IoT and Ensemble Long-Short-Term-Memory-Based Evapotranspiration Forecasting for Riyadh

Muhammad Asif Nauman [1], Mahlaqa Saeed [2], Oumaima Saidani [3,*], Tayyaba Javed [4], Latifah Almuqren [3], Rab Nawaz Bashir [5] and Rashid Jahangir [5]

[1] Department of Computer Science, University of Engineering and Technology, Lahore 54890, Pakistan; asif.nauman.uet@gmail.com
[2] Department of Computer Science, University of South Asia, Lahore 53400, Pakistan; mahlaqa.saeed@usa.edu.pk
[3] Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; laAlmuqren@pnu.edu.sa
[4] Department of Computer Science, Barani Institute of Information Technology, Rawalpindi 46604, Pakistan; tayyaba@biit.edu.pk
[5] Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Vehari 61100, Pakistan; rabnawaz@cuivehari.edu.pk (R.N.B.); rashidjahangir@cuivehari.edu.pk (R.J.)
* Correspondence: ocsaidani@pnu.edu.sa

Abstract: Evapotranspiration (ET) is the fundamental component of efficient water resource management. Accurate forecasting of ET is essential for efficient water utilization in agriculture. ET forecasting is a complex process due to the requirements of large meteorological variables. The recommended approach is based on the Internet of Things (IoT) and an ensemble-learning-based approach for meteorological data collection and ET forecasting with limited meteorological conditions. IoT is part of the recommended approach to collect real-time data on meteorological variables. The daily maximum temperature (T), mean humidity (Hm), and maximum wind speed (Ws) are used to forecast evapotranspiration (ET). Long short-term memory (LSTM) and ensemble LSTM with bagged and boosted approaches are implemented and evaluated for their accuracy in forecasting ET values using meteorological data from 2001 to 2023. The results demonstrate that the bagged LSTM approach accurately forecasts ET with limited meteorological conditions in Riyadh, Saudi Arabia, with the coefficient of determination (R^2) of 0.94 compared to the boosted LSTM and off-the-shelf LSTM with R^2 of 0.91 and 0.77, respectively. The bagged LSTM model is also more efficient with small values of root mean squared error (RMSE) and mean squared error (MSE) of 0.42 and 0.53 compared to the boosted LSTM and off-the-shelf LSTM models.

Keywords: ensemble learning; long short-term memory (LSTM); evapotranspiration (*ET*); boosted LSTM; bagged LSTM

1. Introduction

Evapotranspiration (*ET*) is a significant process of the water cycle [1]. The objective of smart agriculture is to optimize resource management and maximize crop yields with efficient use of resources, especially irrigation water [2]. The smart irrigation system is the basis for achieving high yields from crops with scarce irrigation water resources. The implementation of irrigation water according to *ET* is the foundation of efficient irrigation water. The global scarcity of freshwater resources emphasizes the urgent need for efficient utilization of water resources. The efficiency of the irrigation system is highly dependent on the *ET* process. The accurate forecasting of *ET* can result in better agricultural management through water resource planning and drought management [3,4].

ET is the transformation of water into vapor, called evaporation, and the release of water vapor from the surface of plants, called transpiration [5]. *ET* is crucial in maintaining

the water balance between soil and atmosphere [6]. *ET* helps to balance water levels on the Earth's surface and underground [7]. *ET* maintains the overall hydrological balance in ecosystems. Meteorological conditions like daily maximum temperature (*T*), mean humidity (*Hm*), carbon dioxide (CO_2) concentration, air moisture level, and maximum wind speed (*Ws*) significantly affect the *ET* [8–10]. The major meteorological factors affecting *ET* are illustrated in Figure 1.

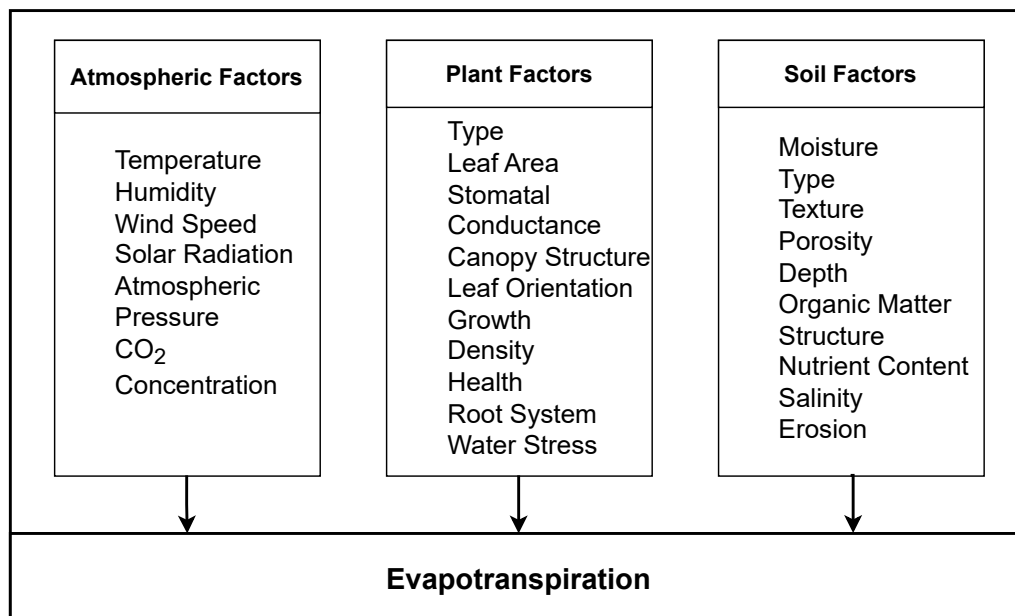| Atmospheric Factors	Plant Factors	Soil Factors
Temperature Humidity Wind Speed Solar Radiation Atmospheric Pressure CO_2 Concentration	Type Leaf Area Stomatal Conductance Canopy Structure Leaf Orientation Growth Density Health Root System Water Stress	Moisture Type Texture Porosity Depth Organic Matter Structure Nutrient Content Salinity Erosion

Evapotranspiration

Figure 1. Factors affecting evapotranspiration (*ET*).

The Penman–Montieth (PM) method is an accurate method of *ET* calculation and is also approved by the Food and Agriculture Organization (FAO) of the United Nations (UN) for *ET* calculations [11]. The PM method is expressed by Equation (1) [12]. The PM method is a widely recognized method of *ET* calculations using the meteorological data of a location. PM calculates the evaporation and transpiration rate for specific meteorological conditions. The PM method is the most accurate method of *ET* calculation [13]. The major issue of the PM method is its requirement for the large number of meteorological variables that make it difficult to use in hydrological cycles [11].

$$ET = \frac{0.408m(R - G) + c\frac{900}{Tm+273}Ws(es - ea)}{m + c(1 + 0.0.34Ws)} \tag{1}$$

where *ET* depicts the evapotranspiration in mm/day, m depicts the slope of the vapor pressure curve in kPa/°C, R depicts the net radiation at the crop surface in MJ/m²/day, G depicts the soil heat flux density in MJ/m²/day, c depicts the psychrometric constant in kPa/°C, Tm depicts the mean daily air temperature in °C, Ws depicts the wind speed at 2 m height in m/s, es depicts the saturated vapor pressure in kPa, and ea depicts the actual vapor pressure in kPa. The *ET* determination with a large number of meteorological inputs is a complex process [14]. The acquisition of large amounts of data are costly and time-consuming. Simplification of *ET* forecasting with limited meteorological data are important for the practical implementation of *ET*-based irrigation water applications.

The Internet of Things (IoT) is widely utilized in modern technology in many applications. IoT provides improved efficiency, automation, optimization, real-time data collection, and monitoring [1]. IoT also has made revolutionary changes in agriculture, precision farming, energy management, industrial automation, and environmental monitoring [15–18]. IoT can address many irrigation issues faced in traditional agricultural practices. IoT can be utilized in smart irrigation systems [19]. The ability of IoT to capture

and integrate contextual factors can be utilized to provide better forecasting of *ET* rates [20]. IoT devices and sensors can collect real-time data on meteorological conditions to make different contextual decisions in agriculture [21]. IoT systems provide valuable inputs for accurate and precise *ET* forecasting models.

