

## Article

# Daily Prediction and Multi-Step Forward Forecasting of Reference Evapotranspiration Using LSTM and Bi-LSTM Models

Dilip Kumar Roy <sup>1,\*</sup> , Tapash Kumar Sarkar <sup>2</sup>, Sheikh Shamshul Alam Kamar <sup>1</sup>, Torsha Goswami <sup>3</sup>, Md Abdul Muktedir <sup>4</sup> , Hussein M. Al-Ghobari <sup>5</sup>, Abed Alataway <sup>6</sup>, Ahmed Z. Dewidar <sup>5,6</sup>, Ahmed A. El-Shafei <sup>5,6,7</sup>  and Mohamed A. Mattar <sup>5,6,8,\*</sup> 

- <sup>1</sup> Irrigation and Water Management Division, Bangladesh Agricultural Research Institute, Joydebpur, Gazipur 1701, Bangladesh; alamkamar91@gmail.com
  - <sup>2</sup> Grain Quality and Nutrition Division, Bangladesh Rice Research Institute, Joydebpur, Gazipur 1701, Bangladesh; tksarkar\_engr@yahoo.com
  - <sup>3</sup> Department of Veterinary Microbiology, Faculty of Veterinary and Animal Sciences, West Bengal University of Animal and Fishery Sciences, Kolkata 700056, West Bengal, India; toshami09@gmail.com
  - <sup>4</sup> Centre for Carbon, Water and Food, The University of Sydney, Camperdown, NSW 2570, Australia; md.muktadir@sydney.edu.au
  - <sup>5</sup> Department of Agricultural Engineering, College of Food and Agriculture Sciences, King Saud University, Riyadh 11451, Saudi Arabia; hghobari@ksu.edu.sa (H.M.A.-G.); adewidar@ksu.edu.sa (A.Z.D.); aelshafei1bn.c@ksu.edu.sa (A.A.E.-S.)
  - <sup>6</sup> Prince Sultan Bin Abdulaziz International Prize for Water Chair, Prince Sultan Institute for Environmental, Water and Desert Research, King Saud University, Riyadh 11451, Saudi Arabia; aalataway@ksu.edu.sa
  - <sup>7</sup> Department of Agricultural Engineering, Faculty of Agriculture (El-Shatby), Alexandria University, Alexandria 21545, Egypt
  - <sup>8</sup> Agricultural Engineering Research Institute (AEnRI), Agricultural Research Centre, Giza 12618, Egypt
- \* Correspondence: dilip.roy@my.jcu.edu.au (D.K.R.); mmattar@ksu.edu.sa (M.A.M.)



**Citation:** Roy, D.K.; Sarkar, T.K.; Kamar, S.S.A.; Goswami, T.; Muktedir, 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. <https://doi.org/10.3390/agronomy12030594>

Academic Editors: Pantazis Georgiou and Dimitris Karpouzou

Received: 29 December 2021

Accepted: 24 February 2022

Published: 27 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Precise forecasting of reference evapotranspiration ( $ET_0$ ) is one of the critical initial steps in determining crop water requirements, which contributes to the reliable management and long-term planning of the world's scarce water sources. This study provides daily prediction and multi-step forward forecasting of  $ET_0$  utilizing a long short-term memory network (LSTM) and a bi-directional LSTM (Bi-LSTM) model. For daily predictions, the LSTM model's accuracy was compared to that of other artificial intelligence-based models commonly used in  $ET_0$  forecasting, including support vector regression (SVR), M5 model tree (M5Tree), multivariate adaptive regression spline (MARS), probabilistic linear regression (PLR), adaptive neuro-fuzzy inference system (ANFIS), and Gaussian process regression (GPR). The LSTM model outperformed the other models in a comparison based on Shannon's entropy-based decision theory, while the M5 tree and PLR models proved to be the lowest performers. Prior to performing a multi-step-ahead forecasting, ANFIS, sequence-to-sequence regression LSTM network (SSR-LSTM), LSTM, and Bi-LSTM approaches were used for one-step-ahead forecasting utilizing the past values of the  $ET_0$  time series. The results showed that the Bi-LSTM model outperformed other models and that the sequence of models in ascending order in terms of accuracies was Bi-LSTM > SSR-LSTM > ANFIS > LSTM. The Bi-LSTM model provided multi-step (5 day)-ahead  $ET_0$  forecasting in the next step. According to the results, the Bi-LSTM provided reasonably accurate and acceptable forecasting of multi-step-forward  $ET_0$  with relatively lower levels of forecasting errors. In the final step, the generalization capability of the proposed best models (LSTM for daily predictions and Bi-LSTM for multi-step-ahead forecasting) was evaluated on new unseen data obtained from a test station, Ishurdi. The model's performance was assessed on three distinct datasets (the entire dataset and the first and the second halves of the entire dataset) derived from the test dataset between 1 January 2015 and 31 December 2020. The results indicated that the deep learning techniques (LSTM and Bi-LSTM) achieved equally good performances as the training station dataset, for which the models were developed. The research outcomes demonstrated the

ability of the developed deep learning models to generalize the prediction capabilities outside the training station.

**Keywords:** deep learning; recurrent neural networks; machine learning algorithms; reference evapotranspiration

## 1. Introduction

Water conservation in irrigated agriculture has been a significant concern, as agriculture consumes the majority of the world's freshwater reserves. A considerable amount of water can be saved through accurate quantification of crop water requirements, which depends on the precise estimation of evapotranspiration (ET), one of the vital elements in computational frameworks of water balance equations. Being an essential element of the surface energy balances and water budgets, ET plays a central role in controlling interactions among soil, vegetation, and the atmosphere [1]. As such, proper design and efficient management of irrigation techniques and reliable planning for the allocation of scarce water resources largely depend on the accurate estimation of the ET [2]. The values of ET can be obtained through direct measurement techniques, including lysimeter methods, eddy covariance techniques, and the Bowen ratio–energy balance approach [3–5], which are expensive and deemed unavailable in many countries [6,7]. Alternatively, ET can be estimated indirectly utilizing a set of accessible climatological variables to determine reference evapotranspiration ( $ET_0$ ). This indirect approach has been extensively used in many parts around the globe in which either unavailability or budgetary constraints prohibit direct estimation of ET. One of the most stable and well-established techniques of  $ET_0$  computation is the FAO-56 Penman–Monteith (FAO-56 PM) equation [6]. It is also utilized to validate alternative  $ET_0$  computation methods, as the equation was validated using lysimeter methods in different climates [8].  $ET_0$  computation using the FAO-56 PM equation requires a few climatological variables, including maximum and minimum air temperatures, wind speed, relative humidity, and solar radiation. Upon estimation of  $ET_0$ , crop evapotranspiration can be obtained by utilizing estimated  $ET_0$  values and crop coefficient values for a particular crop.

Machine learning algorithms have recently been recognized as reliable tools in the prediction and future forecasting of  $ET_0$ . They have been used extensively in providing a reasonably accurate forecast of  $ET_0$  in various hydrologic and climatic settings. The first implementation of  $ET_0$  prediction modeling was based on the usage of artificial neural networks (ANN) [9–13]. Later, different variants of ANN and other machine learning algorithms have attained the researchers' interests. These include the usage of generalized regression neural networks [14,15], neural network with optimum time lags [16], adaptive neuro-fuzzy inference system (ANFIS) [17–23], random forests (RF) [14,24,25], CatBoost [26], hybrid extreme gradient boosting grey wolf optimizer (GWO) [27], extreme learning machine (ELM) [15,17,28–31], support vector regression (SVR) [23–25,31–33], multivariate relevance vector regression [34], genetic programming (GP) [35], Gaussian process regression (GPR) [36], multivariate adaptive regression splines (MARS) [2,9], M5 model tree (M5Tree) [2], radial basis M5Tree [37], gene-expression programming (GEP) [12,18,38–45], hierarchical fuzzy systems (HFS) [46], coupled extreme gradient boosting-whale optimization algorithm [47], coupled natural-extreme gradient boosting [48], hybrid model based on variational mode decomposition-GWO-SVM [49], and inter-model ensemble approaches [50]. Apart from machine learning approaches, there are other approaches of ET estimation, including the application of Sentinel-2 spectral information [51], comparison of different empirical methods [52], utilizing NASA POWER Reanalysis Products [53], and using lysimeter data [54]. Recently, Bellido-Jiménez et al. [55] examined several machine learning approaches to improve  $ET_0$  estimations, considering only the temperature-based data (EnergyT and Hourmin) as inputs, and they determined that ELM outperformed the

others. In another study, Vásquez et al. [56] proposed several methods based on maximum and minimum temperatures to enhance  $ET_0$  computation under scarce data situations in the high tropical Andes. Nourani et al. [57] proposed one-, two-, and three-step-ahead predictions of  $ET_0$  using ensembles of ANFIS, ANN, and MLR models in various climatic stations. This study evaluates deep learning algorithms' daily prediction and multi-step (5 steps)-ahead forecasting abilities.

