RESEARCH

A novel method for estimating daily evapotranspiration based on one, two, and three-day meteorological records using the long short-term memory model combined with feedforward neural networks

Milad Sharafi[1] · Hamed Talebi[2] · Saeed Samadianfard[2] · Ramendra Prasad[3]

Abstract

In agricultural and hydrological applications, daily reference evapotranspiration (ET) is a key parameter. Based on one, two, and three-day antecedent meteorological parameters including average air temperature (T), sunshine hours (S), average relative humidity (RH), average wind speed (W), and reference evapotranspiration (ET), this study develops a new model framework for estimating daily ET magnitudes. The Long Short-Term Memory (LSTM) model was combined with the Feedforward Neural Network (FFNN) model in order to develop the hybrid approach, which was used to estimate daily ET values at two stations, Babolsar and Bandar Anzali, Iran, between 1990 and 2022. Additionally, 70% of the data in the present study were used for training purposes and 30% for testing purposes. Based on the statistical metrics, the proposed LSTM-FFNN model was benchmarked against the standalone LSTM model. Results from the comparison at Babolsar station indicated that the LSTM-FFNN model performed better than all scenarios with an error of approximately 1 millimeter per day in scenarios 12, 15, and 16. According to the standalone LSTM model, scenarios 5, 8, 11, 15, and 16 had a 1.57 mm/day error rate, and likewise, for the Bandar Anzali station, the combined LSTM-FFNN model performed the best in scenarios 14 and 15 with the lowest error rates (RMSE=0.96 mm/day), while the standalone LSTM model performed better in scenarios 15 and 16. It was found, however, that the proposed LSTM-FFNN model outperformed the standalone LSTM model at both sites. These results suggest that the proposed LSTM-FFNN model is capable of accurately estimating daily ET magnitudes, which may assist in the optimal management of water resources within agriculture.

Keywords Long short-term memory · Evapotranspiration · Feedforward neural network · Humid climates

Introduction

Agriculture is the biggest consumer of the world's freshwater reserves. Inadequate water availability and allocations affecting crop growth and harvest would lead to reduced food production or even food scarcity (De Fraiture et al., 2010).

✉ Saeed Samadianfard
s.samadian@tabrizu.ac.ir

[1] Department of Water Engineering, Urmia University, Urmia, Iran

[2] Department of Water Engineering, University of Tabriz, Tabriz, Iran

[3] Department of Science, School of Science and Technology, The University of Fiji, Lautoka, Fiji

The impacts of anthropogenic development, climate change, environmental degradation, and industrial water demand have placed enormous pressure on the already depleting freshwater resource. Hence, water conservation has been a major concern for irrigated agriculture (Hamdy et al., 2003). Planning, management, and regulation of agricultural water resources require an accurate estimation of water use over agricultural (Wanniarachchi & Sarukkalige, 2022). Proper water resources management can help mitigate these issues and promote equitable access to water resources for all activities.

In agricultural water management and the hydrological cycle, one of the most influential parameters is evapotranspiration (ET). ET is the process by which water is transferred from the land surface and vegetation to the atmosphere through evaporation from soil and transpiration from plants (Chang et al., 2018). It plays a crucial role in regulating the

water balance of ecosystems (Wang et al., 2023). In order to optimize the water management system, accurate estimation of ET values is crucial (Reyes-Gonzalez, 2017). Accurate estimation of ET is essential for crop growth modeling, climate change impact studies, ecosystem management, estimating crop water resource requirements, and subsequent irrigation scheduling (Hussain et al., 2020).

Direct point-based measurement of ET can be performed using a Lysimeter and then estimations care performed using the Bowen ratio–energy balance approach, and Eddy covariance techniques (Todd et al., 2000). These methods have several inherent drawbacks such as the high cost of installation and maintenance of Lysimeter, the need for a large area of land with similar actual conditions, and the need for specialized skilled personnel to collect and process the data (Todd et al., 2000). It is also possible to indirectly estimate ET by using a set of climatological variables in order to define reference evapotranspiration (Zouzou & Citakoglu, 2023). Estimating reference ET is commonly used in water management since it represents the amount of water that would be lost to the atmosphere by a hypothetical, well-watered grass reference crop that completely covers the soil surface and experiences minimal water stress under optimal conditions (Hargreaves, 1994; Cobaner et al., 2017). The FAO-56 Penman–Monteith (FAO-56 PM) method is one of the stable and well-established techniques for determining ET (Allen et al., 1998). The FAO-56 PM method combines energy balance and aerodynamic principles to estimate ET based on meteorological data such as temperature, humidity, wind speed, and solar radiation. This method was validated using the lysimeter technique in different climates, it can also be used to validate other ET computation methods (Landeras et al., 2008), yet proper calibration is necessary.

Alternatively, to save on costs, tedious calibration and validation processes, machine learning algorithms can reliably estimate and forecast ET (Roy, 2021; Bayram & Çıtakoğlu, 2023). Using machine learning algorithms, it is possible to identify complex relationships between meteorological parameters and ET values that are difficult to detect using traditional methods (Yamaç & Todorovic, 2020; Citakoglu et al., 2014). In a study, Hu et al. (2022) investigated ET in Pakistan during the period 2015-2021 using the Internet of Things (IoT) based on machine learning (ML) models including k-nearest neighbors (KNN), Gaussian Naive Bayes (GNB), artificial neural network (ANN), and support vector machine (SVM). Comparing and evaluating the results of this research indicated that the KNN model had more suitable performance compared to other ML models with an accuracy of 92%. Mehdizade et al. 2021 used the adaptive neuro-fuzzy inference system (ANFIS) algorithm in combination with the shuffled frog-leaping algorithm (SFLA) and the invasive weed optimization method (IWO) to estimate ET_0. Results showed

that these models outperformed classic ANFIS as well as empirical models, with ANFIS-SFLA demonstrating the best performance (RMSE = 0.15 mm.day^{-1}, R^2 = 0.99). Hadadi et al. (2022) in research, the performance of artificial intelligence models, including ANFIS and its hybrids with SFLA and grey wolf optimization (GWO) optimization algorithms, in determining monthly Actual evapotranspiration (AET) in Iran is evaluated. By combining optimization algorithms with ANFIS, further accuracy was enhanced, resulting in promising results for estimating AET in arid climates. Talebi et al. 2023 developed an algorithm based on multilayer perceptrons (MLP) and MLP optimized with stochastic gradient descent (SGD) (MLP-SGD) for estimating daily ET_0 in two different climates. Results showed that the hybrid model with SGD-MLP performed better than the single model with all input parameters. In another study, Kushwaha et al. (2022) evaluated additive regression (AdR), random subspace (RSS), and the M5P tree models for modeling daily ET. The evaluation of the used parameters demonstrated that with more input variables, the performance of the model increased. Also, among the implemented models, the AdR6 model, which had all meteorological variables as input during the test period, performed better than other models with a coefficient of determination of 0.998. In a recent study by Sabanci et al. (2023), the researchers aimed to estimate ET_0 using different machine learning models, as an improvement over the FAO 56-PM approach. The study focused on 12 stations located in the Central Anatolian Region (CAR), which had diverse climate characteristics. To evaluate the performance of the models, the researchers used commonly used metrics such as R^2, MAE, RMSE, and PI. Among the models tested, the LSTM, ANN, and multivariate adaptive regression splines (MARS) showed the best performance at eight, three, and one station(s), respectively. The selected models achieved impressive values for R^2 (ranging from 0.987 to 0.999), MAE (ranging from 1.948 to 4.567), RMSE (ranging from 2.671 to 6.659), and PI (ranging from 1.544 to 4.018). Overall, this research provided valuable insights into accurately estimating ET0 for various climate conditions in the Central Anatolian Region. In addition, supervised machine learning algorithms from artificial neural networks (ANN) such as feedforward neural networks (FFNN) (Mahesh, 2020) are also very effective for ET forecasting. FFNNs have been used to estimate ET in many different studies. For predicting daily ET in northwest Algeria, Achite et al. (2022) compared the FFNN and Radial Basis Function Neural Network (RBFNN) and found that the FFNN model performed better than the Penman-Monteith model in predicting ET with a coefficient of determination of 0.992. In modeling and predicting water quality using adaptive neural fuzzy inference system (ANFIS) and FFNN Hmoud Al-Adhaileh and Waselallah Alsaade (2021) found that the FFNN model registered higher accuracy (100%)

for water quality classification (WQC) despite the ANFIS model having acceptable accuracy for predicting WQI water quality index values. The review of the conducted studies indicates that the FFNN model has performed well as a reliable model. The FFNN model has four major advantages: 1. Adaptability: even without the presence of the user, there can be a high correspondence between observed and predicted values. 2. Non-linearity: it causes the relationship between parameters to be found in the best way. 3. Mapping between input and output: reduces the difference between input and output to the minimum possible value. 4. High strength: it reduces the amount of noise (Svozil et al., 1997).

With the emergence of advanced artificial intelligence techniques, deep learning approaches have been widely used recently. One such DL model, the Long Short-Term Memory (LSTM) is commonly used for time-series data analysis (Sun et al., 2019; Citakoglu, 2021; Uncuoglu et al., 2022). These DL algorithms can process large amounts of data quickly and with high accuracy, reducing the time required for manual calculations. It is particularly useful for modeling sequences of data where there are long-term dependencies between the inputs and outputs. During the training process, the LSTM model 'learns' to identify patterns in historical hydro-meteorological data that are associated with daily ET changes (Ferreira & da Cunha, 2020). The LSTM model has attracted the attention of scientists due to its appropriate accuracy in recent years (Demir & Citakoglu, 2023). In a recent study, Alibabaei et al. (2021) modeled ET and reference soil water content using climate data by deep learning method. The results of this research indicated that LSTM achieved the best performance with a coefficient of determination of 0.9. Similarly, Chen et al. (2020) estimated daily ET based on limited meteorological data using deep learning and machine learning methods in the northeastern plain of China. The results of this study indicated that when temperature-based features were available, temporal convolution neural network (TCN) and LSTM models performed significantly better than experimental temperature-based models beyond the study regions.

Scholars argue and have shown that hybrid models tend to have more suitable performance in comparison to standalone models. In research conducted by Jia et al. (2023), two hybrid models were introduced to predict ET at four climate stations in Shaanxi province, China. These models combined particle swarm optimization (PSO) with the LSTM network. To train the models, 40 years of historical data were utilized, with PSO optimizing the hyperparameters within the LSTM network. The resulting optimized models were then employed to predict daily ETo in 2019, using different datasets. Therefore, predicting daily reference evapotranspiration with hybrid methods can be more cost-effective and precise than classical models. Therefore, the primary objective of this study is to introduce an innovative hybrid model that integrates Feedforward Neural Networks with Long Short-Term Memory Networks for the accurate prediction of daily reference evapotranspiration (ET_0). Even though coastal areas have a high potential for agricultural industry having humid climates, high rainfall, and fertile soil, estimating ET in these humid climates is a challenge. Hence the proposed model is tested at two distinct sites in the humid climates of Iran. To achieve the mentioned goal, the approach is based on meteorological parameters from one, two, and three-day antecedent periods. The next section of the paper presents materials and methods following are results and conclusions.