Machine learning (ML) with IoT has a wide range of applications in precision agriculture. ML has many advantages that can provide better suitability and efficiency integrated with IoT context-oriented applications [22]. ML algorithms can handle large volumes of IoT-generated data. ML can identify patterns and relationships between meteorological variables to make decisions [23,24]. ML can be utilized for the accurate forecasting of *ET* [20]. From this perspective, this paper proposes an IoT and ensemble learning-based approach for the forecasting of *ET* in Riyadh. Three meteorological variables, daily maximum temperature (*T*), mean humidity (*Hm*), and maximum wind speed (*Ws*), are considered for *ET* forecasting in the proposed approach. The data are collected through the proposed long-range-wireless-area-network (LoRaWAN)-enabled IoT architecture. Long-short-term-memory (LSTM)-based ensemble ML approaches are part of the proposed solution to assess the performance of ML models in handling the complex relationship between meteorological conditions and *ET*. The main contributions of the study are as follows:

- LoRaWAN-enabled IoT architecture to sense meteorological conditions is proposed and implemented to predict accurate *ET* according to real-time meteorological conditions.
- Off-the-shelf LSTM and bagged and boosted ensemble LSTM ML approaches are implemented and evaluated to forecast the *ET* values from real-time meteorological data collected through the proposed LoRaWAN-enabled IoT architecture.
- The evaluation of the performance of ensemble LSTM approach and off-the-shelf LSTM ML models for *ET* forecasting for Riyadh in Saudi Arabia.

IoT is effectively used in acquiring the parameters affecting *ET* values. Moreover, ML algorithms with IoT are proven to be very efficient in forecasting *ET* values [22,25]. The accurate forecasting of *ET* can lead to better irrigation and agriculture productivity [26]. IoT and ML-based approaches for *ET* forecasting are discussed in this section.

Hu et al. [20] proposed an ML model using temperature and humidity. They apply the Internet of Things (IoT) architecture for accurately estimating evapotranspiration (*ET*) for efficient irrigation water management and conservation. They used crop field data from Pakistan from 2015 to 2016. The results suggested that the k-nearest neighbors (KNN) model outperformed the other models and achieved 92% accuracy. Moreover, KNN is successful in reducing the root mean squared error (RMSE) and mean absolute error (MAE) by 16% and 3%, respectively.

Nawandar et al. [27] proposed an artificial-neural-network (ANN)-based model to forecast *ET* for maximum crop yield considering water loss in the irrigation system. The proposed approach is compared with the PM method with fewer inputs. The proposed model successfully produces comparable results to the PM method, with a maximum error of only 0.4 mm day^{-1} in the forecasting of *ET*.

Bellido-Jiménez et al. [28] proposed an ML-based approach for the accurate forecasting of *ET* for regions with limited water resources. The methodology considered two inputs, EnergyT and Hourmin. It was observed that multilayer perceptron (MLP) and extreme learning machine (ELM) performed better across all sites with a determination factor greater than 0.89.

Bashir et al. [14] proposed an ensemble ANN-based approach for estimating *ET* for water conservation in agriculture. They implemented three models named Model-1 (ensemble model), Model-2 (ANN model), and Model-3 (ANN model temperature input only). Model-1 outperformed other models and achieved an accuracy of 91.44%. The predictions from Model-1 also exhibit a stronger correlation with the *ET* values obtained through the PM approach. Moreover, it was observed that Model-1 has a Pearson correlation coefficient (r) of 0.996. However, Model-2 and Model-3 show weaker correlations.

Majumdar et al. [29] recommended an IoT and ML-oriented approach for the forecasting of *ET* at each stage of rice growth. For *ET* prediction, individual ML models and

ensemble learning methods are compared. The results show that individual models suffer from high bias and variance, leading to inadequate prediction accuracy. From their research, it is evident that boosting is the most effective approach for *ET* and soil moisture prediction in different growth stages of rice cultivation. It outperforms other models in terms of accuracy, as demonstrated by lower error metrics and higher determination coefficients, to support better water management decisions throughout the crop cycle.

Torres et al. [30] proposed an IoT-based multilevel data fusion approach, Hydra. They predicted irrigation needs based on soil moisture levels and estimated crop *ET* to forecast the appropriate irrigation timing. Moreover, a novel *ET* model was generated using the quadratic support vector machine (SVM) based model. The results indicate that Hydra, with its multilevel data fusion approach, improves sensor accuracy, identifies target events more accurately, and enables better decision-making.

Chen Zhijun et al. [31] assessed the deep learning approaches named deep neural network (DNN), temporal convolution neural network (TCN), and LSTM to forecast daily *ET* using limited climate data in China. The performance of the deep learning approach was compared with the ML model and empirical equations. The results showed that TCN and RF outperformed daily *ET* measurements in China. Jia Weibing et al. [32] introduced two novel hybrid models, combining particle swarm optimization (PSO) with LSTM neural network to forecast *ET* at four stations in China. Jung Dae-Hyun et al. [33] recommended deep-learning approaches for *ET* prediction in tomato greenhouses. The LSTM model yields low RMSE values of 0.00356 in *ET* forecasting. Roy Dilip Kumar et al. [34] proposed the daily and multistep forward *ET* prediction model using LSTM and bi-directional LSTM (Bi-LSTM) models. The LSTM model outperformed the support vector regressor (SVR) for daily *ET* predictions. Yin Juan et al. [35] proposed a hybrid Bi-LSTM approach to predict daily *ET*. The proposed model is successfully integrated into an intelligent irrigation system in China's central Ningxia region with limited meteorological data. Wang Tianteng et al. [36] proposed *ET* for fruit tree farms.

It is concluded that much research is conducted for the forecasting of *ET* in precision agriculture. The crop yield and productivity can be increased by the automation of *ET* monitoring and smart irrigation systems. In this perspective, IoT and ML-based approaches are proven to be very effective. This research proposes an ensemble learning-based model with bagged LSTM and boosted LSTM for *ET* forecasting in Riyadh.

The interactions between meteorological conditions and *ET* are diverse in nature and nonlinear. These relationships between meteorological conditions and *ET* also change from location to location. Therefore, to handle the complexity between meteorological conditions and *ET*, an LSTM-based ML approach is proposed. To deal with the dynamic nature of the problem, there is also a need to handle seasonal variations. Keeping in view the above issues in *ET* forecasting using the LSTM algorithm to handle the complex relationship between the *ET* and meteorological conditions. LSTM can handle seasonal variations and patterns better due to its ability to handle sequential data with memory capabilities. *ET* forecasting with limited meteorological conditions for Riyadh, Saudi Arabia, is proposed.

The subsequent sections of this research article are structured into materials and methods, results, discussion, and conclusion. The materials and methods section illustrates the flow chart, LoRaWAN-enabled IoT architecture, the data set used for the study, and ML implementation. The results and discussion sections evaluate the performance of the model in terms of different metrics. The conclusion section concludes the significant findings and implications of the proposed solution.

2. Materials and Methods

An IoT and ensemble LSTM-oriented ML-based approach for *ET* forecasting in Riyadh, Saudi Arabia, is proposed with limited meteorological conditions. The meteorological conditions in this regard are daily maximum temperature (*T*), mean humidity (*Hm*), and maximum wind speed (*Ws*). The proposed approach utilizes real-time data collected through the proposed LoRaWAN-enabled IoT architecture. The real-time meteorological data are used

to train the off-the-shelf LSTM and ensemble LSTM models with bagging and boosting approaches. The flowchart of the IoT-ML-based proposed approach is presented in Figure 2.
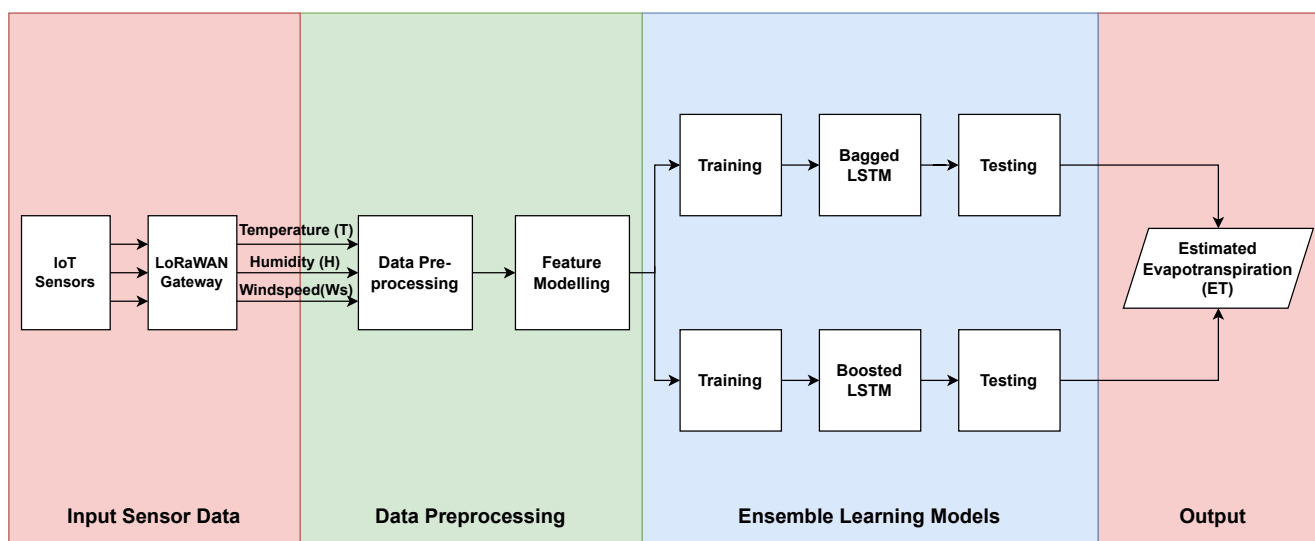

Figure 2. Flowchart of proposed *ET* forecasting.