The deep machine learning (DL) technique has attained substantial attention in recent years, being considered an advanced version of machine learning techniques. The DL technique has been successfully utilized in various research domains, including time series prediction [58–60], computer vision [61], classification of images [62], recognition of speech [63], language processing [64], forecasting of groundwater levels [65,66], and prediction of water quality parameters [67]. The DL techniques are primarily based on the recurrent neural networks (RNN), which, for their ability to preserve and utilize memory from the previous network states, are superior candidates for predicting and forecasting time series data [68–70]. Nevertheless, despite the ability to capture the trends of the time series data, the standard RNN model structures face difficulties in retaining the longer-term dependence among the variables and suffer from vanishing and exploding gradients-related issues [71]. Due to these two inherent problems of the standard RNN, network training becomes unrealistic as the network weights may either become zero or unnecessarily large during network training. The two most important criteria that ensure better network training are retaining necessary information and eluding redundant or unnecessary information among various network states. A long short-term memory (LSTM) network possesses these characteristics to overcome the training shortfall of RNNs. The LSTMs are the variants of standard RNNs and have widely been used in various research domains such as financial time series and language processing [72], traffic congestion, and traveling [73], including the application in the hydrologic time series prediction [74–77].

The application of DL-based models in predicting pan evaporation, reference evapotranspiration, and crop evapotranspiration in different climatic conditions have been found in recent literature. These include daily pan evaporation prediction using deep LSTM model [78], evapotranspiration computation estimation using deep neural network [79], daily reference evapotranspiration prediction using convolutional neural network (CNN) [80], one-step-ahead forecasting of reference evapotranspiration using LSTM [81], multi-step-ahead forecasting of daily reference evapotranspiration using LSTM and CNN-LSTM [82], multi-week-ahead forecasting of  $ET_0$  using CNN-gated recurrent unit optimized with ant colony optimization [83],  $ET_0$  estimation using deep learning-multilayer perceptrons [84], and short-term actual ET prediction using LSTM and NARX [85]. Despite the  $ET_0$  prediction and forecasting application, the DL-based models, especially LSTM models, need to be evaluated for different combinations of input variables that provide better prediction accuracy. Recently, Zhang et al. [26] used only eight input combinations of different meteorological variables to estimate reference crop evapotranspiration using the CatBoost model. Another study by Maroufpoor et al. [86] used optimal input combinations to estimate reference evapotranspiration using a hybridized ANN model. Another study [87] used 29 different combinations of input variables from various meteorological variables to forecast daily reference evapotranspiration using ANN, SVR, and ELM. To the best of our knowledge, none of the previous studies evaluated all possible combinations of available input climatological variables to provide daily and multi-step forward  $ET_0$  estimation using DL-based LSTM models. This is the first effort that has used various possible combinations of input variables using a deep learning model to predict daily and forecast multi-step-ahead reference evapotranspiration.

Another critical aspect of predictive modeling with the machine or deep learning approaches is evaluating the established models' ability to anticipate and forecast data from other meteorological stations. However, the generalization capabilities of the developed models for predicting and forecasting  $ET_0$  in other meteorological stations have been given relatively little attention. For daily prediction of  $ET_0$ , Wang et al. [44] investigated the

generalization capability of RF- and GEP-based machine learning tools, while Roy et al. [46] evaluated the potential of HFS models in generalizing the outputs using data from another meteorological station. For one-step-ahead forecasting of  $ET_0$ , Roy [81] utilized LSTM models; however, the study did not evaluate the generalization capability of the developed LSTM models for a new unseen test dataset. Nevertheless, model generalization has not been used for multi-step-ahead  $ET_0$  forecasting using different combinations of input variables as well as using various machine and deep learning algorithms. To the best of our understanding, this study was the first attempt at providing daily prediction and multi-step-forward forecasting of  $ET_0$  using LSTM and Bi-LSTM models.

Therefore, the prime objective and focus of this research were to (1) explore the capability of DL-based techniques, LSTM, and Bi-LSTM in predicting daily and forecasting multi-step (5 day)-ahead  $ET_0$  estimates in the selected study areas in Bangladesh; (2) compare the prediction and forecasting skill of the proposed LSTM and Bi-LSTM models with that of the commonly used machine learning algorithms; and (3) assess the generalization capability of the proposed LSTM and Bi-LSTM models to predict and forecast  $ET_0$  at a nearby station, at which the models were neither trained nor validated.

## 2. Material and Methods

### 2.1. Study Area and the Data

The study area consists of two upazillas (administrative units) in Gazipur and Pabna districts: Gazipur Sadar Upazilla and Ishurdi Upazilla (Figure 1). Meteorological data, including minimum and maximum daily temperatures, relative humidity, wind speed, and duration of sunshine, were acquired from two weather stations (Gazipur Sadar and Ishurdi). The climatic variables were gathered from different weather stations, as illustrated in Figure 1. A silicon photodiode type global solar radiation recorder (Licor-200SZ, LI-COR Biosciences, USA; accuracy =  $\pm 5\%$ ; range = 0.3–4  $\mu\text{m}$ ; measurement height = 2 m) was used to measure the amount of sunshine along with length of the day. The maximum and minimum temperatures were measured employing the maximum and minimum thermometers (Zeal P1000, G. H. Zeal Ltd., London SW19 3UU, UK; accuracy =  $\pm 0.2$   $^{\circ}\text{C}$ ; range and resolution =  $-50$  to  $+70$   $^{\circ}\text{C}$ , 0.1  $^{\circ}\text{C}$ ; measurement height = 2 m). Relative humidity was measured using a capacitive-type hygrometer (R. M. Young Company, Traverse City, MI 49686, USA; accuracy =  $\pm 3\%$ ; range and resolution = 0–100%, 1%; measurement height = 2 m). The measurement of wind speed was performed using a rotating cup anemometer (Cup Anemometer 4.3018.10.000, Adolf Thies GmbH and Co. KG, Hauptstraße 76, 37083 Göttingen, Germany; accuracy = 1.2 m/s; range and resolution = 0.5–60 m/s, 0.1 m/s; measurement height = 10 m). It is noted that performing a thorough quality assurance procedure is often desirable to ensure the quality of climatic datasets, which enhances the reliability of  $ET_0$  estimations using machine learning tools [88]. Although a detailed quality assurance procedure was not performed, the quality of the obtained climatic data was checked thoroughly for its correctness and completeness. The missing entries (less than 1%) were imputed using the ‘movmedian’ (Matlab MATLAB 2021a) approach of data imputation. Nevertheless, a few adjustments were performed to obtain the FAO-56 PM equation appropriate for local conditions following the recommendations provided in [89]. For instance, the wind speeds obtained at 10 m height (from the weather stations) were converted to wind speeds at the height of 2 m (keeping a lower limit of 0.5 m/s).