Materials and methods

Study area

Effective management of water resources in humid and extremely humid climates is critical to ensure sustainable and equitable utilization. As a result, two stations with a varied climate, including Bandar Anzali and Babolsar in Iran, were selected as the study sites. Bandar Anzali experiences an average annual rainfall of 1733.9 mm and an average annual temperature of 16.4°C. In Babolsar, the average annual rainfall is slightly lower at around 902.9 mm, and the average annual temperature is slightly higher at around 17.3°C (Fallah-Ghalhari & Shakeri, 2023). As the aim of the study was to develop modelling options for arid climatic conditions, the de Martonne aridity index was used to identify the stations. The de Martonne aridity index categorizes the stations on the basis of mean annual precipitation and mean temperature, which shows that both the Bandar Anzali and Babolsar stations are located in extremely humid and humid climates (Mehdizadeh, 2020). Also, in the present study, 70% of the data were used for the training phase and 30% of the data were used for the test phase. Table 1 summarizes the geographical characteristics of the study area, while Fig. 1 illustrates their respective locations.

Table 1 Geographical characteristics of the studied stations.

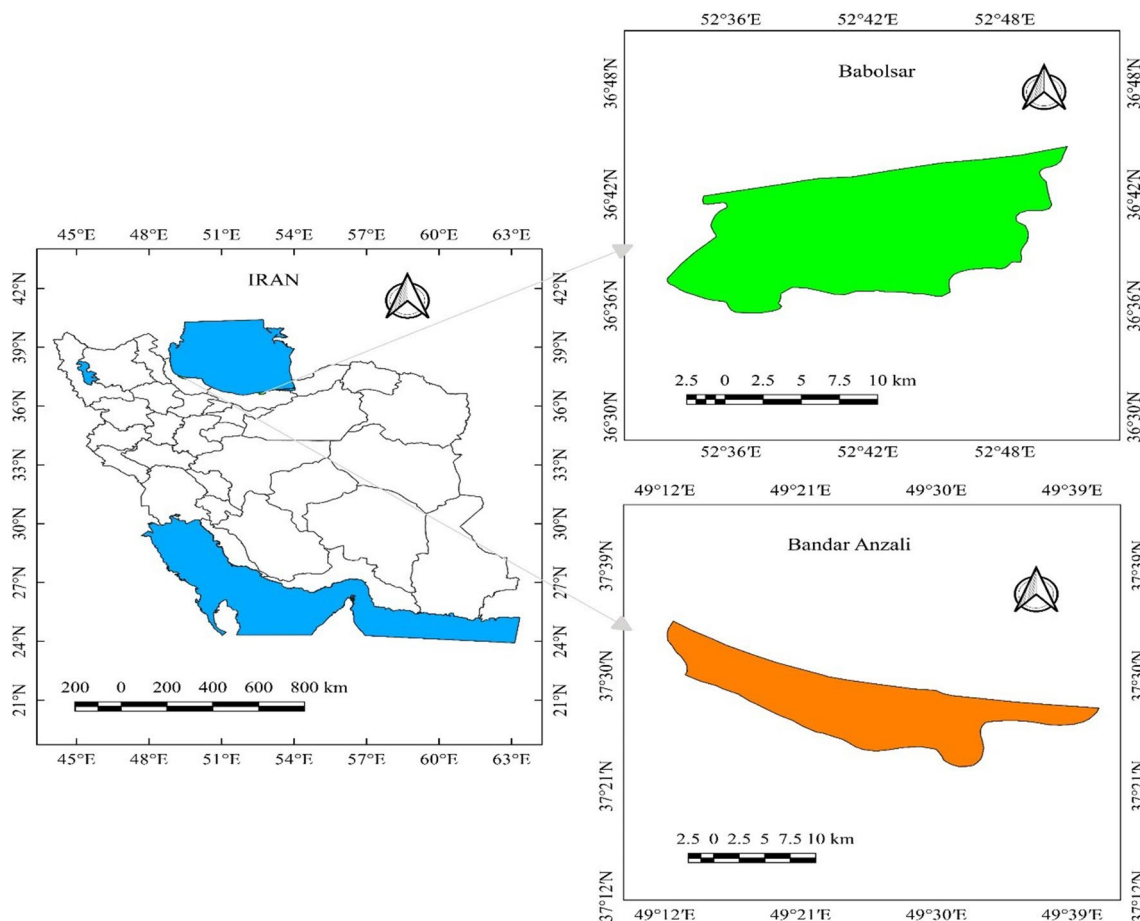| Station name	Latitude (°N)	Longitude (°E)	Altitude (m)	Mean annual precipitation (mm)	Climate type
Bandar Anzali	37.48	49.45	-23.6	1773.89	Extremely humid
Babolsar	36.72	52.65	-21	916.72	Humid

Fig. 1 The location of the studied stations

FAO-penman–Monteith (FAO-PM)

A standard method for estimating ET is the FAO-PM equation. FAO-PM was calculated using Eq. 1.

$$ET = \frac{0.408\Delta\left(R_n - G\right) + \gamma\frac{900}{T_{min}+273}U_2\left(e_s - e_a\right)}{\Delta + \gamma\left(1 + 0.34U_2\right)} \tag{1}$$

where ET is the reference evapotranspiration (mm/day), Δ is the slope of the saturation vapor pressure curve (kPa/°C) at the daily air temperature (°C), R_n and G are the net solar radiation and soil heat flux density (MJ/m^2 day), γ is the psychrometric constant (kPa/°C), T is the daily temperature (°C), U_2 is the wind speed (m/s), e_s is the saturation vapor pressure (kPa), and e_a is the actual vapor pressure (kPa) (Allen et al., 1998).

Equation 1 is used for ET estimations that can only be performed after the magnitude of parameters such as R_n, G, U_2, e_s, e_a, T, are known at the end of the day. In order to make effective decisions, prior knowledge is important at

the beginning of the day. Hence, this study tries to perform these estimations.

Long short-term memory (LSTM) model

Long Short-Term Memory (LSTM) is designed to handle the disappearing gradient problem that exists in traditional recurrent neural networks (RNN) (Hochreiter & Schmidhuber, 1997). LSTM networks use a memory cell to keep data over long periods and selectively forget or remember data based on input signals (Coşkun & Citakoglu, 2023). Three gates rule the memory cell: input gate, forget gate, and output gate (Gers et al., 2000). These gates are controlled by sigmoid activation functions the outputs of which ranges between 0 and 1. The input gate controls the amount of new data that should be added to the cell state and is calculated as follows:

$$i_t = \sigma\left(W_i * \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

where i_t is the input gate vector at time t, σ is the sigmoid function, W_i is the weight matrix for the input gate, h_{t-1} is the previous hidden state, x_t is the current input vector, and b_i is the bias vector for the input gate.

The forget gate is calculated as follows:

$$f_t = \sigma\left(W_f * [h_{t-1}, x_t] + b_f\right) \tag{3}$$

where f_t is the forget gate vector at time t, W_f is the weight matrix for the forget gate, h_{t-1} is the previous hidden state, x_t is the current input vector, and b_f is the bias vector for the forget gate.

The output gate determines the magnitude of the cell state that needs to be expelled to generate a new hidden state, which is calculated as follows:

$$o_t = \sigma\left(W_o * [h_{t-1}, x_t] + b_o\right) \tag{4}$$

where o_t is the output gate vector at time t, W_o is the weight matrix for the output gate, h_{t-1} is the preceding hidden state, x_t is the current input vector, and b_o is the bias vector for the output gate.

The cell state (C_t) at time t can be updated using these gates as follows:

$$C_t = f_t \bullet C_{t-1} + i_t \bullet \tanh\left(W_c * [h_{t-1}, X_t] + b_c\right) \tag{5}$$

where $\bullet$ denotes element-wise multiplication and *tanh* represents hyperbolic tangent function. W_c and b_c are weight matrix and bias vector respectively for updating cell state.

Finally, a new hidden state (h_t) can be generated using this updated cell state as follows:

$$h_t = o_t \bullet \tanh\left(C_t\right) \tag{6}$$

The structure of the LSTM model consists of multiple LSTM cells arranged sequentially. Figure 2 shows the structure of the standalone LSTM model. Each LSTM cell takes an input vector (x) and a hidden state (h) from its previous cell as 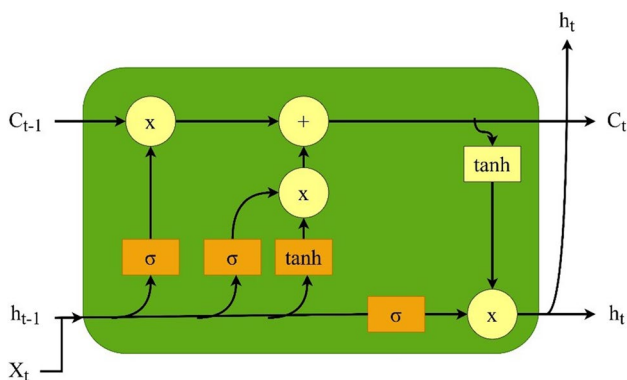inputs. The LSTM cell processes these inputs through its three gates (input gate "i", forget gate "f", and output gate "o") to update its internal memory (cell state "c"). The updated memory then generates a new hidden state (output "h"). LSTM architecture can have multiple layers stacked on top of each other to form a deep LSTM network. The outputs from each layer are fed into subsequent layers terminating in the output layer that produces the final forecasts or classifications. LSTM architecture has been widely used in various applications such as speech recognition, natural language processing (NLP), image captioning, etc., where sequential data needs to be processed with long-term dependencies.

Furthermore, the optimization of the number of neurons and epochs in the LSTM network and the batch size and learning rates was conducted through trial and error to enhance the overall accuracy. The implementation of a dropout layer serves as a regularization approach that is employed to mitigate the issue of overfitting within a neural network. During the training process, a subset of neurons is deliberately excluded or "dropped out" randomly, enabling the neural network to acquire knowledge from diverse designs and neuron configurations. In the present investigation, a dropout rate of 3% was implemented on the input layer. The activation function known as Rectified Linear Unit (ReLU) is commonly used in the middle layer of neural networks. The ADAM optimizer method was utilized to adjust the weights. The number of epochs, batch size, and learning rate were configured at 400, 20, and 0.0075, respectively. The results were selected via testing to enhance the level of accuracy of the network.

Feedforward neural network (FFNN) model

FFNN is a type of artificial neural network that consists of several layers with interconnected nodes so that each node performs a non-linear transformation on its inputs. The FFNN can learn complex patterns in the input data and extract useful features that can be used for prediction.
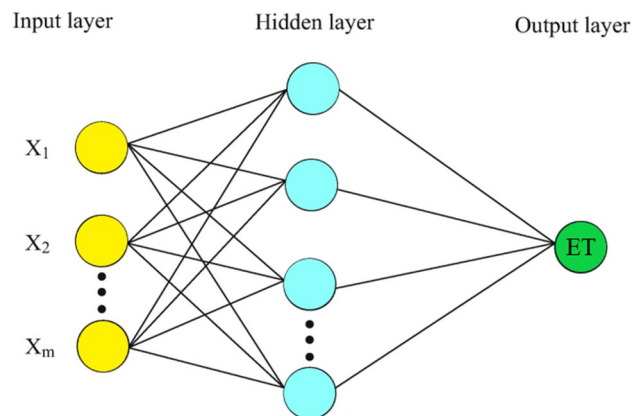

Fig. 2 Structure of a typical LSTM model

Fig. 3 Structure of a one-hidden layer FFNN model

Figure 3 displays the general assembly of the FFNN model. Each FFNN model has an input layer, a number of hidden layers, and output layers. A simple case of one hidden layer is presented in Fig. 3. In this study, after rigorous trial and error, the number of hidden layers was considered as one.