The main steps of the flow chart for the forecasting of *ET* are as follows:

- First, meteorological data (daily maximum temperature (*T*), mean humidity (*Hm*), and maximum wind speed (*Ws*)) were determined from daily sensed temperature, humidity, and wind speed using LoRaWAN-enabled proposed IoT architecture.
- Second, the real-time data are preprocessed by cleaning and normalization to bring all variables to a similar scale.
- From the real-time collected data, the off-the-shelf LSTM and ensemble LSTM-based bagged and boosted ML models are trained.
- Finally, the performance of both ensemble LSTM models for *ET* forecasting in Riyadh is evaluated.

The simplification of the *ET* forecasting process is essential for its successful implementation in smart agriculture applications [20]. Simplification of *ET* forecasting with fewer meteorological conditions is proposed to deal with the difficulty associated with *ET* forecasting. The *ET* forecast based on daily maximum temperature (*T*), mean humidity (*Hm*), and maximum wind speed (*Ws*) is proposed for *ET* forecasting in Riyadh, Saudi Arabia. An IoT ML-based model is implemented to forecast *ET* using daily maximum temperature (*T*), mean humidity (*Hm*), and maximum wind speed (*Ws*) meteorological conditions, which are expressed by Equation (2).

$$ET \rightarrow f(T, Hm, Ws) \tag{2}$$

where *ET* is the evapotranspiration, T is the daily maximum temperature, Hm is the mean humidity, and Ws is the daily maximum wind speed. The proposed LoRaWAN-enabled IoT architecture for sensing meteorological conditions is shown in Figure 3. The sensing meteorological data from the sensor nodes are transmitted to the server through LoRaWAN. Real-time data on meteorological variables (daily maximum temperature (*T*), mean humidity (*Hm*), and daily maximum wind speed (*Ws*) were calculated from meteorological data collected from Riyadh, Saudi Arabia. For this purpose, IoT architecture with LoRaWAN is utilized. LoRaWAN technology is utilized for wireless communication between the sensors and the IoT gateway without the Internet. The architecture allows the collection of meteorological data from remote areas without involving the Internet. The data collected through the IoT sensors is transmitted to the central server through the LoRaWAN gateway. The meteorological data at the server is processed for the training of ML models. Apart

from these functions, different services are also part of the server, like data storage, data analytics, data management, and server configuration, as shown in Figure 3.
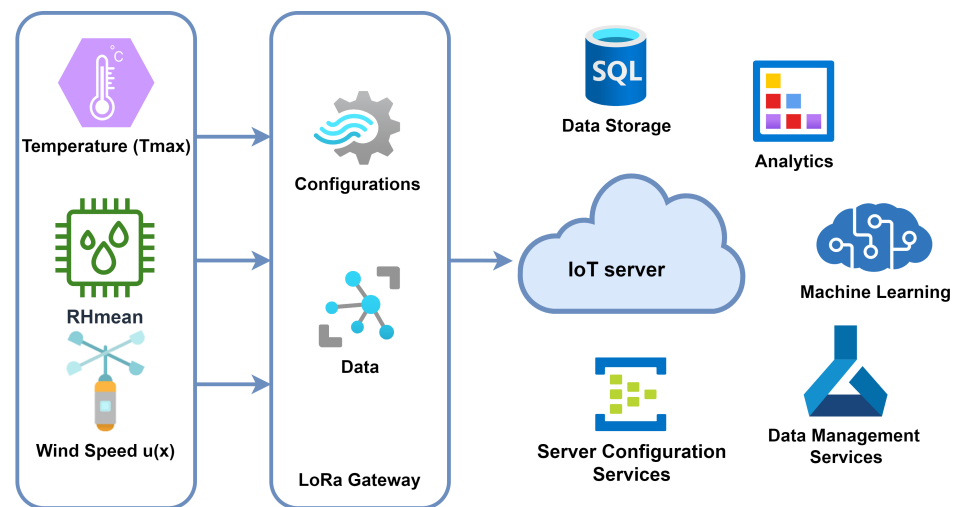

Figure 3. Proposed LoRaWAN-enabled IoT architecture

2.1. Location

The proposed solution intends to forecast the *ET* of Riyadh, Saudi Arabia. Riyadh is situated in the central western region of Saudi Arabia at an altitude of 2000 feet from sea level with longitude and latitude of 24.7136° N, 46.6753° E, respectively, as shown in Figure 4. The meteorological conditions of Riyadh are desert in nature with semi-arid conditions. The desert climatic conditions pose challenges to the efficient use of irrigation water to support agricultural activities in such areas.

Figure 4. Location of Riyadh, Saudi Arabia on world map.

2.2. Data Set

The meteorological data were acquired in Riyadh from the year 2001 to 2023 to forecast the *ET* according to the PM approach. The meteorological data were used to determine the *ET* by PM approach to develop the data set. To simplify the *ET* determination process, this data set is utilized to train the ML model for *ET* forecasting using only daily maximum temperature (*T*), mean humidity (*Hm*), and daily maximum wind speed (*Ws*). The meteorological data selection for *ET* forecasting is made based on their strong correlation with *ET*, shown in Figure 5. The daily maximum temperature (*T*), mean

humidity (*Hm*), and maximum wind speed (*Ws*) are strongly correlated with *ET*; therefore, these meteorological conditions are chosen.
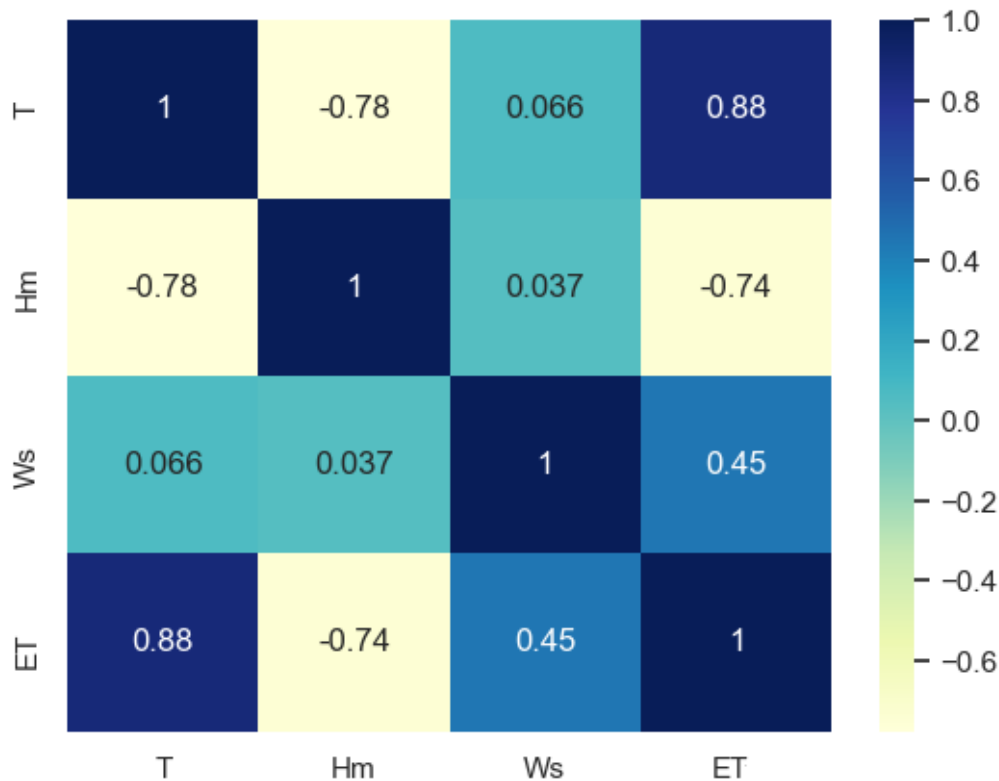

Figure 5. Correlation between climatic conditions and *ET*.

