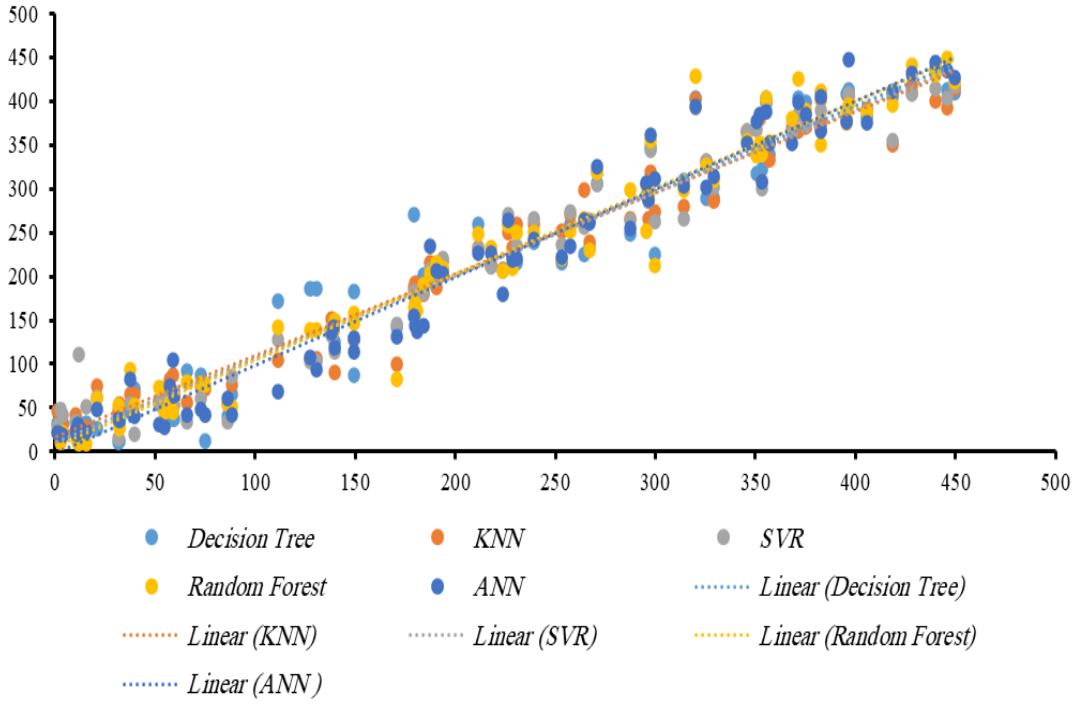

Figure 3. Scatter plots of predicted versus actual evaporation values for five machine learning models. The proximity of points to the 1:1 line indicates the accuracy of the prediction.