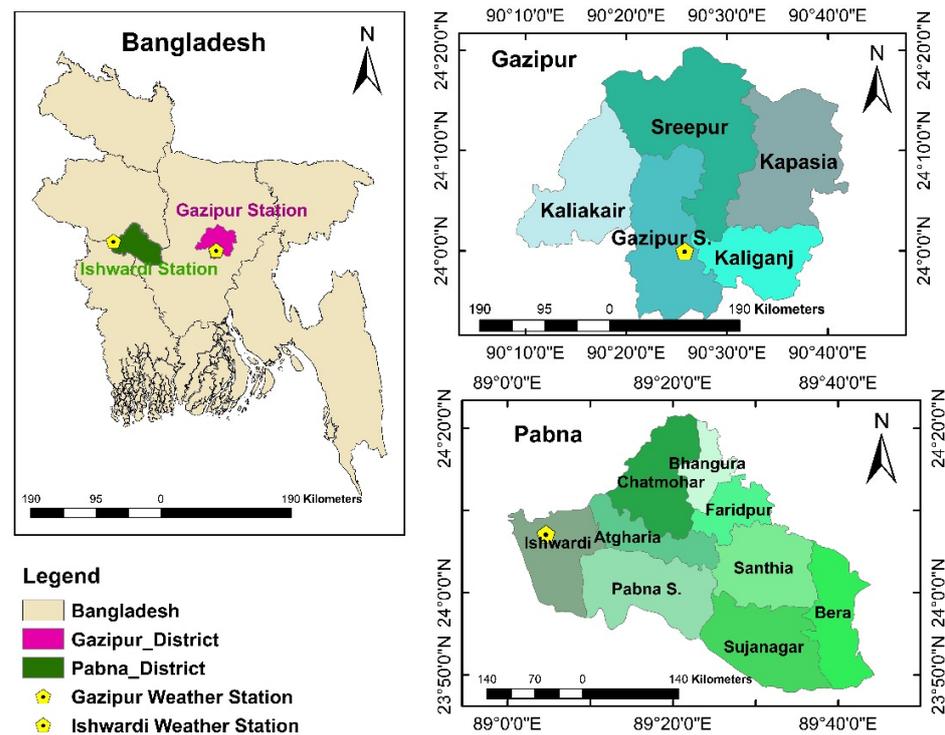


Figure 1. Weather station locations.

The weather station in Gazipur Sadar Upazilla was utilized as the training station for developing the proposed models, whereas data from the weather station in Ishurdi were used to evaluate the produced models’ performance (testing station). The position of the weather station at Gazipur Sadar Upazilla is at 24.00° N latitude and 90.43° E longitude, being located 8.4 m above mean sea level (MSL). On the other hand, the test station is placed between 24.12° N latitude and 89.08° E longitude with an altitude of 18 m from the MSL. The weather data for the training station were obtained for a duration of 15.5 years (from 1 January 2004 to 30 June 2019). Descriptive statistics of the acquired weather data for the training station are presented in Table 1.

Table 1. Descriptive statistics of the weather data for the training station (Gazipur Sadar Upazilla).

Climatic Variables	Min	Max	Mean	Standard Deviation	Skewness	Kurtosis
<i>Data Range: 1 January 2004 to 30 June 2019 (5660 Daily Entries)</i>						
Minimum temperature, °C	4.40	34.50	21.17	5.64	−0.63	−0.88
Maximum temperature, °C	12.00	53.00	30.93	3.92	−1.10	2.11
Relative humidity, %	38.00	89.00	80.22	8.20	−0.63	0.75
Wind speed, m/s	0.68	5.06	2.79	1.05	−0.06	−1.32
Sunshine duration, h	0.00	11.40	5.54	3.09	−0.40	−1.04

The weather data for the test station were obtained for a duration of around 5.5 years (from 1 June 2015 to 31 December 2020). Descriptive statistics of the acquired weather data for the test station are presented in Table 2.

**Table 2.** Descriptive statistics of the entire, first half, and the second half of the weather data for the test station (Ishurdi Upazilla).

Climatic Variables	Mean	Standard Deviation	Skewness	Kurtosis
<i>Entire dataset (1 June 2015 to 31 December 2020: 2041 daily entries)</i>				
Minimum temperature, °C	21.37	5.98	−0.73	−0.76
Maximum temperature, °C	31.46	4.16	−0.83	0.28
Relative humidity, %	78.89	12.18	−1.23	1.93
Wind speed, m/s	1.43	0.23	0.07	0.22
Sunshine duration, h	5.90	3.19	−0.41	−0.71
<i>First half data (1 June 2015 to 3 October 2018: 1221 daily entries)</i>				
Minimum temperature, °C	21.06	6.08	−0.65	−0.92
Maximum temperature, °C	31.27	4.21	−0.71	0.26
Relative humidity, %	80.06	11.30	−1.24	2.25
Wind speed, m/s	1.43	0.23	0.06	0.35
Sunshine duration, h	5.75	3.18	−0.42	−0.98
<i>Second half data (4 October 2018 to 31 December 2020: 820 daily entries)</i>				
Minimum temperature, °C	21.69	5.87	−0.83	−0.56
Maximum temperature, °C	31.66	4.11	−0.95	0.35
Relative humidity, %	77.71	12.89	−1.18	1.54
Wind speed, m/s	1.44	0.23	0.09	0.08
Sunshine duration, h	6.05	3.19	−0.39	−0.44

Weather data acquired from the two weather stations for the specified duration were used to calculate the daily  $ET_0$  values employing the FAO-56 PM equation (Equation (1)). The climatological variables (acquired weather data) and corresponding  $ET_0$  values (computed using FAO-56 PM equation) were used as inputs and outputs from the proposed LSTM, Bi-LSTM, and other machine learning-based models. This approach of estimating  $ET_0$  indirectly using the climatological variables has been a widely accepted method in situations where obtaining  $ET_0$  directly becomes infeasible due to technical and budgetary constraints [6,15,90]. The FAO-56 PM equation is represented by

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T_{\text{mean}} + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

where  $ET_0$  denotes reference evapotranspiration, mm/d;  $R_n$  represents the net radiation at the crop surface, MJ/m<sup>2</sup>/d;  $G$  indicates heat flux density of soil, MJ/m<sup>2</sup>/d;  $\Delta$  represents the slope of the saturation vapor pressure curve, kPa/°C;  $\gamma$  denotes psychrometric constant, kPa/°C;  $e_s$  represents the saturation vapor pressure, kPa;  $e_a$  indicates the actual vapor pressure, kPa;  $u_2$  is the wind speed at the height of 2 m, m/s; and  $T_{\text{mean}}$  is the mean air temperature at 2.0 m height, °C.

The computed  $ET_0$  values at the training station (Gazipur Sadar Upazilla) ranged between 0.92 and 8.02 mm/d, with the mean, standard deviation, skewness, and kurtosis values of 3.80 mm/d, 1.32 mm/d, 0.30, and −0.67, respectively. For the test station (Ishurdi), the computed  $ET_0$  time series was divided into three sub-time series to test the generalization capability of the proposed modeling approach at different regions of the time series. The first time series considered was the entire dataset for which the  $ET_0$  values had the mean, standard deviation, skewness, and kurtosis values of 3.67 mm/d, 1.24 mm/d, 0.28, and −0.62, respectively. The values of the mean, standard deviation, skewness, and kurtosis of the calculated  $ET_0$  for the first half of the dataset were 3.57 mm/d, 1.25 mm/d, 0.35, and −0.62, respectively. The second half of the  $ET_0$  time series contained the mean, standard deviation, skewness, and kurtosis values of 3.76 mm/d, 1.23 mm/d, 0.22, and −0.59, respectively.

For daily  $ET_0$  prediction, meteorological variables and calculated  $ET_0$  values using the FAO-56 PM equation were used as inputs and outputs. On the other hand, calculated  $ET_0$  time series were used to develop the proposed models for one- and multi-step-ahead predictions by obtaining time-lagged characteristics from the time series data. For training the models, the entire dataset was divided into three parts: training data (40% of the entire dataset: 2264 daily entries—from 1 January 2004 to 13 March 2010), validation data (40% of the entire dataset: 2264 daily entries—from 14 March 2010 to 24 May 2016), and test data (remaining 20% of the total dataset: 1132 daily entries—from 25 May 2016 to 30 June 2019). To test the generalization capability of the proposed models, we partitioned the data from the test station as follows: entire dataset (2021  $ET_0$  values and associated meteorological variables ranging from 1 June 2015 to 31 December 2020), the first half of the entire dataset (1221  $ET_0$  values and associated meteorological variables ranging from 1 June 2015 to 3 October 2018), and the first half of the entire dataset (820  $ET_0$  values and associated meteorological variables ranging from 4 October 2018 to 31 December 2020).