The *ET* distribution in the data set calculated by the PM approach for each month is displayed in Figure 6. It is noted that the *ET* for each month exhibits a similar pattern over each year.
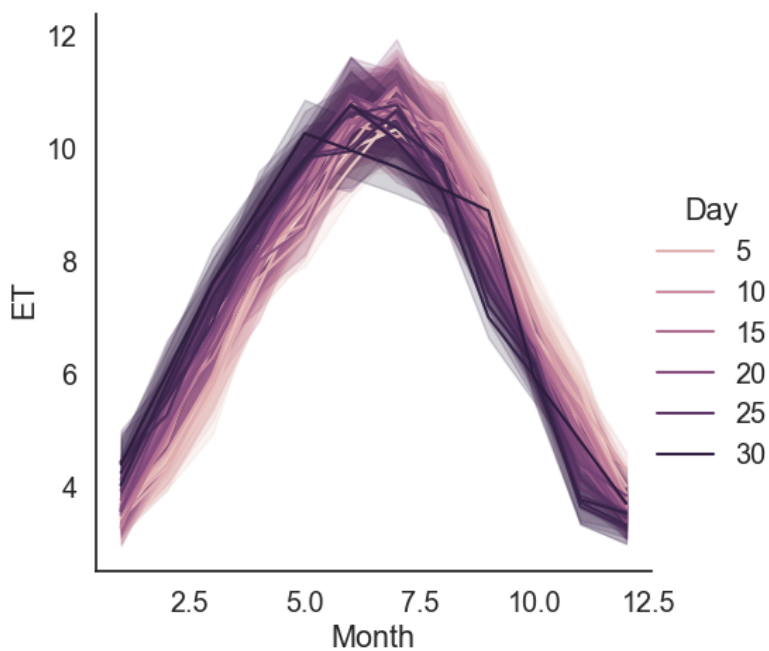

Figure 6. *ET* distributions in data set.

The relationship between *ET* and daily maximum temperature (*T*) for each month from the year 2001 to March of 2023 is shown in Figure 7. The daily maximum temperature (*T*) has a positive relationship with *ET*. The relationship between *ET* and daily mean humidity (*Hm*) for each month from the year 2001 to March of 2023 is shown in Figure 8. The daily mean humidity (*Hm*) has a negative relationship with *ET*. The relationship between *ET* and wind speed (*Ws*) for each month from the year 2001 to March of 2023 is shown in Figure 9. The wind speed (*Ws*) has a positive relationship with *ET*. To deal with the non-linear nature of the problem, an ML-assisted solution is proposed to forecast *ET* with limited meteorological conditions.
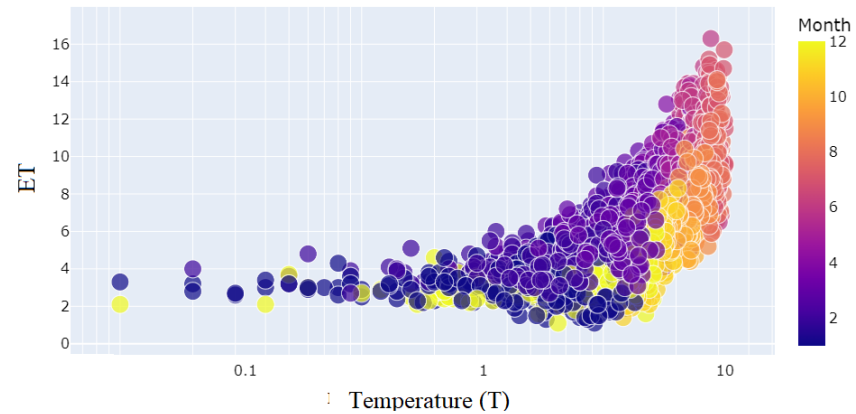

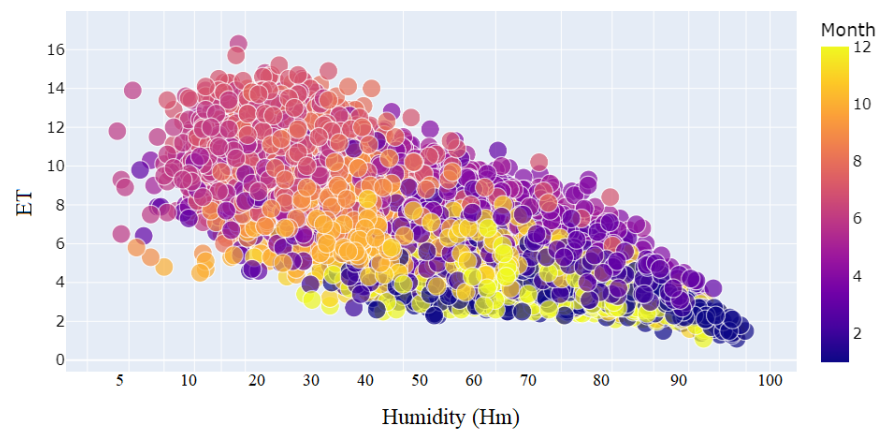**Figure 7.** Distribution of daily maximum temperature (*T*) to months.

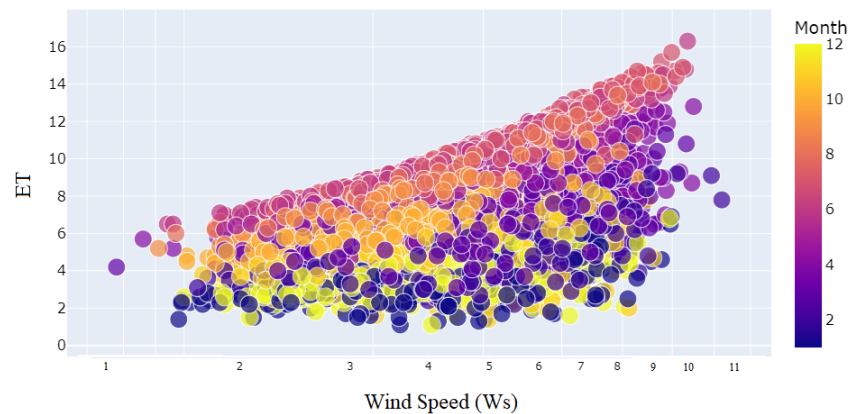**Figure 8.** Distribution of daily mean humidity (*Hm*) to months.

Figure 9. Distribution of daily maximum wind speed (*Ws*) to months.

2.3. Pre-Processing of Data

The collected data are in raw form and have much unnecessary information. The collected data were preprocessed before using them for ML models. The main steps for preprocessing are dealing with missing values, the imputation method, normalization, and standardization of data. The flow chart of the preprocessing is presented in Figure 10. The preprocessing of data can lead to reliable and efficient forecasting of *ET*. The data preprocessing is performed in the following manner.

1. First, the process of handling missing values is performed. The main reasons for these issues are sensor malfunctions or other data collection errors. Such missing values can cause bias in the model's performance. Mean imputation is applied for the handling of missing values. In this method, the missing values are substituted with the average of the values.

2. Second, the outliers are identified and handled from the raw data. Outliers are data values that have a significant deviation from the normal distribution of the variables. In the proposed approach, the sensed data are detected in their respective ranges. Data received out of these ranges is discarded.

3. Third, the normalization process is implemented to make all variables of the data set on a similar scale. It is essential for the performance of the proposed model. The variation in different scale variables can cause inaccurate impacts on the model's performance.

4. Finally, the standardization technique is utilized for the normalization of the data that transforms the variables to have zero mean and unit variance.
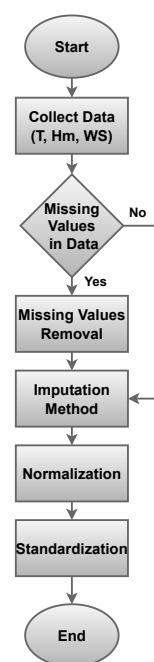
Figure 10. Preprocessing of data.

2.4. Implementation of ML Models

In the proposed approach, the ensemble learning LSTM technique with off-the-shelf LSTM is utilized for accurate *ET* forecasting. The basic architecture of LSTM is presented in Figure 11. For the implementation of the ML model, the Sciket-learn, seaborne, and Matplotlib libraries of Python are used. The data set is divided into training and testing sets using a 70:30 ratio.

The ensemble model tends to enhance the predictive accuracy in complex problems and reduce bias in model training [2,29]. The purpose of the implementation of the ensemble approach is to explore the possibilities of accurate *ET* forecasting with different configurations of the LSTM models.