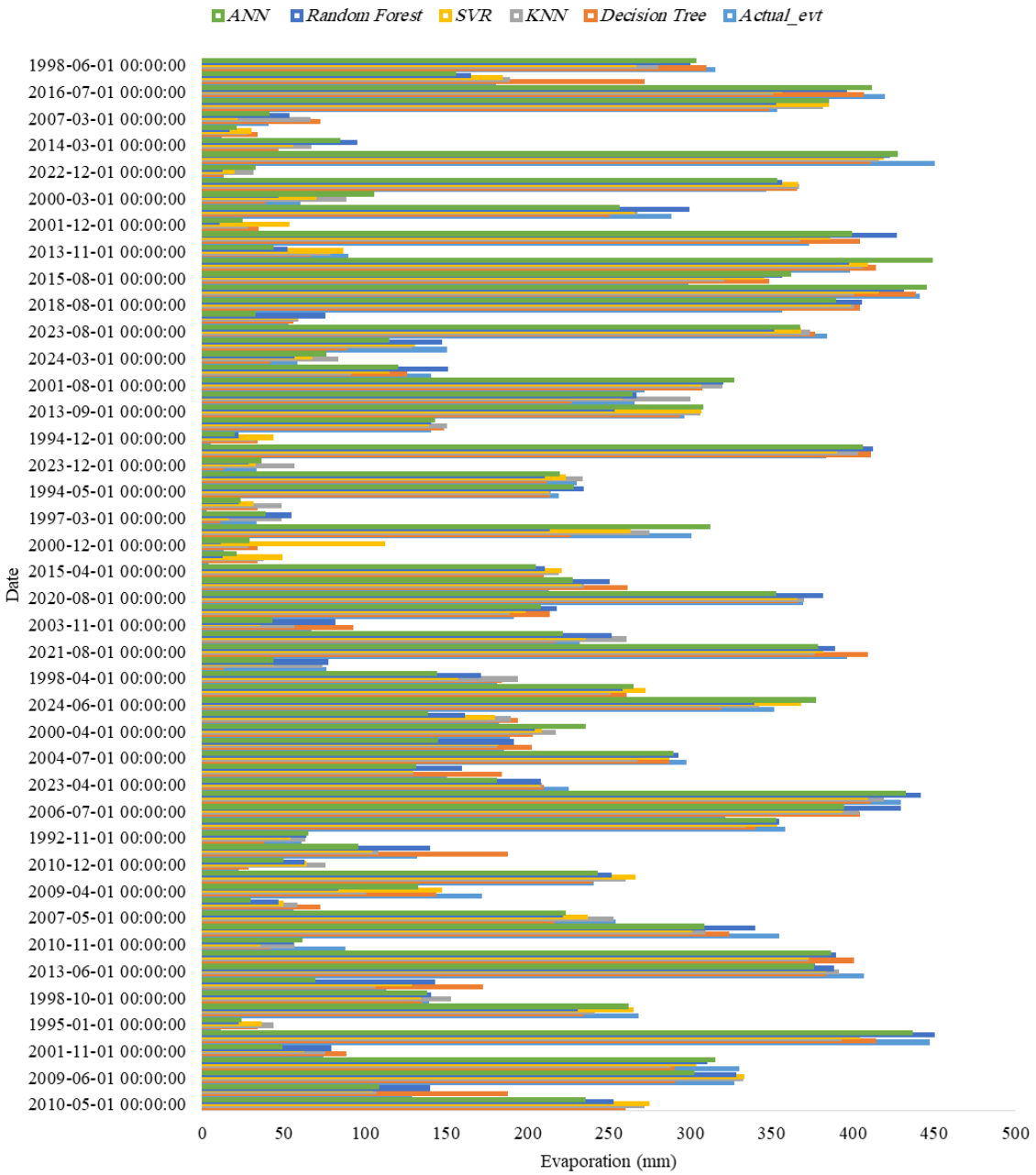

Figure 4. Comparison of predicted evaporation values by five machine learning models with actual observations**.**

4. Conclusions

This research comprehensively evaluated the performance of five widely used machine learning algorithms: Decision Tree, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Random Forest, and Artificial Neural Network (ANN) in predicting monthly evaporation rates in the Shahrood region using long-term meteorological data from the Shahrood Synoptic Station spanning 1992 to April 2025. Evaporation is a critical component of the hydrological cycle, especially in arid and semi-arid regions like Shahrood, where excessive evaporation can lead to significant water loss and exacerbate water scarcity issues. Accurate estimation of evaporation is essential for effective water resource management, agricultural planning, and climate impact mitigation in such vulnerable environments. However, direct measurement of evaporation can be challenging. Synoptic stations sometimes have missing evaporation

data for certain months, and in some cases, evaporation measurements have not been recorded for recent years. Moreover, in some regions, evaporation data might be completely unavailable due to the lack of measurement infrastructure. These limitations motivate the use of advanced machine learning algorithms to predict evaporation from other meteorological variables, enabling reliable estimates even in the absence of direct observations. The dataset, comprising diverse climatic parameters such as temperature, wind speed, precipitation, relative humidity, and evaporation, enabled an in-depth assessment of model performance under real-world conditions. Predictive accuracy and robustness, achieving the lowest Mean Absolute Error (MAE), highest Kling-Gupta Efficiency (KGE), and lowest Average Absolute Relative Deviation (AARD). The Artificial Neural Network also showed strong performance, slightly trailing Random Forest in overall metrics but excelling in modeling intricate evaporation dynamics. Conversely, simpler models such as Decision Trees and KNN exhibited limitations in handling the complexity and variability of evaporation processes, leading to higher prediction errors and less stable outputs. Support Vector Regression achieved balanced performance but did not surpass the ensemble and neural network approaches. The temporal analysis revealed that all models captured the seasonal patterns of evaporation, with peaks generally occurring during summer months and troughs in spring and autumn. However, Random Forest and ANN more precisely tracked these fluctuations, including extreme peaks and sharp declines, underscoring their suitability for hydrological and climatic applications in semi-arid regions such as Shahrood. For future research, it is recommended to explore advanced deep learning techniques, such as Long Short-Term Memory (LSTM), which have shown great potential for capturing complex temporal dependencies and nonlinear patterns in hydrological time series, such as evaporation. Additionally, expanding the application of these machine learning models to other semi-arid and arid regions lacking direct evaporation measurements would provide valuable insights into their generalizability and robustness across different climatic and geographic conditions. Furthermore, a comparative analysis between traditional evaporation estimation methods and state-of-the-art machine learning approaches would help clarify the advantages and limitations of each, guiding more effective selection of prediction tools for water resource management. Such efforts will ultimately contribute to more accurate and reliable evaporation forecasting, essential for optimizing water use and addressing scarcity challenges in vulnerable regions.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

References

Aghelpour, P., Bagheri-Khalili, Z., Varshavian, V., & Mohammadi, B. (2022). Evaluating three supervised machine learning algorithms (LM, BR, and SCG) for daily pan evaporation estimation in a semi-arid region. *Water*, *14*(21), 3435.

Al Sudani, Z. A., & Salem, G. S. A. (2022). Evaporation rate prediction using advanced machine learning models: a comparative study. *Advances in Meteorology*, *2022*(1), 1433835.