## 2.2. Prediction Models

### 2.2.1. Long Short-Term Memory (LSTM) Networks

An LSTM is a variant of the neural network-based modeling approach, an upgraded version of RNNs capable of learning long-term dependence that exists at various steps in the sequential time series data. LSTMs safeguard against the vanishing and exploding gradient issues commonly observed in a standard RNN architecture, making an LSTM an ideal modeling tool to predict and forecast sequential time series data. To eliminate vanishing and exploding gradient problems, an LSTM integrates two important parameters called ‘state dynamics’ and ‘gating functions’ [91]. An LSTM network architecture is made up of several interconnected memory blocks that are connected to each other in a number of layers, each of which consists of many recurrently connected memory cells. The memory cells of LSTM architectures are comprised of three gates [92]: (a) input, (b) forget, and (c) output. For performing a regression task, an LSTM model employs four layers: a sequence input layer, an LSTM layer, a fully connected layer, and a regression layer. The input and fully connected layers correspond to the number of input and output variables, respectively. The LSTM layer accommodates the number of hidden units, whereas the regression layer performs the regression task. The sequence input and LSTM layers are the most important components of a fundamental LSTM network. The input layer is responsible for inputting the sequence data, e.g., time-series data to the network, whereas the LSTM layer facilitates learning long-term dependence among various time-steps of a sequential time series data. A comprehensive explanation of the LSTM model architecture is presented by Roy [81] and is not repeated in this effort. A bidirectional LSTM network (Bi-LSTM) architecture is similar to an LSTM network except that a Bi-LSTM network is associated with bidirectional long-term dependence among various time-steps of a sequential time series data.

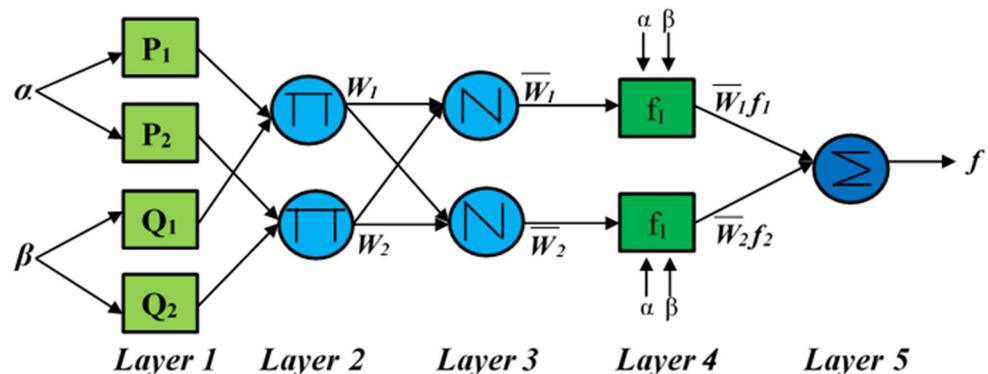
In this study, both networks (LSTM and Bi-LSTM) have three hidden layers, each of which is followed by a dropout layer that is employed to prevent model overfitting. Each of the three hidden layers has a large number of hidden neurons. The first, second, and third hidden layers each had 100, 50, and 20 hidden neurons, respectively. In contrast, the dropout rates assigned for the associated dropout layers were chosen as 0.4, 0.3, and 0.2, respectively. The optimum numbers of hidden layers, hidden neurons, and dropout rates are determined by conducting a series of trials. Numerous combinations of varying numbers of these parameters are tested until a stable network is obtained. In addition, the best training options are selected upon conducting several trials, and similar training options are used for training both the LSTM and Bi-LSTM models for consistency. The training options used for training the LSTM and Bi-LSTM networks are provided in Table 3.

**Table 3.** Training options and the associated parameter values.

Training Options	Corresponding Parameter Values
Solver for optimization	'adam'
Maximum number of epochs	1000
Gradient threshold value	1
Preliminary learning rate	0.01
Minimum size of the batch	150
Length of sequence	1000

2.2.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS, a variant of fuzzy inference systems (FIS), is adaptive in nature, incorporating fuzziness and ambiguity of input variables in developing input–output relationships of nonlinear systems [93]. An ANFIS grab holds the advantageous features of both the artificial neural networks and fuzzy set theory into an adaptive framework to model nonlinear and complex systems quite efficiently and effectively [94,95]. Due to less complexity and better learning ability [93], a Sugeno-type FIS is used to develop the ANFIS model utilizing a fuzzy c-means clustering (FCM) [96] algorithm to reduce the dimensionality of input variables. Detailed descriptions of ANFIS model structures are provided in Jang et al. [93] and are not repeated in this effort. Figure 2 presents an ANFIS model structure derived from a Sugeno-type FIS. The ANFIS models were developed in a MATLAB [97] environment.



**Figure 2.** A schematic diagram of an ANFIS structure derived from a first-order Sugeno-type FIS. Reprinted with permission from Jang [98].

2.2.3. Gaussian Process Regression (GPR)

A GPR is a nonparametric modeling algorithm that is derived from the theories of probability and Gaussian process [99]. Following a Gaussian distribution, a GPR model provides the output,  $Y$  from the input variables, and  $X(i)$  through developing a functional relationship, which can be mathematically represented as [100]

$$Y = f(X(i)) + \epsilon \tag{2}$$

where  $\epsilon$  is a Gaussian noise, the variance of which is denoted by  $\sigma_n^2$ .

The mean,  $m(x_i)$ , and covariance,  $k(x_i, x_j)$ , functions are the two important functional components of a typical GPR model. They can be mathematically expressed as [99]

$$m(x_i) = E[f(x_i)] \tag{3}$$

$$k(x_i, x_j) = E[(f(x_i) - m(x_i))(f(x_j) - m(x_j))] \tag{4}$$

On the basis of these two key functions, the functional relationship using Gaussian process theory is established by the following equation:

$$f(x) \sim gp(m(x_i), k(x_i, x_j)) \tag{5}$$

The prediction probability distribution of a GPR model is governed by the free parameters or hyperparameters, which are in essence the parameters of the mean and covariance functions. The values of free parameters or hyperparameters depend on the training dataset's log-likelihood function values. The GPR models were developed by utilizing the commands and functions of MATLAB [97].

#### 2.2.4. M5 Model Trees (M5 Tree)

The development of the M5 tree is derived from the philosophies associated with the M5 technique [101,102] in building standalone trees. The prediction capabilities of M5 trees were demonstrated and well documented in various research domains [103,104]. In the M5 tree modeling approach, a complex modeling task is sub-divided into numerous sub-tasks via the divide-and-conquer technique, and the final result is the integration of solutions from all the sub-tasks [103]. This splitting technique results in a hierarchy of model trees in which non-terminal nodes are associated with splitting rules, whereas expert models are represented by the tree leaves [104]. Model development using the M5 tree technique is performed using three stepwise procedures: (1) development of an initial tree, (2) pruning of the tree, and (3) smoothing of the tree [105]. In the MATLAB environment, a toolbox "M5PrimeLab" [106] was used to develop M5 trees for predicting daily reference ET<sub>0</sub> values.

#### 2.2.5. Multivariate Adaptive Regression Spline (MARS)

MARS [107] is a nonparametric modeling technique that is adaptive in nature and is believed to be a flexible and rapid approach to developing regression models. The MARS approach partitions the entire decision space into several input parameters on which standalone basis functions or splines are fitted to obtain the final MARS model [108]. Both a forward procedure and a backward procedure are utilized, i.e., MARS initially builds a comparatively complex model using the user-specified maximum number of basis functions in the forward step. In contrast, in the backward step, MARS parsimoniously selects the most significant input variables in predicting the output variable [109]. The backward step eliminates redundant input variables and assists in simplifying the final model while avoiding over-or under-fitting. The relationship between the input and output variables can be represented by the following equation [110]:

$$BF_i(x) = \max(0, x_j - \alpha) \text{ OR } BF_i(x) = \max(0, \alpha - x_j) \quad (6)$$

$$y = f(x) = \beta \pm \gamma_k \times BF_i(x) \quad (7)$$

where  $i$  represents the index of Basis functions,  $j$  denotes the index of input variables,  $BF_i$  symbolizes the  $i^{\text{th}}$  Basis function,  $x_j$  is the  $j^{\text{th}}$  input variable,  $\alpha$  is a threshold value used by the MARS model during model building,  $\beta$  is a constant,  $\gamma_k$  indicates the respective coefficient of  $BF_i(x)$ , and  $y$  denotes the model prediction (output variable).

A MATLAB toolbox 'ARESLab' [106] was employed to build MARS-based ET<sub>0</sub> prediction models. This study used both piecewise-linear and piecewise-cubic modeling approaches to predict daily ET<sub>0</sub> values.