In the present study, the interconnection between neurons in each layer was established through weighted connections to all neurons in the subsequent layer—training the neural network involved iteratively updating the weights. The neurons in this architecture exhibited an inability to preserve the current condition. They only allowed for unidirectional signal propagation, perhaps due to the absence of internal interconnections and return loops. The experiment employed a guided test. The training strategies employed to generate the predictive models include Levenberg-Marquardt Back Propagation (LMBP), Bayesian Regularization, and the Scaled Conjugate Gradient approach. During the training phase, weighted connections propagated the input layer characteristics to the subsequent layer. The data underwent processing within the concealed layers and ultimately arrived at the output neuron. The computation involved determining the discrepancy between the output of the network and the intended target aim. This discrepancy was then used to update the layers' weights by propagating the error in a backward manner. The training procedures were deemed complete when no additional weight updates were observed. The Levenberg algorithm was employed to minimize mistakes within the framework of Levenberg's technique.

Furthermore, the neural network's activity was governed by the transfer function utilized in each layer. The transfer functions employed for the initial and secondary hidden layers were sigmoid-tan and sigmoid-log. In addition, the output layer also incorporated the sigmoid-log transferring mechanism.

The proposed hybrid LSTM-FFNN model architecture

A combination of optimization algorithms improves the accuracy of hydrological time series estimation (Mohammadi, 2023). In this study, the long short-term memory (LSTM) network and the feedforward neural network (FFNN) are combined to develop the proposed hybrid LSTM- FFNN algorithm. The LSTM- FFNN algorithm works by first extracting the pertinent features from the input time series data using the LSTM model. Once the features are extracted, these are fed into the FFNN network. The relationship between LSTM and FFNN lies in their architecture and application. While LSTM is used for tasks that involve sequential data, such as speech recognition or language translation, the FFNN is suitable for tasks that require processing static data, such as image recognition or text classification. In this study, both LSTM and FFNN are combined to create a hybrid model that can

handle both sequential and static data. This hybrid LSTM-FFNN algorithm is being used for time series estimation of daily ET values, which involves extracting pertinent features of a time-dependent variable and building a relationship based on its past values in predicting future values. Figure 4 displays the flowchart of the proposed hybrid LSTM-FFNN hybrid algorithm. Initially, the predictors are channeled into the algorithm and the assembly of the LSTM model is determined. Then the data training process begins. Once the least error criterion in training data is met, the process of testing commences. By determining
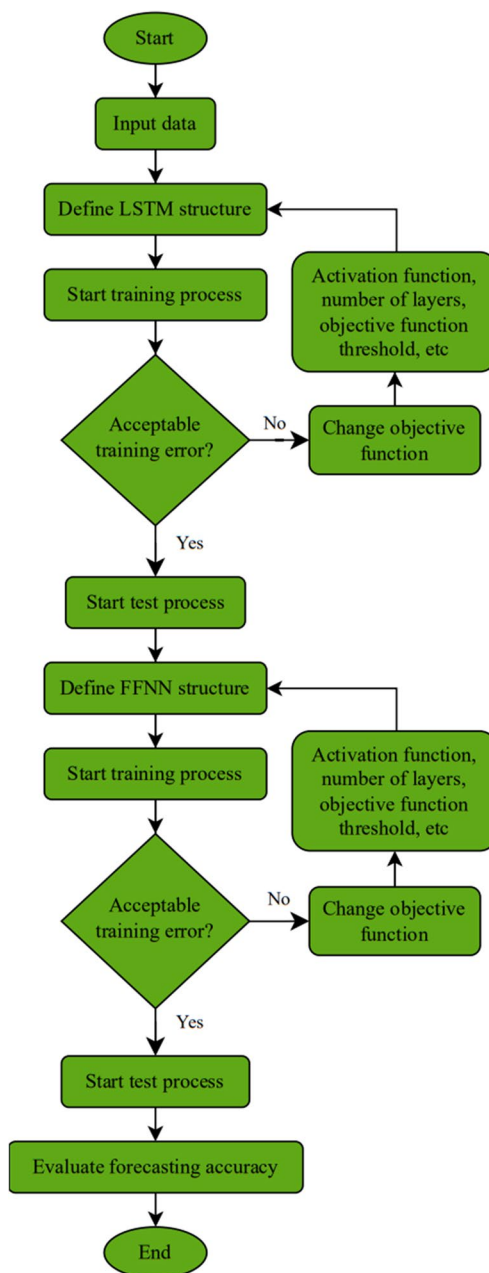

Fig. 4 Flowchart of hybrid LSTM-FFNN modelling framework

the structure of FFNN, the model starts evaluating training and testing again and then determines the prediction accuracy using evaluation criteria. Long Short-Term Memory (LSTM) models and their combination with Feedforward Neural Networks (FFNN) have proven to be effective in a wide range of sequence-based tasks, such as natural language processing and time series forecasting. However, they do have several limitations: Computational complexity, hyperparameter tuning, training time, overfitting (Nguyen et al., 2020).

The scenario-based modelling approach

To test the pertinence of the LSTM-FFNN model, daily ET estimation was performed using one, two, and three-day antecedent meteorological parameters. The parameters used include average air temperature (T), sunshine hours (S), average relative humidity (RH), average wind speed (W), and reference evapotranspiration (ET), while the integers -1, -2, and -3 in the parameters indicate a delay of one, two, and three days, respectively. The historic data from 1990-2022 have been as predictor inputs in this study.

The heat plots based on Pearson's correlation coefficient were used to study the probable impact of different input variables on ET at Babolsar and Anzali stations as shown in Fig. 5. For Babolsar station, the comparison between input parameters demonstrated that temperature has the highest correlation with ET values. On the other hand, relative humidity showed a negative correlation with ET revealing its lowest correlation with ET. For Anzali station, the comparison between the input parameters showed that antecedent evapotranspiration and then temperature had the highest correlation with ET followed by that was relative humidity, and then the wind speed had the lowest correlation with ET.

To have a comprehensive understanding of the influence of antecedent inputs on ET, a Scenario-based approach with incremental combinations of input series was used as shown in Table 2. For instance, in Scenario 1 (ET_{t-1}, T_{t-1}) two input series were used including one-day antecedent ET and one-day antecedent T. However, for Scenario 16 a total of 19 inputs were used for daily ET estimations (Table 2).

Model evaluation criteria

Model evaluation is one of the most important aspects. As such different evaluation parameters were considered to accurately evaluate the performances of the proposed models for ET estimations. The first criterion, coefficient of determination (R^2) varies from 0 to 1, is mathematically described as follows:

$$R^2 = \left(\frac{\left(\sum_{i=1}^{N} \left(Oi - \bar{O} \right) - \frac{1}{N} \left(Pi - \overline{P} \right) \right)}{\sqrt{\sum_{i=1}^{N} \left(O_i - \overline{O} \right)^2 \sum_{i=1}^{N} \left(P_i - \overline{P} \right)^2}} \right)^2 \quad (7)$$

Another broadly used statistical parameter is the root mean square error (*RMSE*), and it can be determined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Pi - Oi)^2} \quad (8)$$

The mean absolute error (*MAE*) is an index that is used to evaluate the model error and varies from 0 to ∞. It can be showed as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{n} |Oi - Pi| \quad (9)$$

Nash-Sutcliffe coefficient (*NS*) can range from -∞ to 1. If NS = 1, it shows complete agreement (Nash & Sutcliffe, 1970), expressed as:

$$\text{NS} = 1 - \left[\frac{\sum_{i=1}^{N} (Oi - Pi)^2}{\sum_{i=1}^{N} \left(Oi - \overline{O}i \right)^2} \right] \quad (10)$$

While the Willmott's index of agreement (*WI*) ranges between 0 to 1. A value of 1 shows a complete correlation, and 0 shows a mismatch between the observed and predicted values (Willmott, 1981), expressed as:

$$\text{WI} = 1 - \left[\frac{\sum_{i=1}^{N} (Oi - Pi)^2}{\sum_{i=1}^{N} \left(\left| Pi - \overline{O}i \right| + \left| Oi - \overline{O}i \right| \right)^2} \right] \quad (11)$$

In Eqs. 7-12, the P_i and O_i are the estimated and observed i^{th} values, respectively. Also, $\overline{P}$ and $\overline{O}$ are the mean of estimated and observed values, respectively.

In addition, graphical evaluation of the model performances was also performed. The Taylor diagram, which is a graphical representation of the observed and predicted data, was used to check the accuracy of the used models (Taylor, 2001) together with the scatter and violin plot.

Results

The performance of the hybrid LSTM-FFNN model is comprehensively evaluated for daily ET values at Babolsar and Anzali stations located in Iran using the inputs from 1990-2022. The performance of the hybrid LSTM-FFNN model is compared with the standalone LSTM model.