2.4.1. Bagged LSTM

Bagged LSTM employs an ensemble learning strategy where several LSTM models are trained using subsets of the training data. The bagged LSTM reduces over-fitting and enhances the model's capacity to generalize. The algorithm of bagged LSTM is presented in Algorithm 1, Where, h_i, c_i, and i are hidden states, cell states, and time steps, respectively. The output equation calculates the predicted value y_{pred_i} based on the hidden state h_i. The loss function L_i assesses the disparity between the predicted value and the target y_i. The model parameters $W_f, W_i, W_o, W_c, W_y, b_f, b_i, b_o, b_c, b_y$ are used to minimize the loss with the help of back-propagation and gradient descent. The implementation of the bagged approach is shown in Figure 12 and by Algorithm 1.
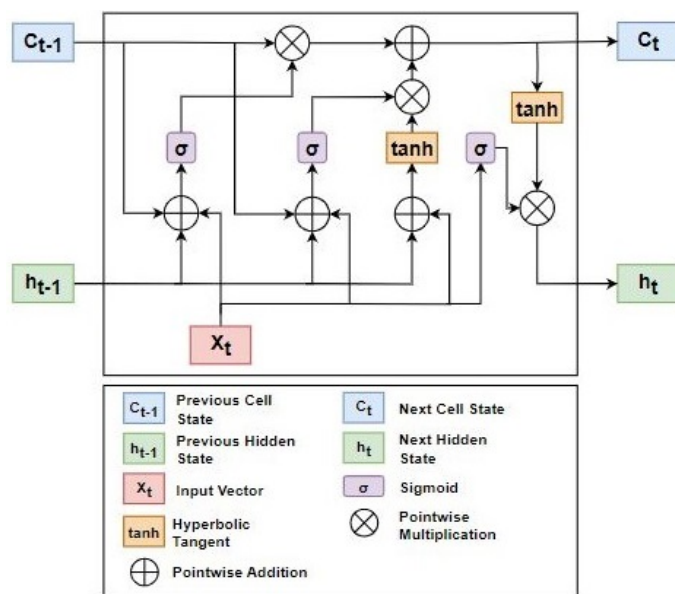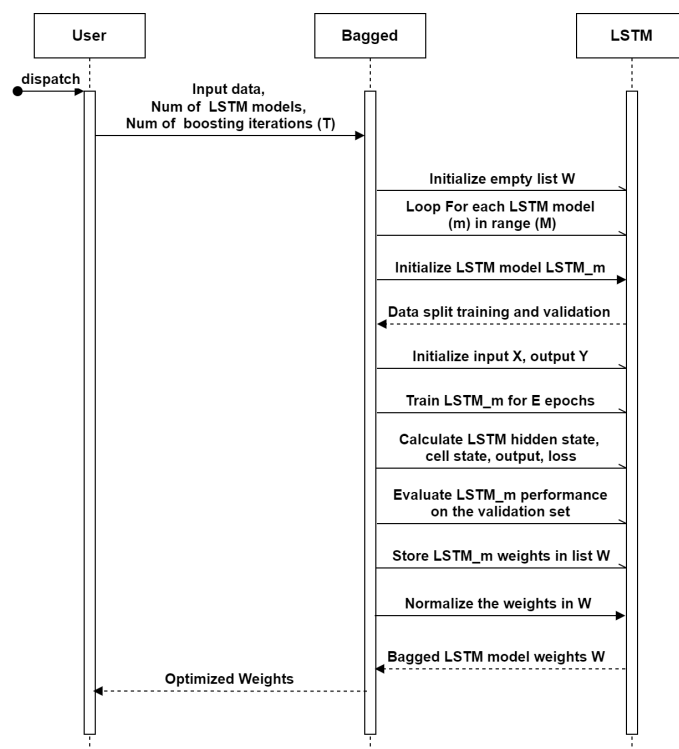

Figure 11. Architecture of LSTM

Figure 12. Sequence diagram of bagged LSTM.

Algorithm 1 Bagged LSTM algorithm for *ET* forecasting

Require:

      Temperature sequence $T = [T_1, T_2, \ldots, T_n]$
      Humidity sequence $Hm = [Hm_1, Hm_2, \ldots, Hm_n]$
      Wind speed sequence $Ws = [Ws_1, Ws_2, \ldots, Ws_n]$
      Output sequence $ET = [ET_1, ET_2, \ldots, ET_n]$
      Number of LSTM models M
      Number of epochs E

Ensure:

      Bagged LSTM model weights $W = [w_1, w_2, \ldots, w_M]$

1: Initialize empty list W
2: **for** m in range(M) **do**
3:      Initialize LSTM model $LSTM_m$
4:      Randomly split the data into training and validation sets
5:      Initialize input sequence $X = [(T_1, Hm_1, Ws_1), (T_2, Hm_2, Ws_2), \ldots, (T_n, Hm_n, Ws_n)]$
6:      Initialize output sequence $Y = [ET_1, ET_2, \ldots, ET_n]$
7:      Train $LSTM_m$ on the training set for E epochs using the following equations:
8:      **for** each epoch e in range(E) **do**
9:           **for** i in range(n) **do**
10:               Set the LSTM input sequence $x_i = (Ws_i, Hm_i, Ws_i)$
11:              Set the LSTM target output $y_i = ET_i$
12:              Calculate the LSTM hidden state h_i and cell state c_i using the LSTM equations:
13:              $f_i = \sigma(W_f \cdot [h_{i-1}, x_i] + b_f)$
14:              $i_i = \sigma(W_i \cdot [h_{i-1}, x_i] + b_i)$
15:              $o_i = \sigma(W_o \cdot [h_{i-1}, x_i] + b_o)$
16:              $\tilde{c}_i = \tanh(W_c \cdot [h_{i-1}, x_i] + b_c)$
17:              $c_i = f_i \cdot c_{i-1} + i_i \cdot \tilde{c}_i$
18:              $h_i = o_i \cdot \tanh(c_i)$
19:              Calculate the LSTM output y_{pred_i} using the output equation:
20:              $y_{pred_i} = W_y \cdot h_i + b_y$
21:              Calculate the LSTM loss L_i using a suitable loss function (e.g., mean squared error):
22:              $L_i = \frac{1}{2}(y_{pred_i} - y_i)^2$
23:              Update the LSTM model parameters $W_f, W_i, W_o, W_c, W_y, b_f, b_i, b_o, b_c, b_y$ using back-propagation and gradient descent
24:      Evaluate the performance of $LSTM_m$ on the validation set
25:      Store the weights of $LSTM_m$ in the list W
26: Normalize the weights in W
27: **return** W

2.4.2. Boosted LSTM

Boosted LSTM is an efficient ensemble learning approach where multiple LSTM models are trained sequentially. In this model, each subsequent LSTM model focuses on the samples that were unclassified by the previous models. This approach leads to improving the overall accuracy. Boosted LSTM combines the forecasting of the individual models using weighted averaging to obtain the final forecasting. The algorithm of boosted LSTM is presented in Algorithm 2 and in the flow chart shown in Figure 13.

Algorithm 2 Boosted LSTM algorithm for *ET* forecasting.

Require:

     Temperature sequence $T = [T_1, T_2, \ldots, T_n]$
     Humidity sequence $Hm = [Hm_1, Hm_2, \ldots, Hm_n]$
     Wind speed sequence $Ws = [Ws_1, Ws_2, \ldots, Ws_n]$
     Output sequence $ET = [ET_1, ET_2, \ldots, ET_n]$
     Number of LSTM models M
     Number of epochs E

Ensure:

     Boosted LSTM model weights $W = [w_1, w_2, \ldots, w_M]$

1: Initialize empty list W
2: **for** m in range(M) **do**
3:      Initialize LSTM model $LSTM_m$
4:      Randomly split the data into training and validation sets
5:      Initialize input sequence $X = [(T_1, Hm_1, Ws_1), (T_2, Hm_2, Ws_2), \ldots, (T_n, Hm_n, Ws_n)]$
6:      Initialize output sequence $Y = [ET_1, ET_2, \ldots, ET_n]$
7:      Train $LSTM_m$ on the training set for E epochs using the following equations:
8:      **for** each epoch e in range(E) **do**
9:         **for** i in range(n) **do**
10:           Set the LSTM input sequence $x_i = (T_i, Hm_i, Ws_i)$
11:           Set the LSTM target output $y_i = ET_i$
12:           Calculate the LSTM hidden state h_i and cell state c_i using the LSTM equations:
13:           $f_i = \sigma(W_f \cdot [h_{i-1}, x_i] + b_f)$
14:           $i_i = \sigma(W_i \cdot [h_{i-1}, x_i] + b_i)$
15:           $o_i = \sigma(W_o \cdot [h_{i-1}, x_i] + b_o)$
16:           $\tilde{c}_i = \tanh(W_c \cdot [h_{i-1}, x_i] + b_c)$
17:           $c_i = f_i \cdot c_{i-1} + i_i \cdot \tilde{c}_i$
18:           $h_i = o_i \cdot \tanh(c_i)$
19:           Calculate the LSTM output y_{pred_i} using the output equation:
20:           $y_{pred_i} = W_y \cdot h_i + b_y$
21:           Calculate the LSTM loss L_i using a suitable loss function (e.g., mean squared error):
22:           $L_i = \frac{1}{2}(y_{pred_i} - y_i)^2$
23:           Update the LSTM model parameters $W_f, W_i, W_o, W_c, W_y, b_f, b_i, b_o, b_c, b_y$ using back-propagation and gradient descent:
24:           $W_f = W_f - \alpha \cdot \frac{\partial L_i}{\partial W_f}$
25:           $W_i = W_i - \alpha \cdot \frac{\partial L_i}{\partial W_i}$
26:           $W_o = W_o - \alpha \cdot \frac{\partial L_i}{\partial W_o}$
27:           $W_c = W_c - \alpha \cdot \frac{\partial L_i}{\partial W_c}$
28:           $W_y = W_y - \alpha \cdot \frac{\partial L_i}{\partial W_y}$
29:           $b_f = b_f - \alpha \cdot \frac{\partial L_i}{\partial b_f}$
30:           $b_i = b_i - \alpha \cdot \frac{\partial L_i}{\partial b_i}$
31:           $b_o = b_o - \alpha \cdot \frac{\partial L_i}{\partial b_o}$
32:           $b_c = b_c - \alpha \cdot \frac{\partial L_i}{\partial b_c}$
33:           $b_y = b_y - \alpha \cdot \frac{\partial L_i}{\partial b_y}$
34:      Evaluate the performance of $LSTM_m$ on the validation set
35:      Calculate the weight w_m based on the validation performance of $LSTM_m$
36:      $w_m = \frac{\text{Validation Accuracy of } LSTM_m}{\text{Total Validation Accuracy}}$
37:      Store the weight w_m in the list W
38: Normalize the weights in W
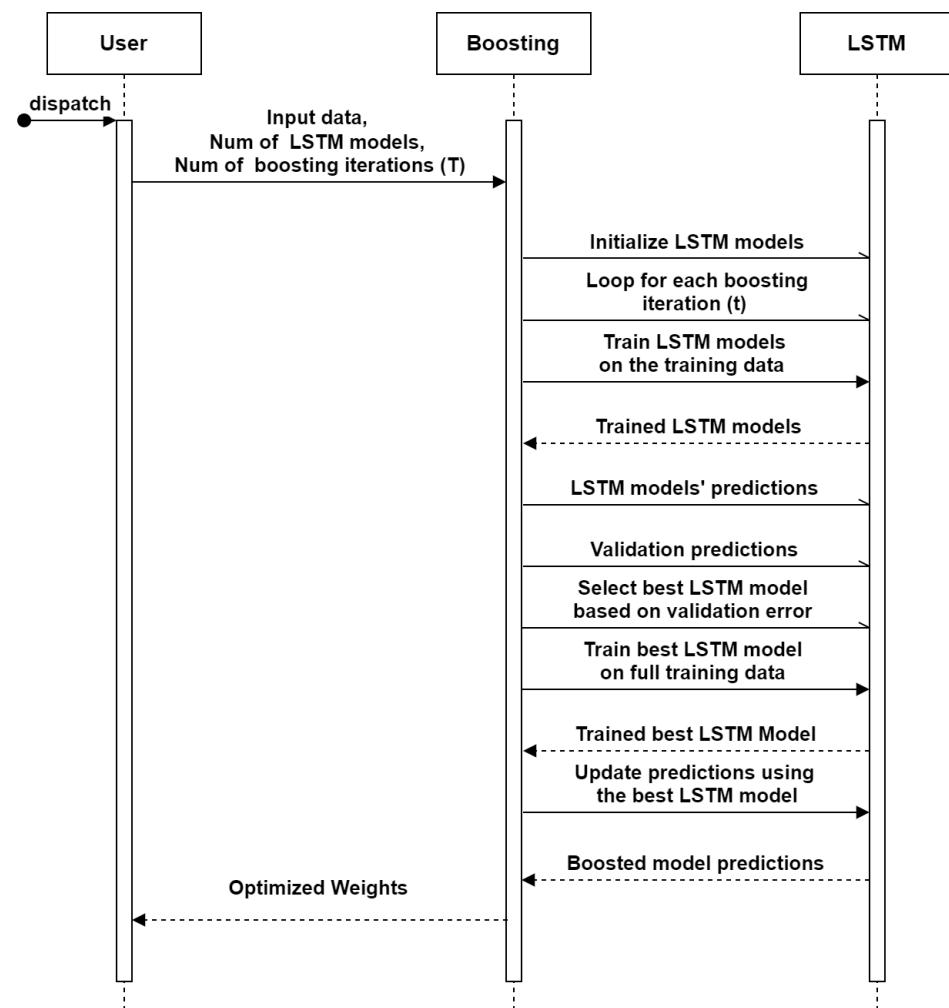39: **return** W

Figure 13. Sequence diagram of boosting LSTM.

3. Results

The performance of ML models for *ET* forecasting is assessed using the 30% data set with the following evaluation metrics:

1. Coefficient of determination (R^2);
2. Pearson correlation coefficient (r);
3. Root mean squared error (RMSE);
4. Mean squared error (MSE).

R^2 is the evaluation metric to judge the accuracy of *ET* forecasting with the ML model. R^2 is expressed by Equation (3), where E_i is the actual *ET* value in the data set at the *i*-th data instance, $\hat{E}_i$ is the forecasted *ET* values by ML model for the *i*-th data instance, $\bar{E}$ is the mean of the actual *ET* values, and n is the total number of *ET* forecasting made by ML model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(E_i - \hat{E})^2}{\sum_{i=1}^{n}(E_i - \bar{E})^2} \tag{3}$$

The Pearson correlation coefficient (r) is used to observe the similarity between the actual *ET* values in the data set and the forecasted *ET* values by the ML model. r is expressed by Equation (4), where X_i is set of meteorological conditions used as input to the ML model to forecast the *ET* for the *i*-th data instance, E_i is the actual *ET* value in the data

set at i-th data instance, $\bar{x}$ is the mean of meteorological conditions, $\bar{E}$ is the mean of actual *ET* values, and n is the total number of *ET* forecasting made with the ML model.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(E_i - \bar{E})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(E_i - \bar{E})^2}} \tag{4}$$

RMSE is used to observe the difference in forecasted *ET* value by the ML model from the actual *ET* value in the data set. RMSE is expressed by Equation (5), where E_i is the actual *ET* values in the data set at the *i*-th data instance, $\hat{E}_i$ represents the forecasted *ET* values for the *i*-th data instance, and n is the total number of *ET* forecasting with the ML model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_i - \hat{E})^2} \tag{5}$$

MSE is the mean of the squared difference between the forecasted *ET* values with the ML model and the actual *ET* values in the data set. MSE is expressed by Equation (6), where E_i is the actual *ET* values in the data set at the *i*-th data instance, $\hat{E}_i$ represents the forecasted *ET* values for the *i*-th data instance and n is the total number of *ET* forecasts made with the ML model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(E_i - \hat{E}_i)^2 \tag{6}$$

The performance metrics of different ML models used for *ET* forecasting are given in Table 1. The ML algorithms considered for *ET* forecasting are off-the-shelf LSTM and ensemble LSTM with a bagged and boosted approach.

Table 1. Performance Matrices of ML Models.

ML Model	R^2	r	MSE	RMSE
Bagged LSTM	0.94	0.97	0.42	0.53
Boosted	0.91	0.95	0.63	0.63
LSTM	0.77	0.87	1.77	1.04

The results indicate that the ensemble LSTM algorithms performed well in forecasting *ET*, with decreasing error compared to the LSTM algorithms. The ensemble LSTM models also outperformed the off-the-shelf LSTM in achieving the lowest MAE and RMSE values of *ET* forecasting. The coefficient of determination (R^2) and Pearson correlation coefficient (r) values were consistently high for the ensemble LSTM algorithms, indicating strong relationships between predicted and observed *ET* values. In the case of the ensemble LSTM approach, the bagged outperformed the boosted approach with high coefficient of determination (R^2) and Pearson correlation coefficient (r) values and low RMSE, and MSE compared to the boosted approach. Overall, the performance of the bagged LSTM model is better in *ET* forecasting using limited meteorological conditions compared to the boosted and off-the-shelf LSTM approaches.

For comparison purposes, the *ET* forecasting with the ML model and actual *ET* values in the data set calculated using the PM method are shown in Figure 14, where it is evident that *ET* forecasting with the ensemble LSTM model from the test data set is more similar to actual *ET* values calculated by PM method compared to the *ET* forecasting with the off-the-shelf LSTM model. As concerns the ensemble LSTM models, the *ET* forecastings with bagged LSTM are more similar to the *ET* by PM method compared to the boosted LSTM.
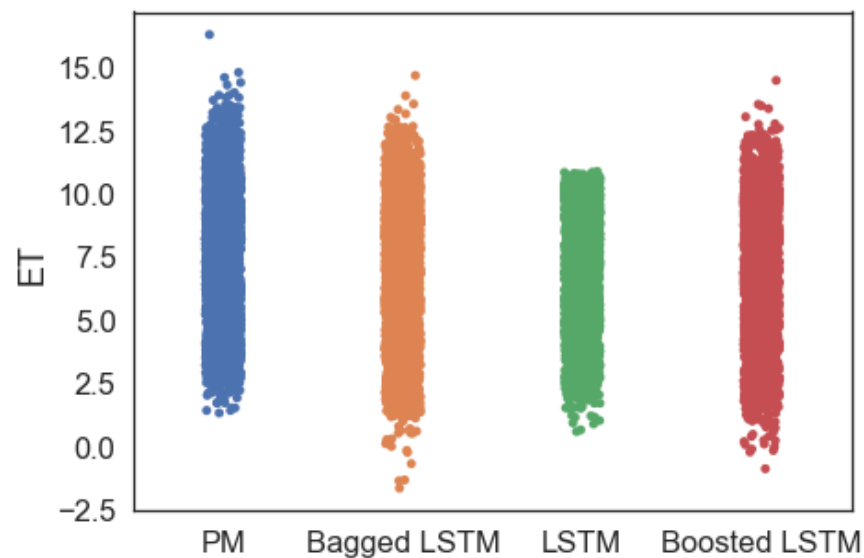
Figure 14. *ET* forecasted with different ML models along with actual *ET* values calculated by PM method in the test data set.