#### 2.2.6. Probabilistic Linear Regression (PLR)

PLR utilizes Bayesian inference techniques to develop prediction models through probabilistically performing linear regression. The PLR approach is often referred to as empirical Bayesian linear regression, using either an expectation-maximization (EM) algorithm [111] or a Mackay fixpoint iteration method [112]. The EM algorithm is generally utilized to formulate the PLR models. As such, the present study used the EM algorithm in developing PLR-based ET<sub>0</sub> prediction models. Mo Chen [113] developed a MATLAB toolbox in this research to develop PLR models.

### 2.2.7. Support Vector Regression (SVR)

SVRs are derived from the principles of the support vector machine (SVM) algorithm [114], which has been attracting significant attention in recent years for its capability to solve a diversified range of regression and classification problems [115]. SVRs are developed via a nonlinear mapping technique that utilizes required data from the input space to a high-dimensional feature space on which linear regressions are executed [116]. An elaborated explanation of the theory of the SVR approach has been provided in Chevalier et al. [117], and only a brief account of the SVR theorem is presented in this effort. The following equation symbolizes the training dataset in developing a linear SVR model:

$$\{(\bar{x}_1, y_1), (\bar{x}_2, y_2), \dots, (\bar{x}_l, y_l)\} \tag{8}$$

$$\bar{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R}, \text{ and } l = \text{number of data entries}$$

In this case, the solution function can be expressed as

$$f(\bar{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle \bar{x}_i, \bar{x} \rangle + b \tag{9}$$

where  $\langle \cdot, \cdot \rangle$  denotes dot product, and  $\alpha_i, \alpha_i^*$ , and  $b$  represent coefficients computed by the SVR model.

A data transformation step is performed to build nonlinear SVR models, including a nonlinear mapping function  $\varnothing$  [118] that transforms low-dimensional input space into a high-dimensional feature space. The computation  $\varnothing$  becomes challenging during progressive mapping of the input–output data into higher dimensions. This limitation is handled using the Mercers theorem, which can be represented by the following equation:

$$\langle \varnothing(\bar{u}), \varnothing(\bar{v}) \rangle = k(\bar{u}, \bar{v}) \tag{10}$$

For a particular mapping  $\varnothing$ , the Mercers theorem introduces the concept of using a kernel function  $k$ , which is used to calculate the dot product of any two points  $(\bar{u}, \bar{v})$ , and the computation of dot products in this approach bypasses the explicit calculation of high-dimensional and nonlinear mapping. The prediction performance of nonlinear SVR models depends on the kernel function, which is regarded as one of the most important parameters in SVR modeling.

### 2.3. Ranking of the ET<sub>0</sub> Prediction Models: Shannon’s Entropy

ET<sub>0</sub> prediction models were ranked using performance-based weights assigned to standalone models using Shannon’s entropy principle. For this, a decision matrix of prediction models ( $m$ ) and performance indices (PI) is formulated, which can be represented in the form of the following equation [119]:

$$ET_{ij} = \begin{bmatrix} ET_{11} & ET_{21} & \dots & ET_{m1} \\ ET_{12} & ET_{22} & \dots & ET_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ ET_{1PI} & ET_{2PI} & \dots & ET_{mPI} \end{bmatrix} \tag{11}$$

To reduce the adverse impacts of index dimensionality, we standardized the performance index values between 0 and 1  $\{S_{ij} \in [0, 1], i = 1, 2, \dots, m; j = 1, 2, \dots, PI\}$ . The standardization component  $S_{ij}$  was performed using the following equation [119]:

$$S_{ij} = \begin{cases} \frac{ET_{ij}}{\max(ET_{i1}, ET_{i2}, \dots, ET_{iPI})}, & \text{for benefit indexes} \\ \frac{X_{ij}}{\min(ET_{i1}, ET_{i2}, \dots, ET_{iPI})}, & \text{for cost indexes} \end{cases} \tag{12}$$

Shannon’s entropy-based ranking was performed using a five-step stepwise procedure described in Roy et al. [21], which was not repeated here.

2.4. Selection of Input Variables for Daily Predictions

All possible combinations of the five input variables (minimum temperatures, maximum temperatures, relative humidity, wind speed, and sunshine hours) were used. A total of 31 models were developed on the basis of the 31 combinations (single, two-input combinations, three-input combinations, four-input combinations, and all five inputs) of input variables. Two-, three-, and four-input combinations are presented in Table 4.

Table 4. Different combinations of two-, three-, and four-input combinations.

Two-Input Combinations	Three-Input Combinations	Four-Input Combinations
Min temp, max temp	Min temp, max temp, humidity	Min temp, max temp, humidity, wind speed
Min temp, humidity	Min temp, max temp, wind speed	Min temp, max temp, humidity, sunshine hours
Min temp, wind speed	Min temp, max temp, sunshine hours	Min temp, max temp, wind speed, sunshine hours
Min temp, sunshine hours	Min temp, humidity, wind speed	Min temp, humidity, wind speed, sunshine hours
Max temp, humidity	Min temp, humidity, sunshine hours	Max temp, humidity, wind speed, sunshine hours
Max temp, wind speed	Min temp, wind speed, sunshine hours	
Max temp, sunshine hours	Max temp, humidity, wind speed	
Humidity, wind speed	Max temp, humidity, sunshine hours	
Humidity, sunshine hours	Max temp, wind speed, sunshine hours	
Wind speed, sunshine hours	Humidity, wind speed, sunshine hours	

These combinations of input variables were evaluated for two deep learning algorithms (LSTM and Bi-LSTM). The 62 models (31 LSTM + 31 Bi-LSTM) developed were ranked on the basis of their prediction accuracies using Shannon’s entropy by incorporating a number of benefit (correlation coefficient, Nash–Sutcliffe efficiency coefficient, Willmott’s index of agreement) and cost (normalized or relative root mean squared error, maximum absolute error, median absolute deviation) performance evaluation indices. The best-input combinations thus obtained were used to develop the other shallow machine learning algorithms.

2.5. Model Performance Evaluation

The performances of the proposed models were evaluated using various statistical evaluation indices as follows:

- Correlation coefficient, R

$$R = \frac{\sum_{i=1}^n (ET_{i,a} - \overline{ET}_a) (ET_{i,p} - \overline{ET}_p)}{\sqrt{\sum_{i=1}^n (ET_{i,a} - \overline{ET}_a)^2} \sqrt{\sum_{i=1}^n (ET_{i,p} - \overline{ET}_p)^2}} \tag{13}$$

- Nash–Sutcliffe efficiency coefficient, NS [120]

$$NS = 1 - \frac{\sum_{i=1}^n (ET_{i,a} - ET_{i,p})^2}{\sum_{i=1}^n (ET_{i,a} - \overline{ET}_a)^2} \tag{14}$$

- Index of agreement, IOA [121]

$$IOA = 1 - \frac{\sum_{i=1}^n (ET_{i,a} - ET_{i,p})^2}{\sum_{i=1}^n (|ET_{i,p} - \overline{ET}_a| + |ET_{i,a} - \overline{ET}_a|)^2} \tag{15}$$

- Root mean square error, RMSE [122]

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{ET}_{i,a} - \text{ET}_{i,p})^2} \quad (16)$$

- Normalized RMSE, NRMSE

$$\text{NRMSE} = \frac{\text{RMSE}}{\overline{\text{ET}_a}} \quad (17)$$

- Maximum absolute error, MAE

$$\text{MAE} = \max [|\text{ET}_{i,a} - \text{ET}_{i,p}|] \quad (18)$$

- Median absolute deviation, MAD

$$\text{MAD}(\text{ET}_a, \text{ET}_p) = \text{median}(|\text{ET}_{1,a} - \text{ET}_{1,p}|, |\text{ET}_{2,a} - \text{ET}_{2,p}|, \dots, |\text{ET}_{n,a} - \text{ET}_{n,p}|) \quad (19)$$

for  $i = 1, 2, \dots, n$

where  $\text{ET}_{i,a}$  and  $\text{ET}_{i,p}$  are  $\text{ET}_0$  quantities at the  $i^{\text{th}}$  data points acquired from the FAO-56 PM computed and model predicted values, respectively;  $\overline{\text{ET}_a}$  represents the arithmetic mean of the FAO-56 PM computed  $\text{ET}_0$  values; and  $n$  is the amount of input-output data.