Fig. 5 Evaluation of ET forecasting using Pearson correlation coefficient heat maps for Babolsar and Anzali stations
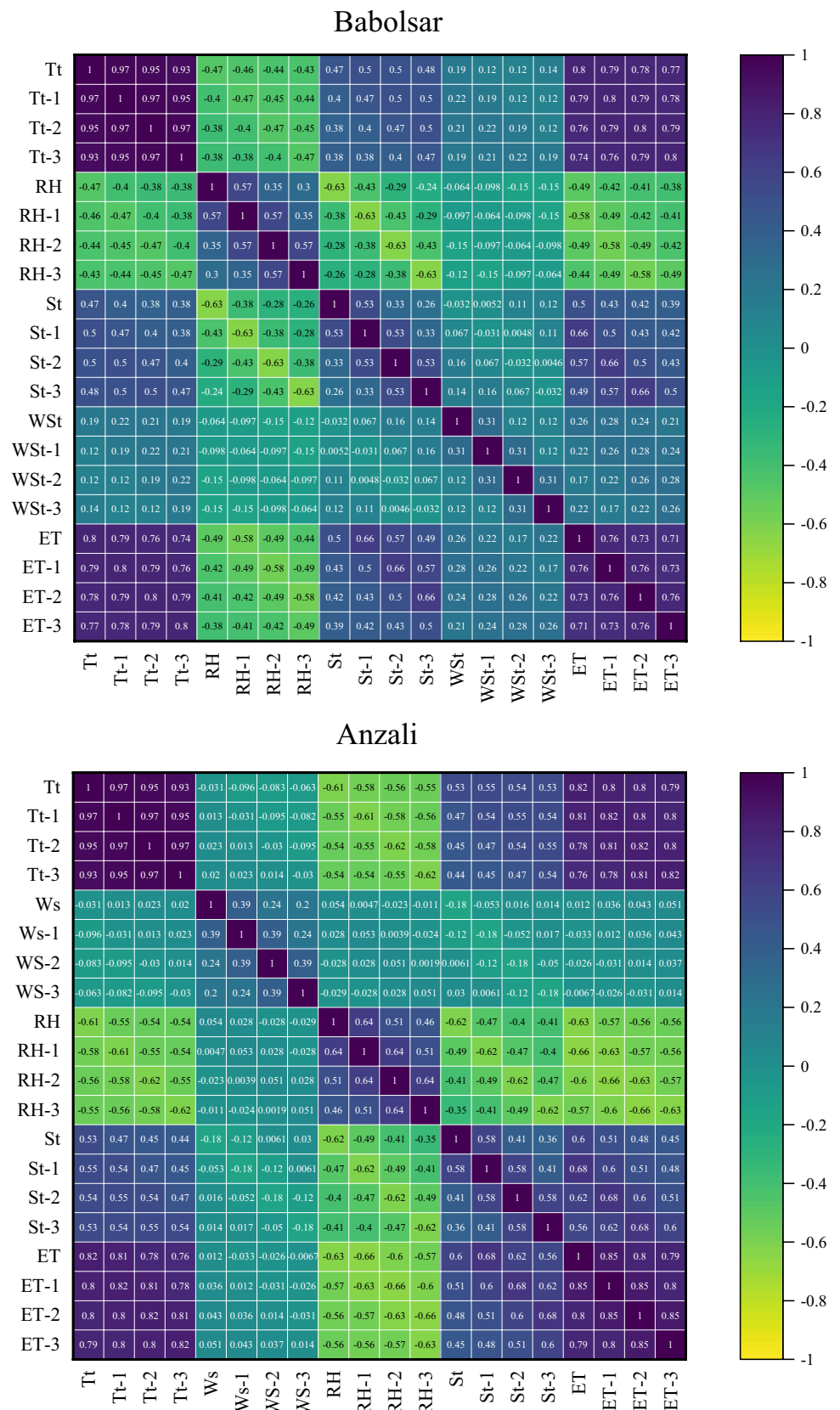
Table 2 Scenarios used in the case study of daily evapotranspiration estimations

Scenario	
1	ET_{t-1}, T_{t-1}
2	ET_{t-1}, ET_{t-2}, T_{t-1}, T_{t-2}
3	ET_{t-1}, ET_{t-2}, ET_{t-3}, T_{t-1}, T_{t-2}, T_{t-3}
4	ET_{t-1}, T_{t-1}, S_{t-1}
5	ET_{t-1}, ET_{t-2}, T_{t-1}, T_{t-2}, S_{t-1}, S_{t-2}
6	ET_{t-1}, ET_{t-2}, ET_{t-3}, T_{t-1}, T_{t-2}, T_{t-3}, S_{t-1}, S_{t-2}, S_{t-3}
7	ET_{t-1}, T_{t-1}, S_{t-1}, RH_{t-1}
8	ET_{t-1}, ET_{t-2}, S_{t-1}, S_{t-2}, RH_{t-1}, RH_{t-2}
9	ET_{t-1}, ET_{t-2}, ET_{t-3}, T_{t-1}, T_{t-2}, T_{t-3}, S_{t-1}, S_{t-2}, S_{t-3}, RH_{t-1}, RH_{t-2}, RH_{t-3}
10	ET_{t-1}, T_{t-1}, S_{t-1}, W_{t-1}
11	ET_{t-1}, ET_{t-2}, S_{t-1}, S_{t-2}, RH_{t-1}, RH_{t-2}, W_{t-1}, W_{t-2}
12	ET_{t-1}, ET_{t-2}, ET_{t-3}, T_{t-1}, T_{t-2}, T_{t-3}, S_{t-1}, S_{t-2}, S_{t-3}, RH_{t-1}, RH_{t-2}, RH_{t-3}, W_{t-1}, W_{t-2}, W_{t-3}
13	ET_{t-1}, ET_{t-2}, ET_{t-3}, T_{t-1}, T_{t-2}, T_{t-3}, T_t
14	ET_{t-1}, ET_{t-2}, ET_{t-3}, T_{t-1}, T_{t-2}, T_{t-3}, T_t, S_{t-1}, S_{t-2}, S_{t-3}, S_t
15	ET_{t-1}, ET_{t-2}, ET_{t-3}, T_{t-1}, T_{t-2}, T_{t-3}, T_t, S_{t-1}, S_{t-2}, S_{t-3}, S_t, RH_{t-1}, RH_{t-2}, RH_{t-3}, RH_t
16	ET_{t-1}, ET_{t-2}, ET_{t-3}, T_t, T_{t-1}, T_{t-2}, T_{t-3}, S_t, S_{t-1}, S_{t-2}, S_{t-3}, RH_t, RH_{t-1}, RH_{t-2}, RH_{t-3}, W_t, W_{t-1}, W_{t-2}, W_{t-3}

An evaluation of model performances on the basis of R^2, RMSE, MAE, NS, and WI criteria for Babolsar station is presented in Table 3. Based on the results, the models performed better during the training period than during the test period in all scenarios. In the first scenario, LSTM and FFNN-LSTM models had the highest error compared to other scenarios. According to scenario one, the RMSE values for the LSTM model are 1.62 mm/day during training and 1.65 mm/day during testing, while those for the FFNN-LSTM model are 1.22 mm/day during training and 1.26 mm/day during testing. In the second and third scenarios, all the evaluation criteria for the LSTM model had a constant value, which shows that both scenarios generated similar performance, however, for the FFNN-LSTM model, the error value (RMSE) increased from 1.21 to 1.23 mm/day and the other metrics were almost constant. This shows that ET_{t-3} and T_{t-3} parameters did not have much positive effect on the accuracy of both models. In the 4th scenario, the addition of the S_{t-1} parameter to the inputs caused the error criterion to decrease by 3.7% in the standalone model and by 6.7% in the hybrid model. From the 5th to the 12th scenario, the accuracy of the LSTM model remained constant and the addition of parameters did not produce a noteworthy effect on the accuracy of the standalone model. This lack of change indicates that the variables such as relative humidity and wind speed do not have much effect on the prediction accuracy of ET values for the standalone model. On the other hand, for the combined LSTM-FFNN model, the addition of the S_{t-2} parameter in the 5th scenario decreased the model error from 1.15 to 1.02 mm/day whilst a slight improvement in NS and WI criteria was observed. In the 6th and 7th scenarios, the addition of S_{t-3} and RH_{t-1} parameters did not

increase the accuracy of the combined LSTM-FFNN model and it also increase the RMSE values by about 3.9% reducing the model performances. In the 8th scenario, as in the previous two scenarios, the error of the combined LSTM-FFNN model has increased from 1.07 to 1.11 mm/day and has caused the model's accuracy to decrease by 3.7%. In the 9th scenario, although the R^2 and WI criteria remained constant, the RMSE and NS criteria improved by 8.5% and 5.3%, respectively. On the contrary, in the 10th scenario, the addition of the W_{t-1} input series increased the RMSE by 8.5% and decreased NS by 5.3%, although R^2 and WI criteria remained unchanged. Both the 11th and 12th scenarios showed improved accuracy of the LSTM-FFNN model. In the 11th scenario, the error reduced from 1.11 to 1.05 mm/day and in the 12th scenario the error reduced from 1.05 to 1 mm/day. For the 13th scenario, adding the T_t input series increased the RMSE value by 3.1% and 18.2%, respectively, for both the standalone LSTM and combined LSTM-FFNN models, and other criteria also decreased relatively. On the other hand, in the 14th scenario, the RMSE for the LSTM model decreased from 1.63 to 1.58 mm/day and for the FFNN-LSTM model the decrease was from 1.2 to 1.02 mm/day, which can be attributed to the addition of the S_t input series. The 15th and 16th scenarios had similar accuracy, yet in comparison to the 14th scenario, the error rate of both of the models decreased slightly.

Comparing and evaluating the results in the Babolsar station for the LSTM model also showed that the 5th, 8th, 11th, 15th, and 16th scenarios performed better than other scenarios of the standalone model. The results of this study are in agreement with the results of Roy (2021), who used short-term memory networks to predict one step ahead

Table 3 Statistical evaluation of model performances during the training and testing period at Babolsar station

Model	Scenario	Training					Testing				
		R^2	RMSE (mm/day)	MAE (mm/day)	NS	WI	R^2	RMSE (mm/day)	MAE (mm/day)	NS	WI
LSTM	1	0.76	1.62	1.34	0.47	0.70	0.67	1.65	1.36	0.43	0.66
	2	0.77	1.61	1.33	0.48	0.71	0.69	1.64	1.35	0.44	0.67
	3	0.76	1.61	1.33	0.48	0.71	0.69	1.64	1.35	0.44	0.67
	4	0.86	1.52	1.26	0.52	0.73	0.77	1.58	1.31	0.48	0.70
	5	0.86	1.52	1.26	0.52	0.74	0.77	1.57	1.30	0.48	0.70
	6	0.86	1.52	1.26	0.52	0.72	0.77	1.58	1.31	0.48	0.69
	7	0.86	1.52	1.26	0.52	0.73	0.77	1.58	1.30	0.48	0.70
	8	0.86	1.51	1.25	0.52	0.72	0.77	1.57	1.30	0.48	0.70
	9	0.86	1.52	1.26	0.52	0.71	0.77	1.58	1.30	0.48	0.69
	10	0.86	1.52	1.26	0.52	0.73	0.77	1.58	1.30	0.48	0.70
	11	0.86	1.51	1.25	0.52	0.72	0.77	1.57	1.30	0.48	0.70
	12	0.86	1.52	1.26	0.52	0.71	0.77	1.58	1.30	0.48	0.69
	13	0.79	1.59	1.31	0.48	0.70	0.71	1.63	1.35	0.44	0.67
	14	0.85	1.54	1.27	0.52	0.72	0.77	1.58	1.30	0.48	0.70
	15	0.85	1.51	1.25	0.52	0.72	0.77	1.57	1.30	0.48	0.70
	16	0.86	1.5	1.24	0.53	0.73	0.77	1.57	1.30	0.48	0.70
FFNN-LSTM	1	0.74	1.22	0.89	0.70	0.92	0.67	1.26	0.92	0.67	0.89
	2	0.76	1.17	0.83	0.72	0.93	0.69	1.21	0.86	0.69	0.90
	3	0.77	0.2	0.15	0.71	0.93	0.69	1.23	0.91	0.68	0.90
	4	0.85	1.11	0.79	0.75	0.94	0.77	1.15	0.82	0.72	0.92
	5	0.85	0.99	0.70	0.81	0.96	0.77	1.02	0.73	0.78	0.94
	6	0.85	1.03	0.78	0.79	0.95	0.77	1.06	0.80	0.76	0.93
	7	0.85	1.04	0.73	0.78	0.95	0.77	1.07	0.76	0.76	0.93
	8	0.86	1.08	0.85	0.77	0.95	0.79	1.11	0.87	0.74	0.93
	9	0.86	1	0.74	0.80	0.95	0.79	1.02	0.75	0.78	0.93
	10	0.83	1.06	0.76	0.78	0.95	0.77	1.11	0.79	0.74	0.93
	11	0.85	1.01	0.70	0.80	0.95	0.79	1.05	0.73	0.77	0.93
	12	0.83	0.97	0.70	0.81	0.96	0.79	1	0.72	0.79	0.94
	13	0.74	1.15	0.82	0.73	0.93	0.71	1.2	0.85	0.70	0.90
	14	0.85	1.01	0.72	0.81	0.97	0.79	1.02	0.72	0.78	0.94
	15	0.85	0.98	0.70	0.81	0.97	0.79	1	0.72	0.79	0.94
	16	0.86	0.97	0.68	0.82	0.98	0.79	1	0.72	0.79	0.94

evapotranspiration at Ghazipur station located in Bangladesh during 2004-2019 and showed that the bi-directional LSTM (Bi-LSTM) model with a correlation coefficient of 0.99 and then the LSTM model with a correlation coefficient of 0.69 have high accuracy in predicting evapotranspiration. Both the standalone and the hybrid FFNN-LSTM models showed that the 12th, 15th, and 16th scenarios had the best performance compared to the rest of the scenarios. The performance of the FFNN-LSTM model was higher in these scenarios with $R^2 = 0.79$, RMSE = 1 mm/day, MAE=0.72 mm/day, NS = 0.79, and WI = 0.94.