Ali, J., & Saraf, S. (2015). Evaporation modelling by using artificial neural network and multiple linear regression technique. *International Journal of Agricultural and Food Science*, *5*(4), 125-133.

Amer, Z., & Farah, B. (2025). Evaporation forecasting using different machine learning models in Beni Haroun Dam, Algeria. *Theoretical and applied climatology*.

Deo, R., Samui, P., & Kim, D. (2016). Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. *Stochastic Environmental Research and Risk Assessment*, *30*, 1769-1784.

Deswal, S., & Pal, M. (2008). Modeling Pan Evaporation Using a Support Vector Machine. *ISH Journal of Hydraulic Engineering*, *14*(1), 104-116.

Dong, H., Geng, Y., Sarkis, J., Fujita, T., Okadera, T., & Xue, B. (2013). Regional water footprint evaluation in China: a case of Liaoning. *Science of the Total Environment*, *442*, 215-224.

Ehteram, M., Barzegari Banadkooki, F., & Afshari Nia, M. (2024). Gaussian mutation-alpine skiing optimization algorithm-recurrent attention unit-gated recurrent unit-extreme learning machine model: an advanced predictive model for predicting evaporation. *Stochastic Environmental Research and Risk Assessment*, *38*(5), 1803-1830.

Emamgholizadeh, S., Bahman, K., Bateni, S. M., Ghorbani, H., Marofpoor, I., & Nielson, J. R. (2017). Estimation of soil dispersivity using soft computing approaches. *Neural Computing and Applications*, *28*, 207-216.

Emamgholizadeh, S., & Demneh, R. K. (2019). A comparison of artificial intelligence models for the estimation of daily suspended sediment load: a case study on the Telar and Kasilian rivers in Iran. *Water Supply*, *19*(1), 165-178.

Emamgholizadeh, S., Parsaeian, M., & Baradaran, M. (2015). Seed yield prediction of sesame using artificial neural network. *European Journal of Agronomy*, *68*, 89-96.

Emamgholizadeh, S., & Rahimi, M. A. (2022). Prediction of the scour depth of bridge pier using artificial neural network model and comparison with empirical equations. *Advanced Technologies in Water Efficiency*, *1*(1), 70-90.

Ercin, A. E., & Hoekstra, A. Y. (2014). Water footprint scenarios for 2050: A global analysis. *Environment International*, *64*, 71-82.

Falkenmark, M. (1995). Land–water linkages: a synopsis. In *Land and Water Integration and River Basin Management: Proceedings of an FAO Informal Workshop, Vol. 1* (pp. 15-16). Food and Agriculture Organization of the United Nations.

Gelete, G., & Yaseen, Z. M. (2024). Hybridization of deep learning, nonlinear system identification and ensemble tree intelligence algorithms for pan evaporation estimation. *Journal of Hydrology*, *640*, 131704.

Hashemi, G., Mirheidari, S. P., & Santivanez, C. G. D. (2018). Urbanization Impact on the Water and Food Security and Assessment of Wheat Production and its Irrigation Water Requirements Using CROPWAT Model in IRAN: A Case Study of City Tehran. *Asian Journal of Advanced Science*, *6*(1), 7-15.

Hoekstra, A. Y., & Chapagain, A. K. (2008). *Globalization of Water: Sharing the Planet's Freshwater Resources*. Blackwell.

Kisi, O., Mirboluki, A., Naganna, S. R., Malik, A., Kuriqi, A., & Mehraein, M. (2022). Comparative evaluation of deep learning and machine learning in modelling pan evaporation using limited inputs. *Hydrological Sciences Journal*, *67*(9), 1309-1327.

Latif, S. D. (2024). Evaluating deep learning and machine learning algorithms for forecasting daily pan evaporation during COVID-19 pandemic. *Environment, Development and Sustainability*, *26*(5), 11729-11742.

Moghaddamnia, A., Ghafari, M., Piri, J., & Han, D. (2009). Evaporation estimation using support vector machines technique. *International Journal of Engineering and Applied Sciences*, *5*(7), 415-423.

Shabani, S., Samadianfard, S., Sattari, M. T., Mosavi, A., Shamshirband, S., Kmet, T., & Várkonyi-Kóczy, A. R. (2020). Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis. *Atmosphere*, *11*(1), 66.

Sudheer, K. P., Gosain, A. K., Mohana Rangan, D., & Saheb, S. M. (2002). Modelling evaporation using an artificial neural network algorithm. *Hydrological Processes*, *16*(16), 3189-3202.

Tezel, G., & Buyukyildiz, M. (2016). Monthly evaporation forecasting using artificial neural networks and support vector machines. *Theoretical and applied climatology*, *124*, 69-80.

Wu, L., Huang, G., Fan, J., Ma, X., Zhou, H., & Zeng, W. (2020). Hybrid extreme learning machine with meta-heuristic algorithms for monthly pan evaporation prediction. *Computers and electronics in agriculture*, *168*, 105115.

Yang, Y., & Chui, T. F. M. (2021). *Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods* (Vol. 25).