### 3. Results and Discussion

#### 3.1. Daily Prediction of $\text{ET}_0$ Using Various Machine Learning Algorithms at the Training Station (Gazipur Sadar)

To determine the optimum numbers of input variables combinations, we used 31 possible combinations of five input variables to develop 31 LSTM and 31 Bi-LSTM models. Learning (training) and testing of the  $\text{ET}_0$  models were performed simultaneously. Prediction errors on the test dataset in terms of RMSE criterion for the 31 developed models are presented in Table 5. As evidenced by the numerical values presented in Table 5, the LSTM model predictions were slightly better than those of the Bi-LSTM models when the RMSE criterion was used as a deciding factor. It was also observed that both the LSTM- and Bi-LSTM-based  $\text{ET}_0$  prediction models produced the lowest RMSE values (best daily  $\text{ET}_0$  predictions) when all five variables were used. The performance of LSTM (RMSE = 0.081 mm/d) was slightly better than that of the Bi-LSTM (RMSE = 0.087 mm/d) model. However, in situations where adequate data are not available, the use of fewer input variables may be employed to achieve a realistically precise prediction of  $\text{ET}_0$  values. For instance, four climatological variables (a combination of maximum temperature, relative humidity, wind speed, and sunshine hours) could be used to obtain sufficiently accurate daily  $\text{ET}_0$  predictions using LSTM (test error in terms of RMSE value equals 0.107 mm/d) and Bi-LSTM (test error in terms of RMSE value equals 0.116 mm/d) models. Other combinations of four meteorological variables, e.g., (minimum temperature, maximum temperature, relative humidity, sunshine hours) and (minimum temperature, relative humidity, wind speed, sunshine hours) provided reasonably accurate daily  $\text{ET}_0$  predictions (Table 5). In addition, combinations of three meteorological variables (relative humidity, wind speed, sunshine hours) and (minimum temperature, relative humidity, sunshine hours) produced reasonable accurate predictions, with test RMSE values ranging between 0.333 and 0.377 mm/d.

**Table 5.** Prediction errors of deep learning-based ET<sub>0</sub> models (LSTM and Bi-LSTM) with different input combinations on the test dataset.

Model No.	Different Input Combinations	Test RMSE, mm/d	
		LSTM	Bi-LSTM
<i>Single Input Combinations</i>			
M1	Min temp	0.880	0.964
M2	Max temp	0.775	0.781
M3	Humidity	1.124	<b>1.211</b>
M4	Wind speed	<b>1.177</b>	1.105
M5	Sunshine hours	0.732	0.807
<i>Two Inputs combinations</i>			
M6	Min temp, max temp	0.765	0.779
M7	Min temp, humidity	0.729	0.751
M8	Min temp, wind speed	1.004	1.049
M9	Min temp, sunshine hours	0.527	0.514
M10	Max temp, humidity	0.634	0.602
M11	Max temp, wind speed	0.734	0.743
M12	Max temp, sunshine hours	0.501	0.430
M13	Humidity, wind speed	0.727	0.760
M14	Humidity, sunshine hours	0.531	0.983
M15	Wind speed, sunshine hours	0.527	0.627
<i>Three Inputs Combinations</i>			
M16	Min temp, max temp, humidity	0.570	0.574
M17	Min temp, max temp, wind speed	0.729	0.722
M18	Min temp, max temp, sunshine hours	0.512	0.447
M19	Min temp, humidity, wind speed	0.726	0.723
M20	Min temp, humidity, sunshine hours	0.337	0.377
M21	Min temp, wind speed, sunshine hours	0.470	0.501
M22	Max temp, humidity, wind speed	0.567	0.566
M23	Max temp, humidity, sunshine hours	0.300	0.239
M24	Max temp, wind speed, sunshine hours	0.409	0.394
M25	Humidity, wind speed, sunshine hours	0.337	0.333
<i>Four Inputs Combinations</i>			
M26	Min temp, max temp, humidity, wind speed	0.577	0.561
M27	Min temp, max temp, humidity, sunshine hours	0.262	0.229
M28	Min temp, max temp, wind speed, sunshine hours	0.382	0.404
M29	Min temp, humidity, wind speed, sunshine hours	0.271	0.238
M30	Max temp, humidity, wind speed, sunshine hours	0.107	0.116
<i>All Inputs</i>			
M31	Min temp, max temp, humidity, wind speed, sunshine hours	<b>0.081</b>	<b>0.087</b>

RMSE = root mean squared error, LSTM = long short-term memory networks, Bi-LSTM = bi-directional long-short term memory networks. The numbers in boldface indicate the best performance, whereas the numbers in boldface and italicized represent the worst performance.

Nonetheless, decision making in such situations is challenging, as the RMSE criterion alone is insufficient as a decision-making tool. To assist in the decision-making process, we used three benefit (the higher numeric values indicate better model performances: R, NS, IOA) and three cost (the lower the numeric values, the better the model performance: NRMSE, MAE, MAD) performance evaluation indices in the decision-making process with the aid of Shannon's entropy. On the testing dataset, we computed the R, NS, IOA, NRMSE, MAE, and MAD criteria for all 31 LSTM and 31 Bi-LSTM models. These evaluation indices were used to rank proposed models using Shannon's entropy-based decision theory. Table 6 shows the ranking results together with the corresponding ranking values.

**Table 6.** Ranking of the LSTM and Bi-LSTM models using Shannon's entropy.

Sl. No.	LSTM		Bi-LSTM	
	Model	Ranking Value	Model	Ranking Value
1	M31	0.996	M31	0.966
2	M30	0.906	M30	0.913
3	M27	0.702	M27	0.704
4	M23	0.687	M23	0.696
5	M20	0.657	M29	0.688
6	M29	0.652	M25	0.642
7	M25	0.640	M20	0.621
8	M28	0.604	M24	0.600
9	M24	0.600	M28	0.594
10	M21	0.584	M12	0.581
11	M12	0.563	M18	0.576
12	M18	0.561	M21	0.563
13	M14	0.560	M26	0.557
14	M22	0.558	M9	0.555
15	M26	0.556	M22	0.551
16	M15	0.555	M16	0.551
17	M9	0.555	M10	0.535
18	M16	0.554	M15	0.522
19	M10	0.535	M17	0.488
20	M11	0.496	M19	0.485
21	M17	0.493	M11	0.482
22	M19	0.491	M7	0.478
23	M13	0.491	M13	0.475
24	M7	0.483	M6	0.462
25	M5	0.482	M2	0.460
26	M6	0.470	M5	0.451
27	M2	0.470	M14	0.384
28	M1	0.415	M1	0.376
29	M8	0.364	M8	0.336
30	M3	0.306	M4	0.311
31	M4	0.209	M3	0.256

It is perceived from the results presented in Table 6 that models that used all five input variables (M31) were the top-ranked predictors, followed by M30, M27, and M23 for both LSTM and Bi-LSTM algorithms. Models M3 and M4 appeared to be the worst performers when using LSTM or Bi-LSTM algorithms for model development. The findings are in accordance with the work of Kisi et al. [37], who indicated that considering all input variables greatly increased the accuracy of the prediction model (radial basis M5Tree) for the data acquired from the three weather stations. Therefore, the results suggest that all input variables would be employed to better predict the daily  $ET_0$  for the meteorological data and the corresponding  $ET_0$  values presented in this study. Consequently, to arrange for an impartial comparison, we developed other prediction modeling algorithms (ANFIS, GPR, M5Tree, MARS, PLR, and SVR) using all five input variables available for the study area. Similar evaluation indices were computed for all the other prediction modeling algorithms proposed in this research. The prediction results are presented in Table 7.