In addition to statistical evaluation, the graphical evaluation via the scatter plots for the best scenarios of both LSTM and LSTM- FFNN models for Babolsar station are

illustrated in Fig. 6. The scatter plots reveal that the 16th scenario has a slightly higher correlation than the 5th, 8th, 11th, and 15th scenarios registered by the LSTM model. In concurrence with the outcomes of Table 3, the 16th scenario has slightly higher accuracy than the 15th and 12th scenarios for the LSTM-FFNN model (Fig. 6). With that, the combined LSTM-FFNN models registered better performance than the standalone LSTM model.

In the case of Bandar Anzali station, the performance evaluation of the proposed LSTM-FFNN and standalone LSTM models during the test phase are presented in Table 4. A comparison of the first and second scenarios for the LSTM model shows that the R^2 criterion has increased a little and other criterion have remained unchanged. For the 3rd and 4th

Fig. 6 Scatter plots of the best scenarios in the testing period at Babolsar station
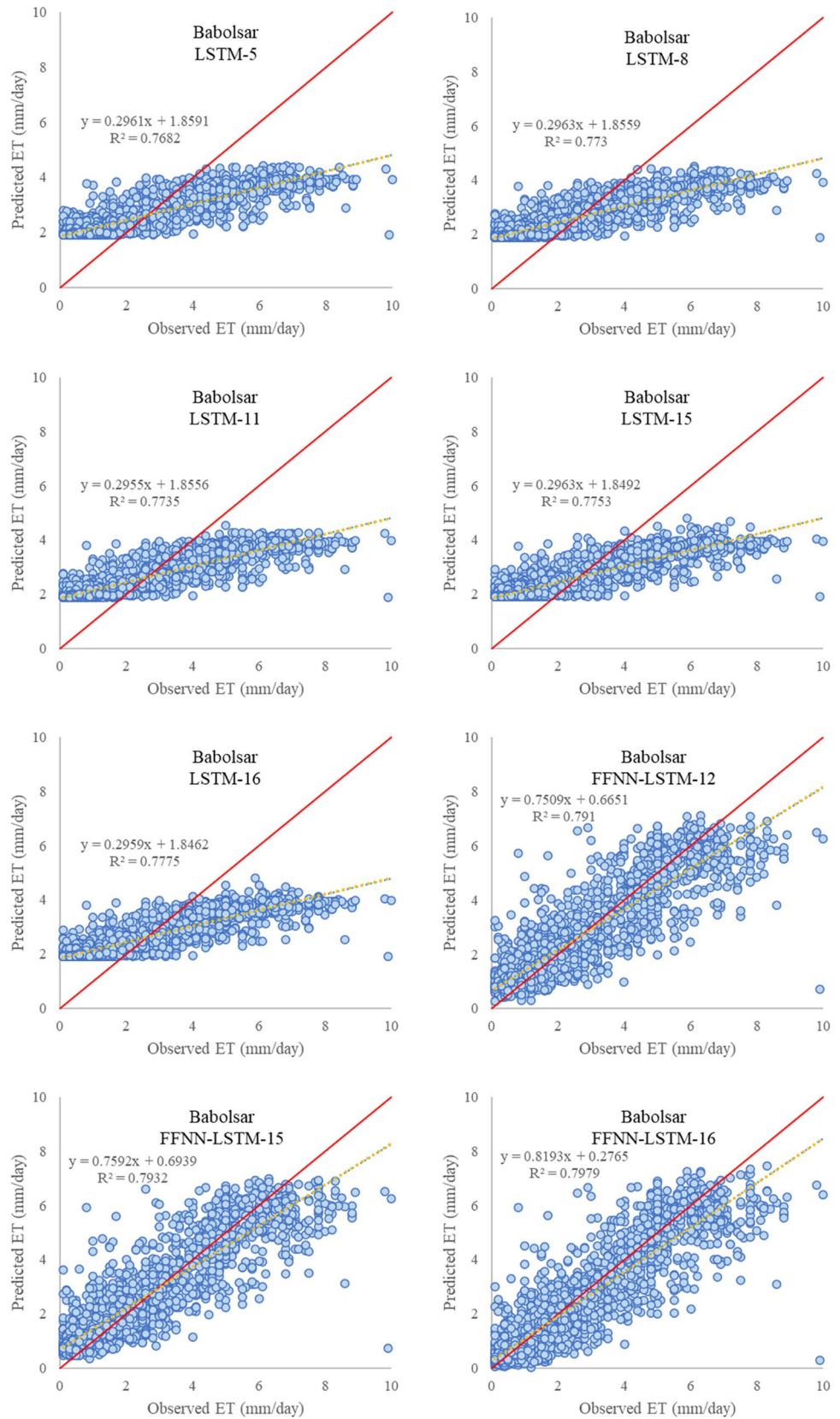
Table 4 Statistical evaluation of model performances during the training and testing period at Bandar Anzali station

Model	Scenario	Training					Testing				
		R^2	RMSE (mm/day)	MAE (mm/day)	NS	WI	R^2	RMSE (mm/day)	MAE (mm/day)	NS	WI
LSTM	1	0.83	1.69	1.45	0.52	0.74	0.76	1.71	1.47	0.49	0.71
	2	0.85	1.68	1.45	0.53	0.74	0.77	1.71	1.47	0.49	0.71
	3	0.85	1.69	1.46	0.51	0.74	0.77	1.71	1.47	0.48	0.71
	4	0.86	1.66	1.43	0.53	0.74	0.79	1.69	1.46	0.50	0.71
	5	0.86	1.66	1.43	0.54	0.73	0.81	1.69	1.45	0.50	0.71
	6	0.88	1.66	1.43	0.54	0.75	0.81	1.69	1.45	0.50	0.72
	7	0.85	1.66	1.43	0.54	0.75	0.79	1.69	1.45	0.50	0.72
	8	0.86	1.66	1.42	0.54	0.75	0.81	1.69	1.45	0.50	0.72
	9	0.85	1.65	1.42	0.54	0.75	0.81	1.68	1.45	0.50	0.72
	10	0.83	1.66	1.42	0.54	0.75	0.79	1.69	1.45	0.50	0.72
	11	0.86	1.66	1.42	0.54	0.75	0.81	1.69	1.45	0.50	0.72
	12	0.85	1.65	1.42	0.54	0.75	0.81	1.68	1.45	0.50	0.72
	13	0.86	1.64	1.41	0.55	0.75	0.81	1.68	1.45	0.51	0.72
	14	0.86	1.63	1.40	0.55	0.76	0.83	1.67	1.44	0.51	0.73
	15	0.88	1.62	1.40	0.55	0.76	0.83	1.66	1.43	0.52	0.73
	16	0.90	1.62	1.39	0.56	0.76	0.83	1.66	1.43	0.52	0.73
FFNN-LSTM	1	0.86	1.1	0.77	0.8	0.96	0.77	1.14	0.79	0.77	0.93
	2	0.86	0.11	0.08	0.8	0.96	0.77	1.15	0.79	0.77	0.93
	3	0.88	0.11	0.08	0.81	0.96	0.79	1.15	0.85	0.77	0.93
	4	0.90	0.09	0.07	0.8	0.96	0.81	1.14	0.90	0.77	0.93
	5	0.90	0.05	0.04	0.81	0.97	0.81	1.08	0.82	0.79	0.94
	6	0.92	1.08	0.75	0.85	0.98	0.83	1.01	0.70	0.82	0.95
	7	0.90	1.09	0.75	0.85	0.97	0.81	1.02	0.70	0.82	0.95
	8	0.90	1.08	0.84	0.81	0.96	0.83	1.12	0.87	0.78	0.94
	9	0.92	0.95	0.66	0.86	0.97	0.83	0.99	0.69	0.83	0.95
	10	0.90	0.99	0.68	0.85	0.97	0.81	1.03	0.70	0.81	0.95
	11	0.92	0.98	0.68	0.87	0.97	0.83	1.01	0.70	0.82	0.95
	12	0.92	0.96	0.67	0.87	0.98	0.83	1	0.70	0.82	0.95
	13	0.90	1.06	0.82	0.8	0.96	0.81	1.11	0.85	0.78	0.94
	14	0.94	0.93	0.66	0.86	0.97	0.85	0.96	0.68	0.84	0.95
	15	0.94	0.93	0.67	0.86	0.98	0.85	0.96	0.68	0.84	0.95
	16	0.96	1.15	0.93	0.79	0.96	0.85	1.18	0.96	0.76	0.93

scenarios, the WI criterion was constant, while the RMSE decreased by 1.2%. On the other hand, the NS indices increased by 4.1%, and the R^2 also increased slightly. The WI index in the 6th scenario showed a slight increase compared to the 5th scenario, while the other parameters did not change. In the 7th and 8th scenarios, only the R^2 increased slightly among all criteria. The comparison between the 9th and 10th scenarios also shows that the addition of the W_{t-1} parameter did not improve the accuracy of the standalone LSTM model (slight increase in the RMSE index) and also a decrease in the R^2 criterion was noted. Also, the examination of the performance of the 11th and 12th scenarios shows that the 12th scenario has performed better by reducing the amount of error to a small extent. For the 13th and 14th

scenarios, the values of all criteria except NS improved by about 1% and the RMSE error values reduced. These results persisted in the last three scenarios (Scenarios 14-16).