Furthermore, Figure 15 illustrates the difference in *ET* forecasting with different models compared to the actual *ET* values determined by the PM method in the test data set. The absolute difference in actual *ET* values calculated by the PM method (E_i) against the forecasted *ET* ($\hat{E}_i$) for each data instance 'i' in the data set by each ML model is determined by Equation (7) and plotted in 15.

$$Difference = Absolute(E_i - \hat{E}_i) \tag{7}$$

It is observed that the absolute differences in *ET* forecasting with the actual *ET* values with the PM method are low in the case of ensemble LSTM compared to the *ET* forecasting with the off-the-shelf LSTM model. In the case of the ensemble approaches, the differences in *ET* with the bagged LSTM with the *ET* by PM are fewer compared to the boosted approach.
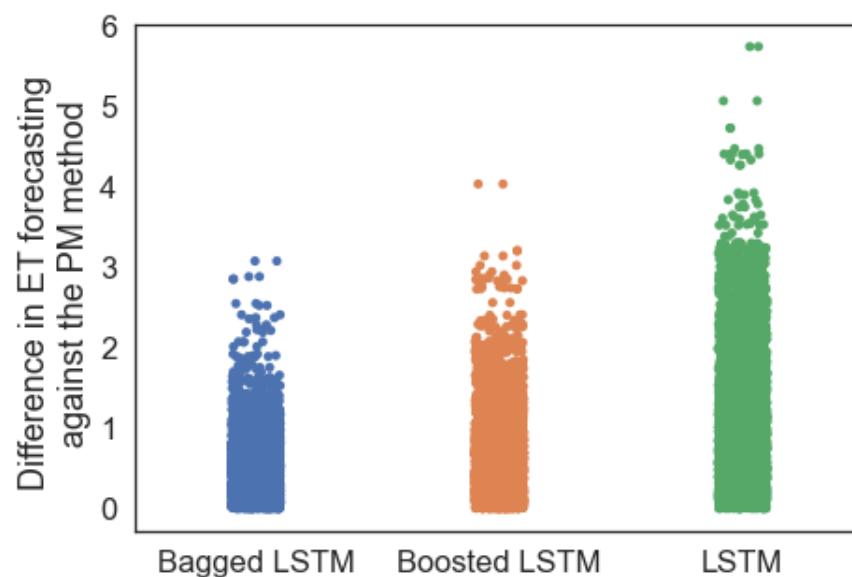

Figure 15. Difference in *ET* forecasted with different ML models compared to actual *ET* calculated using PM method.

4. Discussion

The performance of ensemble models in the forecasting of *ET* is attributed to their ability to handle nonlinear complex relationships between meteorological conditions and *ET*. The ensemble LSTM model is implemented with a bagged and boosted approach. The performance of the *ET* with different configurations of LSTM algorithms is observed in terms of RMSE and MSE in *ET* forecasting. The evaluation of all models on the test data set is made using 30% of the data set. The bagged LSTM ensemble model outperformed with high R^2 of 0.94 and low RMSE and MSE of 0.42 and 0.53 compared to the boosted ensemble LSTM and off-the-shelf LSTM model.

The study also proposed LoRaWAN-enabled architecture for real-time meteorological data sensing and ensemble LSTM models with bagged and boosted LSTM to forecast *ET* with limited meteorological conditions. The LoRaWAN enabled the sensing of data from remote areas without a public network and the Internet. The real-time data enable accurate *ET* forecasting with fewer meteorological conditions. The meteorological conditions considered for the proposed solution are daily maximum temperature (*T*), mean humidity (*Hm*), and wind speed (*Ws*).

ET forecasting with limited meteorological conditions has significant value for water management practices. The proposed solution overcomes the difficulty associated with the implementation of *ET* for large meteorological conditions with high accuracy in *ET* forecasting. The proposed solution is limited in terms of exploiting stacked LSTM and other ensemble deep learning models. Experimentation with other meteorological conditions and the use of other deep learning models is recommended for future work.

5. Conclusions

The study proposed *ET* forecasting in Riyadh, Saudi Arabia, with limited meteorological conditions using LSTM and ensemble LSTM models. The meteorological data from the years 2001 to 2023 for Riyadh, Saudi Arabia, was used for the training and evaluation of ML model in *ET* forecasting. The bagged LSTM outperformed the boosted LSTM and off-the-shelf LSTM in *ET* forecasting with limited meteorological conditions in Riyadh, Saudi Arabia. The proposed solution has several applications in irrigation water management by dealing with the complexity associated with *ET* forecasting. The proposed approach has limitations in terms of the exploration of limited ML algorithms. The exploitation of more ML models and implementation in other parts of the world is recommended for future work.

Author Contributions: Conceptualization, Ideas, and formulation are provided by M.A.N. (first author); Investigation, proofreading, Bugs tracking, and Data Visualization are performed by M.S. (second author); Literature review, Quality assurance, Conceptualization are done by O.S. (third author); Validation, Writing editing, and tracking of changes suggested by reviewers are done by T.J. (fourth Author); Methodology, Testing of code components, Results generation and verifications, writing-review and editing, project management, and Funding acquisition are done by L.A. (fifth author); Project administration, Resources, writing draft, Preparation, creation, and presentation of the published work, specifically writing the initial draft are done by R.N.B. (sixth author); Supervision, Resources Management, Preparation, creation, and presentation of the published work by those from the original research group, specifically critical review, commentary, or revision—including pre- or post-publication stages are done by R.J. (seventh author). All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: This study does not involve any animals or human beings.

Data Availability Statement: Data will be available on request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ANN	Artificial Neural Network
Bi-LSTM	Bi-Directional Long Short-Term Memory
c	Psychrometric Constant
CO_2	Carbon Dioxide
DNN	Deep Neural Network
ELM	Extreme Learning Machine
ET	Evapotranspiration
ea	Actual Vapor Pressure
es	Saturated Vapor Pressure
FAO	Food and Agriculture Organization
G	Soil Heat Flux Density
Hm	Humidity
IoT	Internet of Things
KNN	k-Nearest Neighbors
LoRaWAN	Long-Range Wireless Area Network
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
MLP	Multilayer Perceptron
m	Slope of the Vapor Pressure Curve
PSO	Particle Swarm Optimization
PM	Penman–Montieth
R	Net Radiation at the Crop Surface
r	Pearson Correlation Coefficient
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Squared Error
SVM	Support Vector Machine
SVR	Support Vector Regression
TCN	Temporal Convolution Neural Network
T	Daily Maximum Temperature
Tm	Mean Daily Air Temperature
Ws	Wind Speed

References

1. Khan, A.A.; Nauman, M.A.; Bashir, R.N.; Jahangir, R.; Alroobaea, R.; Binmahfoudh, A.; Alsafyani, M.; Wechtaisong, C. Context Aware Evapotranspiration (*ETs*) for Saline Soils Reclamation. *IEEE Access* **2022**, *10*, 110050–110063. [CrossRef]
2. Martin, J.; Saez, J.A.; Corchado, E. On the suitability of stacking-based ensembles in smart agriculture for evapotranspiration prediction. *Appl. Soft Comput.* **2021**, *108*, 107509. [CrossRef]
3. Beguería, S.; Vicente-Serrano, S.M.; Reig, F.; Latorre, B. Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *Int. J. Climatol.* **2014**, *34*, 3001–3023. [CrossRef]
4. Hui-Mean, F.; Yusop, Z.; Yusof, F. Drought analysis and water resource availability using standardised precipitation evapotranspiration index. *Atmos. Res.* **2018**, *201*, 102–115. [CrossRef]
5. Wang, L.; Caylor, K.K.; Villegas, J.C.; Barron-Gafford, G.A.; Breshears, D.D.; Huxman, T.E. Partitioning evapotranspiration across gradients of woody plant cover: Assessment of a stable isotope technique. *Geophys. Res. Lett.* **2010**, *37*, L09401. [CrossRef]
6. Xu, C.Y.; Singh, V. Evaluation of three complementary relationship evapotranspiration models by water balance approach to estimate actual regional evapotranspiration in different climatic regions. *J. Hydrol.* **2005**, *308*, 105–121. [CrossRef]