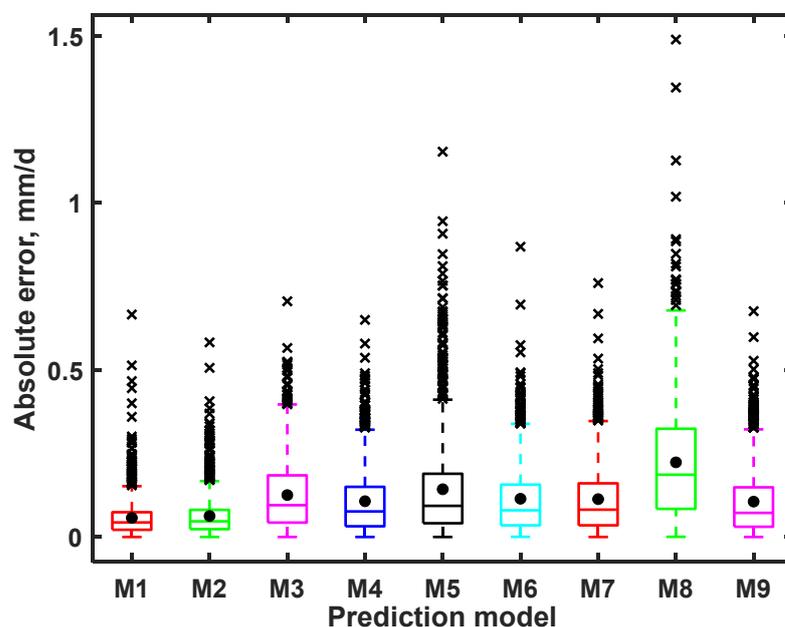
The prediction results in Table 7 indicated that all  $ET_0$  prediction models are reasonably accurate at predicting daily  $ET_0$  values, as evidenced by the different performance indices computed on the testing dataset. While no standalone model exhibited the best performance for all evaluation indices, the individual prediction models provided the estimates of daily  $ET_0$  values superior to others. All  $ET_0$  models had satisfactory prediction accuracy as all models had better (higher) values R, NS, and IOA and lower NRMSE, MAE, and MAD values. LSTM and Bi-LSTM models had superior performance in comparison with others according to all performance evaluation indices. PLR was found to be the worst-performing model.

**Table 7.** Performance indices of the developed ET<sub>0</sub> prediction models for the testing dataset.

Model	Performance Evaluation Indices					
	R	NS	IOA	NRMSE	MAE, mm/d	MAD, mm/d
LSTM	0.998	0.995	0.999	0.021	0.666	0.025
Bi-LSTM	0.998	0.995	0.999	0.023	0.582	0.027
ANFIS	0.991	0.981	0.995	0.043	0.706	0.061
GPR	0.993	0.985	0.996	0.038	0.650	0.052
M5 Tree	0.985	0.970	0.993	0.054	1.153	0.062
MARS_C	0.992	0.983	0.996	0.041	0.869	0.054
MARS_L	0.992	0.983	0.996	0.040	0.760	0.054
PLR	0.973	0.943	0.985	0.075	1.489	0.114
SVR	0.993	0.985	0.996	0.038	0.676	0.050

MARS\_C = piecewise cubic, MARS\_L = piecewise linear.

To provide an additional evaluation regarding the prediction capabilities of the proposed machine learning algorithms (ET<sub>0</sub> prediction models), we presented and compared the absolute error boxplots. Figure 3 illustrates the absolute error boxplots for all the developed models. Absolute error boxplots represent a relative assessment of the statistical distributions of the absolute errors between the FAO-56 PM-computed and model-predicted ET<sub>0</sub> values and supports the evaluation of the degree of general distributions of the inaccuracies provided by the models. The horizontal lines inside the boxplots represent the median values of the absolute errors, whereas the black circles mark the mean (average) of the absolute errors. Absolute error boxplots also demonstrated the superior performance of the LSTM- and Bi-LSTM-based models.



**Figure 3.** Absolute error boxplots. M1–M9 represent LSTM, Bi-LSTM, ANFIS, GPR, M5 tree, MARS\_C, MARS\_L, PLR, and SVR models, respectively.

As far as the two best models are considered, the LSTM model performed better than Bi-LSTM when NRMSE and MAD criteria were considered. In contrast, Bi-LSTM outperformed the LSTM model according to the MAE criterion. On the other hand, both LSTM and Bi-LSTM performed equally well with respect to R, NS, and IOA criteria. Therefore, it is concluded that ET<sub>0</sub> prediction models showed differing precisions depending on the model evaluation indices calculated on the FAO-56 PM and model predicted ET<sub>0</sub> values, which indicated an inconsistency in the model performance when divergent or non-identical evaluation indices were employed. Decision making in this situation is extremely arduous

and can be smoothed by employing a decision theory that integrates a number of different model evaluation indices in decision making. This study employed Shannon's entropy as a decision-making tool.

The ranking of the proposed  $ET_0$  models computed using Shannon's entropy is presented in Table 8. The greater the values of Shannon's entropy, the better the model's performance. Table 8 suggests that LSTM was the top-performing model followed by Bi-LSTM, although the difference between the ranking values of these two models was negligible.

**Table 8.** Shannon's entropy values for different models and their corresponding ranks.

Model	Shannon's Entropy Value	Rank
LSTM	0.979	1
Bi-LSTM	0.978	2
ANFIS	0.807	6
GPR	0.839	3
M5 tree	0.734	8
MARS_C	0.794	7
MARS_L	0.810	5
PLR	0.665	9
SVR	0.836	4

The performance index values for the best model (LSTM) are as follows (Table 7):  $R = 0.998$ ,  $NS = 0.995$ ,  $IOA = 0.999$ ,  $NRMSE = 0.021$ ,  $MAE = 0.666$  mm/d, and  $MAD = 0.025$  mm/d. Although an explicit comparison between the findings of this research and other studies is not possible due to variations in study conditions (modeling tools and geographical locations), the numeric values of various performance indices were observed as being comparable to or even better than those found in the recent literature on  $ET_0$  modeling. For instance, the present study's findings are superior to those obtained by Tao et al. [123], who obtained  $NRMSE$  and  $R^2$  values of 0.043 and 0.97, respectively, using an optimization algorithm-tuned ANFIS model to predict  $ET_0$  in the Bur Dedougou, Burkina Faso. The LSTM model proposed in this study also shows better performance than the optimization algorithm tuned SVR model developed in Ahmadi et al. [32], who obtained the following performance indices at various stations:  $RMSE = 0.540$  mm/d and  $R = 0.983$  at Mashhad station;  $RMSE = 0.404$  mm/d and  $R = 0.980$  at Arak station;  $RMSE = 0.299$  mm/d and  $R = 0.989$  at Shiraz station;  $RMSE = 0.559$  mm/d and  $R = 0.978$  at Tehran station;  $RMSE = 0.457$  mm/d and  $R = 0.962$  at Bandar Abbas station; and  $RMSE = 0.399$  mm/d and  $R = 0.986$  at Yazd station. The present study's findings are also in good agreement with the findings presented in Chia et al. [124], who obtained  $RMSE$  and  $R^2$  values of 0.001–0.197 mm/d and 1.000–0.949, respectively, at three stations using an optimization algorithm-tuned ELM model. The findings are also compared with those presented in Mohammadi and Mehdizadeh [125] that are based on  $RMSE$  and  $R^2$  criteria. Our proposed LSTM model shows superior performance over the best models developed with the daily data in Ferreira and da Cunha [80], who reported  $NS$  values of 0.69 to 0.84 and  $R^2$  values of 0.79 to 0.88. The present study's findings are superior to the optimization algorithm-tuned ELM model developed by Wu et al. [30] that reported  $R^2$  and  $NRMSE$  values of 0.993 and 0.0554, respectively. Elbeltagi et al. [126] reported  $R$  values of 0.94, 0.95, and 0.95 at the Ad Daqahliyah, Kafr ash Shaykh, and Ash Sharqiyah regions, respectively, using the DNN model. These  $R$  values were lower than the  $R$ -value obtained using the proposed LSTM model in the present study ( $R = 0.998$ ). The  $NS$  value of the present study ( $NS = 0.995$ ) is also superior to the  $NS$  value ( $NS = 0.959$ ) presented in Gao et al. [127], indicating the better performance of the proposed LSTM model. The findings of our study are also comparable to those presented in Chia et al. [50], who reported minimum  $MAE$  and  $RMSE$  values of 0.444 mm/d and 0.543 mm/d, respectively.