For the proposed hybrid LSTM-FFNN model, the evaluation between the first and second scenarios showed that the addition of ET_{t-2} and T_{t-2} parameters to the first scenario caused a slight increase in error while the other metrics remained unchanged. A comparison between the 3rd and 4th scenarios revealed that the error of the 4th scenario decreased compared to the third scenario and reached the value of 1.14 mm/day, and the correlation coefficient R^2 also increased slightly. A comparison between the 5th and 6th scenarios also indicated that the 6th scenario had a better performance registering an increase in R^2, NS, and WI criteria by 2.4%,
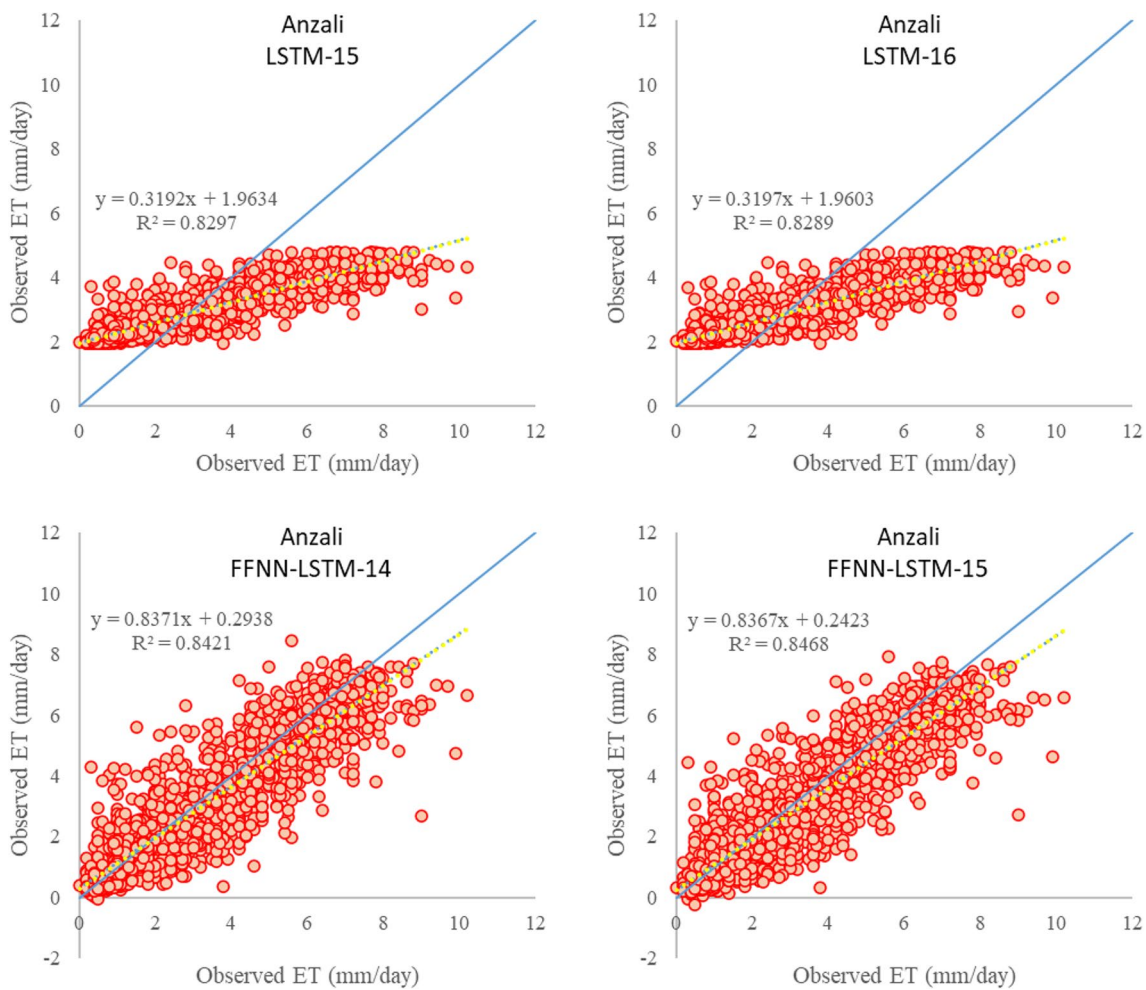
Fig. 7 Scatter plots of the best scenarios in the testing period at Bandar Anzali station

3.7%, and 1.1%, respectively, and a reduction in RMSE, and MAE error criteria by 6.7%, and 15.8%, respectively. On the other hand, the evaluation of the 7th and 8th scenarios showed that the 8th scenario had lower accuracy than the 7th scenario due to an increase in error by 9.4% and a decrease in the NS criterion by 5%. By increasing the RMSE value from 0.99 to 1.03 mm/day, the 10th scenario also showed a decrease in performance accuracy. Examining the performance metrics of the 11th and 12th scenarios showed that the values remained unchanged except for the RMSE, which showed a very slight decrease in the 12th scenario. In comparison to the 13th scenario, the 14th scenario indicated that it has produced a better accuracy by reducing the RMSE, and MAE criterion by 14.5%, and 22.2 respectively, while an increase in R^2, NS, and WI was recorded. The results in the 15th scenario were the same as in the 14th scenario. In contrast, in the 16th scenario, a decrease in the accuracy of the LSTM-FFNN model was noted whereby the RMSE error increased by 20.6% compared to the 15th scenario.

Moreover, an evaluation of all scenarios for LSTM and hybrid LSTM-FFNN models revealed that the 14th and 15th scenarios registered the highest accuracy with the lowest error (RMSE=0.96 mm/day) for the combined LSTM-FFNN model. For the standalone model, the 15th and 16th scenarios had the highest performance (RMSE=1.66 mm/day) compared to other scenarios. In a study, Granata and Di Nunno (2021) predicted evapotranspiration in different climates in the state of Florida using a set of recurrent neural networks including LSTM and nonlinear autoregressive networks with exogenous inputs (NARX). The results showed that in the humid subtropical climate of South Florida, the model based on LSTM performed better than the model based on NARX and showed good accuracy with R^2=0.81. Therefore, in this study, the standalone LSTM model with R^2=0.83 agrees with the study by Granata & Di Nunno (2021). Our study further ascertains that a hybrid modelling approach as the proposed LSTM-FFNN model is able to have better accuracies in comparison to the standalone one with R^2=0.85.
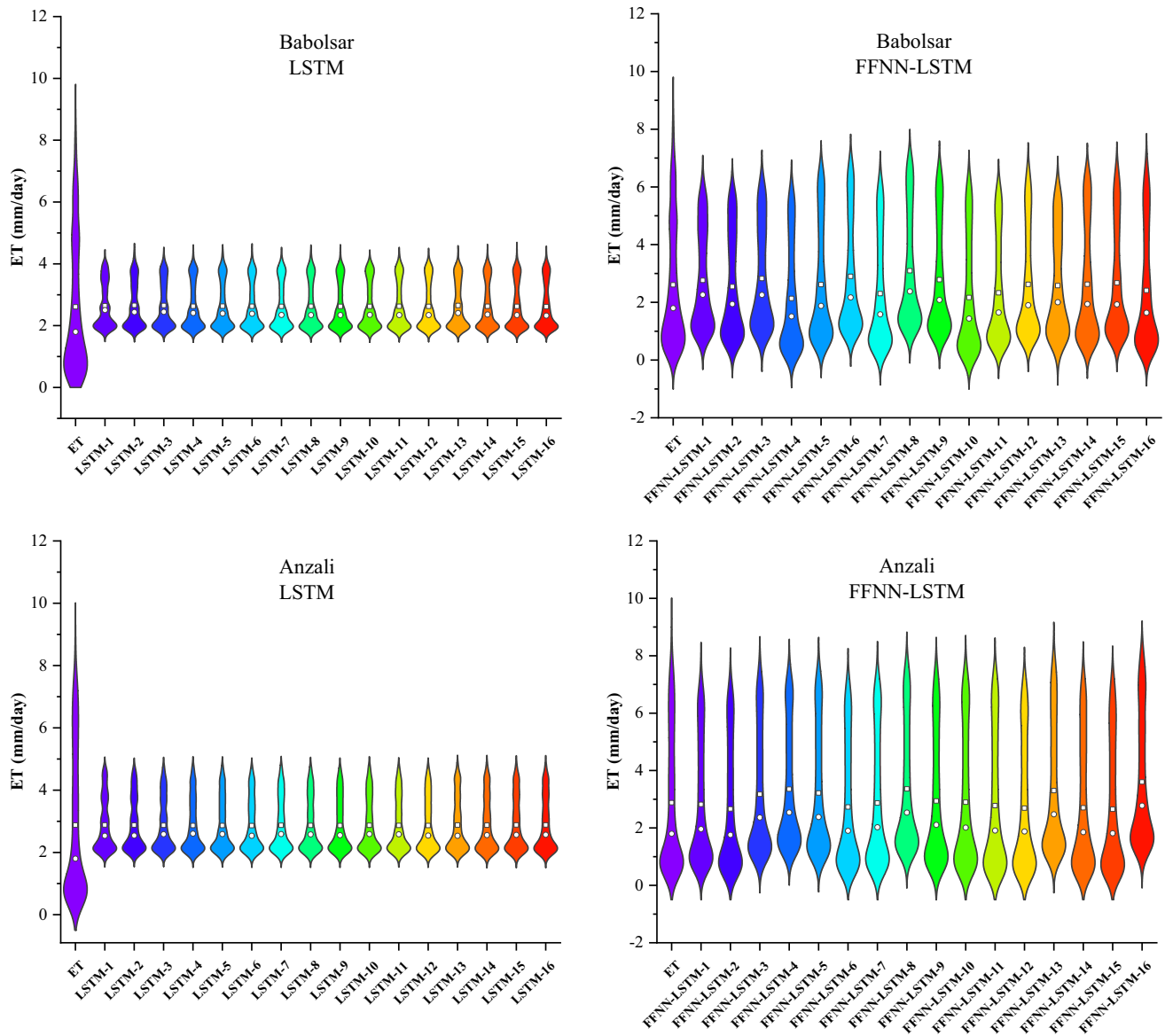
Fig. 8 Violin plots for the best scenario in each model during the testing period

To further attest the better accuracy of the hybrid LSTM-FFNN model, Fig. 7 shows the scatter plots of the best scenarios of standalone LSTM and the hybrid LSTM-FFNN models for the Bandar Anzali station. Although the accuracy in both the 15th and 16th scenarios is almost the same for the standalone LSTM model, the observed and predicted values are more correlated with each other. This important evaluation could only be seen graphically. Also, for the combined hybrid LSTM-FFNN model, the scatter points in the 15th scenario are closer to the 45° ($y=x$) line with the gradient very close to unity ($m = 0.847$) and therefore have higher better accuracy.

Moreover, the violin plots for the best scenarios for respective models at Babolsar and Anzali stations are displayed in Fig. 8. In all the violin plots, the white squares and circles indicate the mean and median of the data series. The results of the LSTM models for the Babolsar station under all scenarios show that the forecasted values ranged between 1.5 to 2.5 mm/day, which are not as the observed ET values are distributed. The comparison of the scenarios also shows that there was the highest number of predicted data around ET=2 mm/day. For the FFNN-LSTM model, the estimated values in Scenario 16 closely followed the observed values ET with a narrow tail (extreme values) and more values concentrated around ET=1 mm/day. This revealed the better performance of the FFNN-LSTM model.