7. Krishnaswamy, J.; Bonell, M.; Venkatesh, B.; Purandara, B.K.; Rakesh, K.; Lele, S.; Kiran, M.; Reddy, V.; Badiger, S. The groundwater recharge response and hydrologic services of tropical humid forest ecosystems to use and reforestation: Support for the "infiltration-evapotranspiration trade-off hypothesis". *J. Hydrol.* **2013**, *498*, 191–209. [CrossRef]

8. Feng, S.; Liu, J.; Zhang, Q.; Zhang, Y.; Singh, V.P.; Gu, X.; Sun, P. A global quantitation of factors affecting evapotranspiration variability. *J. Hydrol.* **2020**, *584*, 124688. [CrossRef]

9. Gates, D.M.; Hanks, R. Plant factors affecting evapotranspiration. *Irrig. Agric. Lands* **1967**, *11*, 506–521.

10. Ritchie, J.; Johson, B. Soil and plant factors affecting evaporation. *Agronomy* **1990**, *30*, 363–390.

11. Spontoni, T.A.; Ventura, T.M.; Palácios, R.S.; Curado, L.F.A.; Fernandes, W.A.; Capistrano, V.B.; Fritzen, C.L.; Pavão, H.G.; Rodrigues, T.R. Evaluation and Modelling of Reference Evapotranspiration Using Different Machine Learning Techniques for a Brazilian Tropical Savanna. *Agronomy* **2023**, *13*, 2056. [CrossRef]

12. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; FAO: Rome, Italy, 1998; Volume 300, p. D05109.

13. Maia, R.F.; Lurbe, C.B.; Hornbuckle, J. Machine learning approach to estimate soil matric potential in the plant root zone based on remote sensing data. *Front. Plant Sci.* **2022**, *13*, 2923. [CrossRef] [PubMed]

14. Bashir, R.N.; Khan, F.A.; Khan, A.A.; Tausif, M.; Abbas, M.Z.; Shahid, M.M.A.; Khan, N. Intelligent optimization of Reference Evapotranspiration (ETo) for precision irrigation. *J. Comput. Sci.* **2023**, *69*, 102025. [CrossRef]

15. Chong, J.L.; Chew, K.W.; Peter, A.P.; Ting, H.Y.; Show, P.L. Internet of Things (IoT)-Based Environmental Monitoring and Control System for Home-Based Mushroom Cultivation. *Biosensors* **2023**, *13*, 98. [CrossRef] [PubMed]

16. Saleh, S.; Cherradi, B.; El Gannour, O.; Gouiza, N.; Bouattane, O. Healthcare monitoring system for automatic database management using mobile application in IoT environment. *Bull. Electr. Eng. Inform.* **2023**, *12*, 1055–1068. [CrossRef]

17. Sitharthan, R.; Vimal, S.; Verma, A.; Karthikeyan, M.; Dhanabalan, S.S.; Prabaharan, N.; Rajesh, M.; Eswaran, T. Smart microgrid with the internet of things for adequate energy management and analysis. *Comput. Electr. Eng.* **2023**, *106*, 108556. [CrossRef]

18. Al Shahrani, A.M.; Alomar, M.A.; Alqahtani, K.N.; Basingab, M.S.; Sharma, B.; Rizwan, A. Machine Learning-Enabled Smart Industrial Automation Systems Using Internet of Things. *Sensors* **2023**, *23*, 324. [CrossRef]

19. Khan, A.A.; Faheem, M.; Bashir, R.N.; Wechtaisong, C.; Abbas, M.Z. Internet of Things (IoT) Assisted Context Aware Fertilizer Recommendation. *IEEE Access* **2022**, *10*, 129505–129519. [CrossRef]

20. Hu, Z.; Bashir, R.N.; Rehman, A.U.; Iqbal, S.I.; Shahid, M.M.A.; Xu, T. Machine Learning Based Prediction of Reference Evapotranspiration (ET 0) Using IoT. *IEEE Access* **2022**, *10*, 70526–70540. [CrossRef]

21. Bashir, R.N.; Bajwa, I.S.; Shahid, M.M.A. Internet of Things and Machine-Learning-Based Leaching Requirements Estimation for Saline Soils. *IEEE Internet Things J.* **2020**, *7*, 4464–4472. [CrossRef]

22. Bashir, R.N.; Bajwa, I.S.; Iqbal, M.W.; Ashraf, M.U.; Alghamdi, A.M.; Bahaddad, A.A.; Almarhabi, K.A. Leaching Fraction (LF) of Irrigation Water for Saline Soils Using Machine Learning. *Intell. Autom. Soft Comput.* **2023**, *36*, 1915–1930. [CrossRef]

23. Tang, Y.; Wang, Y.; Wu, D.; Liu, Z.; Zhang, H.; Zhu, M.; Chen, Z.; Sun, J.; Wang, X. An experimental investigation and machine learning-based prediction for seismic performance of steel tubular column filled with recycled aggregate concrete. *Rev. Adv. Mater. Sci.* **2022**, *61*, 849–872. [CrossRef]

24. Feng, W.; Wang, Y.; Sun, J.; Tang, Y.; Wu, D.; Jiang, Z.; Wang, J.; Wang, X. Prediction of thermo-mechanical properties of rubber-modified recycled aggregate concrete. *Constr. Build. Mater.* **2022**, *318*, 125970. [CrossRef]

25. Ben Abdallah, E.; Grati, R.; Boukadi, K. Towards an explainable irrigation scheduling approach by predicting soil moisture and evapotranspiration via multi-target regression. *J. Ambient. Intell. Smart Environ.* **2023**, *15*, 89–110. [CrossRef]

26. Hilmi, M.Z.B.; Anwar, T.; Rambli, D.R.B.A.; Salma, S.; Ashwitha, A.; Prayogo, P.H.; Rahyadi, I.; La Mani, Y.; Siagian, M.; Melenia, E.A.; et al. Long short-term memory with gated recurrent unit based on hyperparameter settings and hybridization for reference evapotranspiration rate prediction. *J. Theor. Appl. Inf. Technol.* **2022**, *100*, 6702–6714.

27. Nawandar, N.K.; Cheggoju, N.; Satpute, V. ANN-based model to predict reference evapotranspiration for irrigation estimation. In *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2020*; Springer: Singapore, 2021; pp. 671–679.

28. Bellido-Jiménez, J.A.; Estévez, J.; García-Marín, A.P. New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain. *Agric. Water Manag.* **2021**, *245*, 106558. [CrossRef]

29. Majumdar, P.; Bhattacharya, D.; Mitra, S. Prediction of evapotranspiration and soil moisture in different rice growth stages through improved salp swarm based feature optimization and ensembled machine learning algorithm. *Theor. Appl. Climatol.* **2023**, *153*, 649–673. [CrossRef]

30. Torres, A.B.; da Rocha, A.R.; da Silva, T.L.C.; de Souza, J.N.; Gondim, R.S. Multilevel data fusion for the internet of things in smart agriculture. *Comput. Electron. Agric.* **2020**, *171*, 105309. [CrossRef]

31. Chen, Z.; Zhu, Z.; Jiang, H.; Sun, S. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *J. Hydrol.* **2020**, *591*, 125286. [CrossRef]

32. Jia, W.; Zhang, Y.; Wei, Z.; Zheng, Z.; Xie, P. Daily reference evapotranspiration prediction for irrigation scheduling decisions based on the hybrid PSO-LSTM model. *PLoS ONE* **2023**, *18*, e0281478. [CrossRef]

33. Jung, D.H.; Lee, T.S.; Kim, K.; Park, S.H. A Deep Learning Model to Predict Evapotranspiration and Relative Humidity for Moisture Control in Tomato Greenhouses. *Agronomy* **2022**, *12*, 2169. [CrossRef]

34.   Roy, D.K.; Sarkar, T.K.; Kamar, S.S.A.; Goswami, T.; Muktadir, M.A.; Al-Ghobari, H.M.; Alataway, A.; Dewidar, A.Z.; El-Shafei, A.A.; Mattar, M.A. Daily Prediction and Multi-Step Forward Forecasting of Reference Evapotranspiration Using LSTM and Bi-LSTM Models. *Agronomy* **2022**, *12*, 594. [CrossRef]

35.   Yin, J.; Deng, Z.; Ines, A.V.M.; Wu, J.; Rasu, E. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). *Agric. Water Manag.* **2020**, *242*, 106386. [CrossRef]

36.   Wang, T.; Wang, X.; Jiang, Y.; Sun, Z.; Liang, Y.; Hu, X.; Li, H.; Shi, Y.; Xu, J.; Ruan, J. Hybrid machine learning approach for evapotranspiration estimation of fruit tree in agricultural cyber-physical systems. *IEEE Trans. Cybern.* **2022**, *53*, 5677–5691. [CrossRef] [PubMed]