For the Anzali station, the evaluation of the plot related to the standalone LSTM model showed that the highest

dispersion of the LSTM estimated values in the best scenarios was around ET=2.1 mm/day while the observed ET values had the highest dispersion around ET=1 mm/day. Also, even though the scenarios have a mean value almost the same as the observed value, all the scenarios have a higher median than the observed value. However, the combined FFNN-LSTM model again showed a better performance. The comparison of the scenarios showed that the combined FFNN-LSTM model was able to predict the ET values with acceptable accuracy. The violin plots also revealed that the 14th and 15th scenarios were very similar and were closest to the distribution of observed ET values providing the best prediction scenario.

Discussions

Taylor diagrams (Fig. 9) are also excellent at evaluating the performances of respective models. The Taylor diagrams in Fig. 9 present the best-case scenario of both the models during testing period. For Anzali station, a comparison between the models clearly showed that LSTM-FFNN under Scenario-15 had the highest correlation compared to other models. It provided the best forecast, followed by this was LSTM-FFNN under Scenario-14 with slightly lower correlation and standard deviation. Among the standalone model scenarios, LSTM under Scenario-15 had the best performance with the highest correlation. For Babolsar station, the comparison between the combined LSTM-FFNN models showed that although all three scenarios 12, 15, and 16 have the same correlation coefficient, the 16th scenario has a better performance than the competing 15th and 12th scenarios. Among the standalone LSTM model scenarios, all scenarios had similar standard deviations, but the correlation coefficients of scenarios 5, 8, 11, 15, and 16 slightly increased.

Comparing and evaluating the accuracy of scenarios for both stations showed that scenario 15 was the best scenario for both LSTM and hybrid LSTM-FFNN models registering the highest accuracy. Also, the investigation of the effectiveness of the input variables for both stations showed that the sunshine hours were the most important variable in increasing the accuracy of ET estimation, which significantly increased the accuracy of the models. The importance of the sunshine hours for daily ET estimations has been highlighted in other studies as well. Petković et al. (2015) determined the most influential weather parameters on reference evapotranspiration estimation with the adaptive neuro-fuzzy inference system (ANFIS) method. For this research, the weather data set from 12 meteorological stations in Serbia between 1980 and 2010 was used. The results indicated that sunshine hour is the most influential single parameter for estimating evapotranspiration (RMSE = 0.4398 mm/day). In another study, Biazar et al. (2019) analyzed the sensitivity
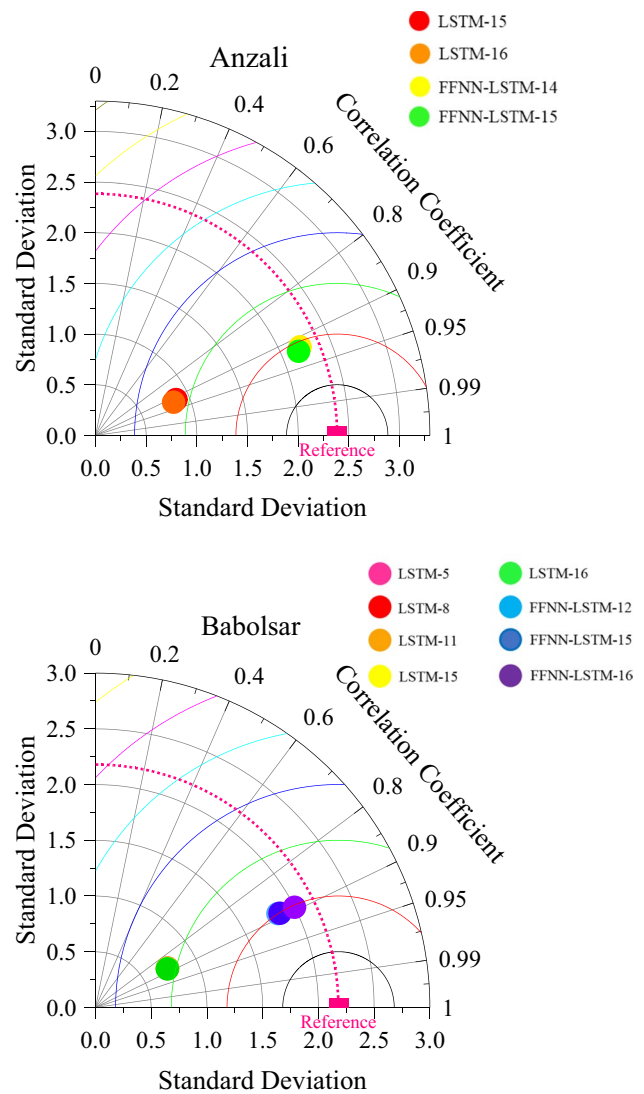

Fig. 9 Taylor diagrams for the best modelling Scenarios at both study sites

of crop evapotranspiration in a humid region in the north of Iran. The results showed that the most important evapotranspiration parameter at Lahijan station was sunshine hours. One limitation of the current study was the unavailability of solar radiation data, hence was not used in this study. It is suggested solar radiation could be used as one of the inputs to the models in further studies.

The Wilcoxon signed-rank test is a non-parametric statistical test that compares two paired samples from a single population (Taheri & Hesamian, 2013). The Wilcoxon signed-rank test was conducted to compare the performance of two different models: LSTM and FFNN-LSTM in estimating ET_0 at two stations, Anzali (Table 5) and Babolsar (Table 6). In summary, the Wilcoxon signed-rank test results clearly indicate that the FFNN-LSTM model consistently and significantly outperforms the

Table 5 Wilcoxon signed-rank test for Anzali station

Scenario	+/-	LSTM						FFNN-LSTM					
		Ranks			Test statistics			Ranks				Test statistics	
		N	Mean Rank	Sum Rank	W	Z	P-Value	N	Mean Rank	Sum Rank	W	Z	P-Value
1	Positive	1177	886.21	1043066	957934	-1.65	0.10	1047	924.24	967679	1033321	1.27	0.20
	Negative	823	1163.95	957934				953	1084.28	1033321			
2	Positive	1173	887.50	1041035	959965	-1.57	0.12	826	915.41	756126	1244874	9.46	0.00
	Negative	827	1160.78	959965				1174	1060.37	1244874			
3	Positive	1175	882.57	1037025	963975	-1.41	0.16	1352	1010.85	1366663	634337	-14.18	0.00
	Negative	825	1168.45	963975				648	978.92	634337			
4	Positive	1174	880.46	1033663	967337	-1.28	0.20	1463	1051.37	1538150	462850	-20.82	0.00
	Negative	826	1171.11	967337				537	861.92	462850			
5	Positive	1170	883.71	1033938	967062	-1.29	0.20	1397	1021.96	1427683	573317	-16.54	0.00
	Negative	830	1165.13	967062				603	950.77	573317			
6	Positive	1170	882.74	1032803	968197	-1.25	0.21	912	917.68	836923	1164077	6.33	0.00
	Negative	830	1166.50	968197				1088	1069.92	1164077			
7	Positive	1174	883.90	1037699	963301	-1.44	0.15	1030	980.59	1010012	990988	-0.37	0.71
	Negative	826	1166.22	963301				970	1021.64	990988			
8	Positive	1171	884.61	1035882	965118	-1.37	0.17	1496	1045.52	1564103	436897	-21.82	0.00
	Negative	829	1164.20	965118				504	866.86	436897			
9	Positive	1169	885.59	1035249	965751	-1.35	0.18	1159	966.46	1120122	880878	-4.63	0.00
	Negative	831	1162.16	965751				841	1047.42	880878			
10	Positive	1174	884.18	1038029	962971	-1.45	0.15	1038	995.59	1033420	967580	-1.27	0.20
	Negative	826	1165.82	962971				962	1005.80	967580			
11	Positive	1174	883.26	1036949	964051	-1.41	0.16	862	993.43	856338	1144662	5.58	0.00
	Negative	826	1167.13	964051				1138	1005.85	1144662			
12	Positive	1168	885.91	1034747	966253	-1.33	0.18	779	957.29	745726	1255274	9.86	0.00
	Negative	832	1161.36	966253				1221	1028.07	1255274			
13	Positive	1175	886.27	1041367	959633	-1.58	0.11	1465	1029.21	1507797	493203	-19.64	0.00
	Negative	825	1163.19	959633				535	921.87	493203			
14	Positive	1180	881.93	1040675	960325	-1.56	0.12	833	935.72	779456	1221544	8.56	0.00
	Negative	820	1171.13	960325				1167	1046.74	1221544			
15	Positive	1181	884.02	1044033	956967	-1.69	0.09	755	927.62	700350	1300650	11.62	0.00
	Negative	819	1168.46	956967				1245	1044.70	1300650			
16	Positive	1182	883.05	1043761	957239	-1.67	0.09	1653	1064.40	1759450	241550	-29.38	0.00
	Negative	818	1170.22	957239				347	696.11	241550			

LSTM model in both positive and negative cases at both Anzali and Babolsar stations in all scenarios. These results highlight the effectiveness of the hybrid FFNN-LSTM approach in improving the accuracy of daily ET predictions in these regions. It's worth noting that the "Test statistics" values in the tables (W and Z) indicate the degree of significance, with values significantly deviating from zero implying stronger statistical significance in favor of the FFNN-LSTM model.

Conclusions

The ever-increasing reduction of water resources requires precise information for prudent water resource management decisions. The ET provides valuable information that can aid tasks such as irrigation planning, ultimately contributing to more efficient and effective practices in the agricultural sector. Therefore, this study proposes and evaluates a new combined LSTM-FFNN method to predict daily ET in some coastal areas of Iran. The modelling is performed under 16

Table 6 Wilcoxon signed-rank test for Babolsar station

Scenario	+/-	LSTM						FFNN-LSTM					
		Ranks			Test statistics			Ranks				Test statistics	
		N	Mean Rank	Sum Rank	W	Z	P-Value	N	Mean Rank	Sum Rank	W	Z	P-Value
1	Positive	1238	903.55	1118601	882399	-4.57	0.00	1237	995.74	1231731	769269	-8.95	0.00
	Negative	762	1158.00	882399				763	1008.22	769269			
2	Positive	1235	910.63	1124630	876370	-4.81	0.00	1052	948.91	998253	1002747	0.09	0.93
	Negative	765	1145.58	876370				948	1057.75	1002747			
3	Positive	1236	909.15	1123707	877293	-4.77	0.00	1282	1008.01	1292268	708732	-11.30	0.00
	Negative	764	1148.29	877293				718	987.09	708732			
4	Positive	1208	914.42	1104625	896375	-4.03	0.00	623	794.47	494956	1506044	19.57	0.00
	Negative	792	1131.79	896375				1377	1093.71	1506044			
5	Positive	1212	911.74	1105024	895976	-4.05	0.00	1110	964.61	1070720	930280	-2.72	0.01
	Negative	788	1137.03	895976				890	1045.26	930280			
6	Positive	1212	910.57	1103609	897391	-3.99	0.00	1388	1012.13	1404839	596161	-15.65	0.00
	Negative	788	1138.82	897391				612	974.12	596161			
7	Positive	1212	910.63	1103689	897311	-3.99	0.00	787	857.63	674954	1326046	12.60	0.00
	Negative	788	1138.72	897311				1213	1093.20	1326046			
8	Positive	1210	913.26	1105050	895950	-4.05	0.00	1505	1051.94	1583167	417833	-22.56	0.00
	Negative	790	1134.11	895950				495	844.11	417833			
9	Positive	1214	908.21	1102571	898429	-3.95	0.00	1299	994.07	1291294	709706	-11.26	0.00
	Negative	786	1143.04	898429				701	1012.42	709706			
10	Positive	1215	909.27	1104768	896232	-4.04	0.00	604	819.39	494909	1506091	19.57	0.00
	Negative	785	1141.70	896232				1396	1078.86	1506091			
11	Positive	1211	912.16	1104627	896373	-4.03	0.00	835	865.19	722434	1278566	10.77	0.00
	Negative	789	1136.09	896373				1165	1097.48	1278566			
12	Positive	1210	911.81	1103288	897712	-3.98	0.00	1147	963.17	1104761	896239	-4.04	0.00
	Negative	790	1136.34	897712				853	1050.69	896239			
13	Positive	1234	908.08	1120576	880424	-4.65	0.00	1085	950.38	1031163	969837	-1.19	0.24
	Negative	766	1149.38	880424				915	1059.93	969837			
14	Positive	1218	905.90	1103387	897613	-3.98	0.00	1110	973.23	1080281	920719	-3.09	0.00
	Negative	782	1147.84	897613				890	1034.52	920719			
15	Positive	1219	903.61	1101495	899505	-3.91	0.00	1179	988.84	1165837	835163	-6.40	0.00
	Negative	781	1151.73	899505				821	1017.25	835163			
16	Positive	1215	904.64	1099132	901868	-3.82	0.00	843	919.33	774993	1226007	8.73	0.00
	Negative	785	1148.88	901868				1157	1059.64	1226007			

scenarios and the hybrid LSTM-FFNN is benchmarked with the standalone LSTM method. The correlation analysis of meteorological parameters showed that air temperature had a high correlation with ET values. The outcomes reveal that the models at Babolsar station the LSTM-FFNN models (RMSE=1.57,1 mm/day) had the best performance in the 15th and 16th scenarios while for Bandar Anzali station, the 15th scenario had better performance for both the hybrid LSTM-FFNN model (RMSE=1.66, 0.96 mm/day) and the standalone LSTM models. The evaluation of the two models showed that the hybrid LSTM-FFNN model can significantly improve the prediction accuracy of ET in humid conditions. The use of this LSTM-FFNN hybrid model is suggested to be explored in further studies at varied sites to comprehensively evaluate model performance. There are some limitations to the study. The study focuses on specific coastal areas in Iran, and the model's applicability to other regions with different climatic conditions has not been explored. Data limitations may affect the generalizability of the findings since the accuracy of the models greatly depends on the quality and availability of meteorological data. The optimized hybrid models have significant potential to assist farmers and irrigation planners in making more informed and precise decisions as accurate prediction of the reference evapotranspiration parameter will lead to the design of efficient and high-efficiency irrigation systems.

Data availability The obtained data can be made available on request of interested parties under the condition of approval of the request by the authors of the article.

Declarations

Competing interests The authors declare no competing interests.

References

Achite M, Jehanzaib M, Sattari MT, Toubal AK, Elshaboury N, Wałęga A, Kim T-W (2022) Modern techniques to modeling reference evapotranspiration in a semiarid area based on ANN and GEP models. Water 14(8):1210

Alibabaei K, Gaspar PD, Lima TM (2021) Modeling soil water content and reference evapotranspiration from climate data using deep learning method. Appl Sci 11(11):5029

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration-guidelines for computing crop water requirements-FAO irrigation and drainage paper 56. Fao, Rome 300(9):D05109

Bayram S, Çıtakoğlu H (2023) Modeling monthly reference evapotranspiration process in Turkey: application of machine learning methods. Environ Monit Assess 195(1):67

Biazar SM, Dinpashoh Y, Singh VP (2019) Sensitivity analysis of the reference crop evapotranspiration in a humid region. Environ Sci Pollut Res 26:32517–32544

Chang LL, Dwivedi R, Knowles JF, Fang YH, Niu GY, Pelletier JD, Meixner T (2018) Why do large-scale land surface models produce a low ratio of transpiration to evapotranspiration? J Geophys Res: Atmospheres 123(17):9109–9130

Chen Z, Zhu Z, Jiang H, Sun S (2020) Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. J Hydrol 591:125286

Citakoglu H (2021) Comparison of multiple learning artificial intelligence models for estimation of long-term monthly temperatures in Turkey. Arab J Geosci 14:1–16

Citakoglu H, Cobaner M, Haktanir T, Kisi O (2014) Estimation of monthly mean reference evapotranspiration in Turkey. Water Resour Manag 28:99–113

Cobaner, M., Citakoğlu, H., Haktanir, T., & Kisi, O. (2017). Modifying Hargreaves–Samani equation with meteorological variables for estimation of reference evapotranspiration in Turkey. Hydrol Res, 48(2), 480-497.

Coşkun Ö, Citakoglu H (2023) Prediction of the standardized precipitation index based on the long short-term memory and empirical mode decomposition-extreme learning machine models: the case of Sakarya, Türkiye. Phys Chem Earth, Parts A/B/C 131:103418

De Fraiture C, Molden D, Wichelns D (2010) Investing in water for food, ecosystems, and livelihoods: an overview of the

comprehensive assessment of water management in agriculture. Agric Water Manag 97(4):495–501

Demir V, Citakoglu H (2023) Forecasting of solar radiation using different machine learning approaches. Neural Comput Applic 35(1):887–906

Fallah-Ghalhari G, Shakeri F (2023) Assessing the consequences of climate change on potential evapotranspiration in Iran in the coming decades. Arab J Geosci 16(4):225

Ferreira LB, da Cunha FF (2020) Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. Comput Electron Agric 178:105728

Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. Neural Comput 12(10):2451–2471

Granata F, Di Nunno F (2021) Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. Agric Water Manag 255:107040

Hadadi F, Moazenzadeh R, Mohammadi B (2022) Estimation of actual evapotranspiration: a novel hybrid method based on remote sensing and artificial intelligence. J Hydrol 609:127774

Hamdy A, Ragab R, Scarascia-Mugnozza E (2003) Coping with water scarcity: water saving and increasing water productivity. Irrig Drainage: J Int Comm Irrig Drainage 52(1):3–20

Hargreaves GH (1994) Defining and using reference evapotranspiration. J Irrig Drain Eng 120(6):1132–1139

Hmoud Al-Adhaileh M, Waselallah Alsaade F (2021) Modelling and prediction of water quality by using artificial intelligence. Sustainability 13(8):4259

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Hu Z, Bashir RN, Rehman AU, Iqbal SI, Shahid MMA, Xu T (2022) Machine learning based prediction of reference evapotranspiration (ET 0) using IoT. IEEE Access 10:70526–70540

Hussain S, Ahmad A, Wajid A, Khaliq T, Hussain N, Mubeen M, Awais M (2020) Irrigation scheduling for cotton cultivation. Agronomy, Crop Protection, and Postharvest Technologies, Cotton Production and Uses, pp 59–80

Jia W, Zhang Y, Wei Z, Zheng Z, Xie P (2023) Daily reference evapotranspiration prediction for irrigation scheduling decisions based on the hybrid PSO-LSTM model. PLoS One 18(4):e0281478

Kushwaha NL, Rajput J, Sena D, Elbeltagi A, Singh D, Mani I (2022) Evaluation of data-driven hybrid machine learning algorithms for modelling daily reference evapotranspiration. Atmosphere-Ocean 60(5):519–540

Landeras G, Ortiz-Barredo A, López JJ (2008) Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (northern Spain). Agric Water Manag 95(5):553–565

Mahesh B (2020) Machine learning algorithms-a review. Int J Sci Res (IJSR)[Internet] 9:381–386

Mehdizadeh S (2020) Using AR, MA, and ARMA time series models to improve the performance of MARS and KNN approaches in monthly precipitation modeling under limited climatic data. Water Resour Manag 34:263–282

Mehdizadeh S, Mohammadi B, Pham QB, Duan Z (2021) Development of boosted machine learning models for estimating daily reference evapotranspiration and comparison with empirical approaches. Water 13(24):3489

Mohammadi B (2023) Modeling various drought time scales via a merged artificial neural network with a firefly algorithm. Hydrology 10(3):58

Nguyen H-P, Liu J, Zio E (2020) A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by tree-structured Parzen estimator and

applied to time-series data of NPP steam generators. Appl Soft Comput 89:106116

Petković D, Gocic M, Trajkovic S, Shamshirband S, Motamedi S, Hashim R, Bonakdari H (2015) Determination of the most influential weather parameters on reference evapotranspiration by adaptive neuro-fuzzy methodology. Comput Electron Agric 114:277–284

Reyes-Gonzalez A (2017) Using remote sensing to estimate crop water use to improve irrigation water management. South Dakota State University

Roy DK (2021) Long short-term memory networks to predict one-step ahead reference evapotranspiration in a subtropical climatic zone. Environ Process 8:911–941

Sabanci D, Yurekli K, Comert MM, Kilicarslan S, Erdogan M (2023) Predicting reference evapotranspiration based on hydro-climatic variables: comparison of different machine learning models. Hydrol Sci J:1–14

Sun Z, Di L, Fang H (2019) Using long short-term memory recurrent neural network in land cover classification on Landsat and cropland data layer time series. Int J Remote Sens 40(2):593–614

Svozil D, Kvasnicka V, Pospichal J (1997) Introduction to multilayer feed-forward neural networks. Chemom Intell Lab Syst 39(1):43–62

Taheri S, Hesamian G (2013) A generalization of the Wilcoxon signed-rank test and its applications. Stat Pap 54:457–470

Talebi H, Samadianfard S, Kamran KV (2023) Investigating the roles of different extracted parameters from satellite images in improving the accuracy of daily reference evapotranspiration estimation. Appl Water Sci 13(2):59

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res: Atmospheres 106(D7):7183–7192

Todd RW, Evett SR, Howell TA (2000) The Bowen ratio-energy balance method for estimating latent heat flux of irrigated alfalfa evaluated in a semi-arid, advective environment. Agric For Meteorol 103(4):335–348

Uncuoglu E, Citakoglu H, Latifoglu L, Bayram S, Laman M, Ilkentapar M, Oner AA (2022) Comparison of neural network, Gaussian regression, support vector machine, long short-term memory, multi-gene genetic programming, and M5 trees methods for solving civil engineering problems. Appl Soft Comput 129:109623

Wang W, Huo Z, Rong Y, Wang C, Zhang C, Wang X (2023) A novel water use efficiency model based on stomata coupling crop growth and farmland water cycle processes in arid area. J Hydrol 617:128974

Wanniarachchi S, Sarukkalige R (2022) A review on evapotranspiration estimation in agricultural water management: past, present, and future. Hydrology 9(7):123

Willmott CJ (1981) On the validation of models. Phys Geogr 2(2):184–194

Yamaç SS, Todorovic M (2020) Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. Agric Water Manag 228:105875

Zouzou Y, Citakoglu H (2023) General and regional cross-station assessment of machine learning models for estimating reference evapotranspiration. Acta Geophysica 71(2):927–947