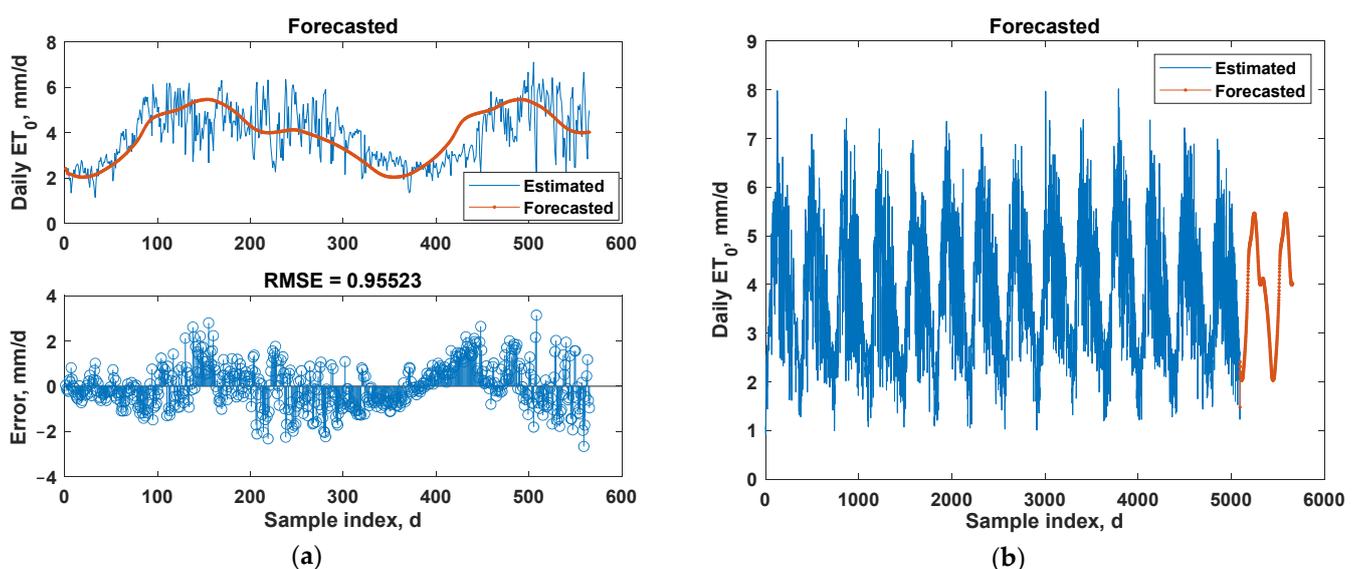
Nevertheless, an apple-to-apple comparison can be performed between the findings obtained from the LSTM model presented in this effort with the models investigated in

Roy et al. [21] (an ensemble of ANFIS models) and in Roy et al. [20] (optimization algorithm tuned ANFIS model). With the optimization algorithm-tuned ANFIS model for the same study area, Roy et al. [20] obtained the following performance indices:  $R = 0.993$ ,  $NS = 0.986$ ,  $IOA = 0.996$ ,  $MAD = 0.054$  mm/d,  $NRMSE = 0.038$ . Our proposed LSTM model performed better than the ANFIS model presented by Roy et al. [20] with respect to all of these performance indices ( $R = 0.998$ ,  $NS = 0.995$ ,  $IOA = 0.999$ ,  $NRMSE = 0.021$ , and  $MAD = 0.025$  mm/d in the present study). Statistical indices provided by the LSTM model ( $R = 0.998$ ,  $NS = 0.995$ ,  $IOA = 0.999$ , and  $MAD = 0.025$  mm/d) proposed in this research also appeared to be superior than those presented by Roy et al. [21] using ensemble of ANFIS models ( $R = 0.993$ ,  $NS = 0.985$ ,  $IOA = 0.996$ , and  $MAD = 0.054$  mm/d). Furthermore, the proposed LSTM model's performance is superior to the performance of the optimization algorithm tune hierarchical fuzzy systems (HFS) presented by Roy et al. [46] with respect to  $R$  (LSTM = 0.998, HFS = 0.987),  $NRMSE$  (LSTM = 0.021, HFS = 0.052), and  $MAD$  (LSTM = 0.025 mm/d, HFS = 0.068 mm/d) criteria.

### 3.2. One-Step-Ahead Prediction of $ET_0$ Using Different Modeling Approaches at the Training Station (Gazipur Sadar)

#### 3.2.1. One-Step-Ahead Forecast Using Sequence to Sequence Regression LSTM (SSR-LSTM) Network

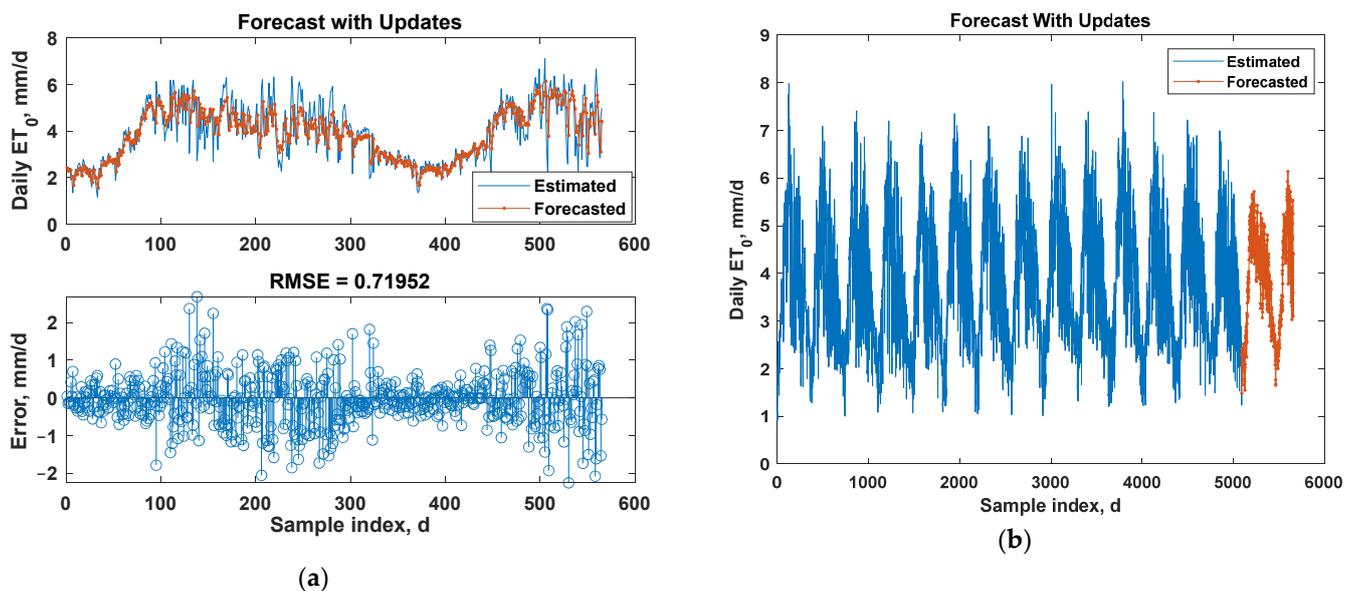
An SSR-LSTM network-based model was trained by employing the historical  $ET_0$  dataset (time series) computed using the FAO-56 PM equation from the meteorological variables. In an SSR-LSTM model, the outputs from the model correspond to the training sequences ( $ET_0$  time series) with  $ET_0$  values moved to a one-time step ahead. At every time step of the  $ET_0$  sequence, an SSR-LSTM network learns how to predict  $ET_0$  values for the next time step. For training the proposed SSR-LSTM model, the historical  $ET_0$  time series was partitioned into training and test sets (90% of the entire data was used for training, whereas the remaining 10% was used for testing the model). Model parameters including the number of hidden layers and neurons were decided upon by conducting several trials. An SSR-LSTM model with one hidden layer having 200 hidden neurons in the hidden layer provided the best results for both the model training and testing phases. The optimal values of other model parameters were solver = 'adam', number of epochs = 250, gradient threshold = 1, initial learning rate = 0.005, and multiplying factor for the learn rate dropping = 0.2. Model performance is presented in Figure 4.



**Figure 4.** SSR-LSTM performance: (a) estimated (FAO-56 PM-computed) and SSR-LSTM-forecasted  $ET_0$  values for the test dataset; (b) future projections beyond the observed  $ET_0$  values.

It is observed from Figure 4 that even though the SSR-LSTM model adequately apprehended the trends of the  $ET_0$  time series for the test set of the data (Figure 4b), the SSR-LSTM forecasts were comparatively flat compared to the original  $ET_0$  time-series data (Figure 4a). This necessitates the improvement in the forecasting performance of the initial SSR-LSTM model. One way of improving performance is to update the SSR-LSTM network state using the observed  $ET_0$  values instead of the predicted  $ET_0$  values. Resetting the network's state is used in this study to prevent previous predictions from impacting the results.

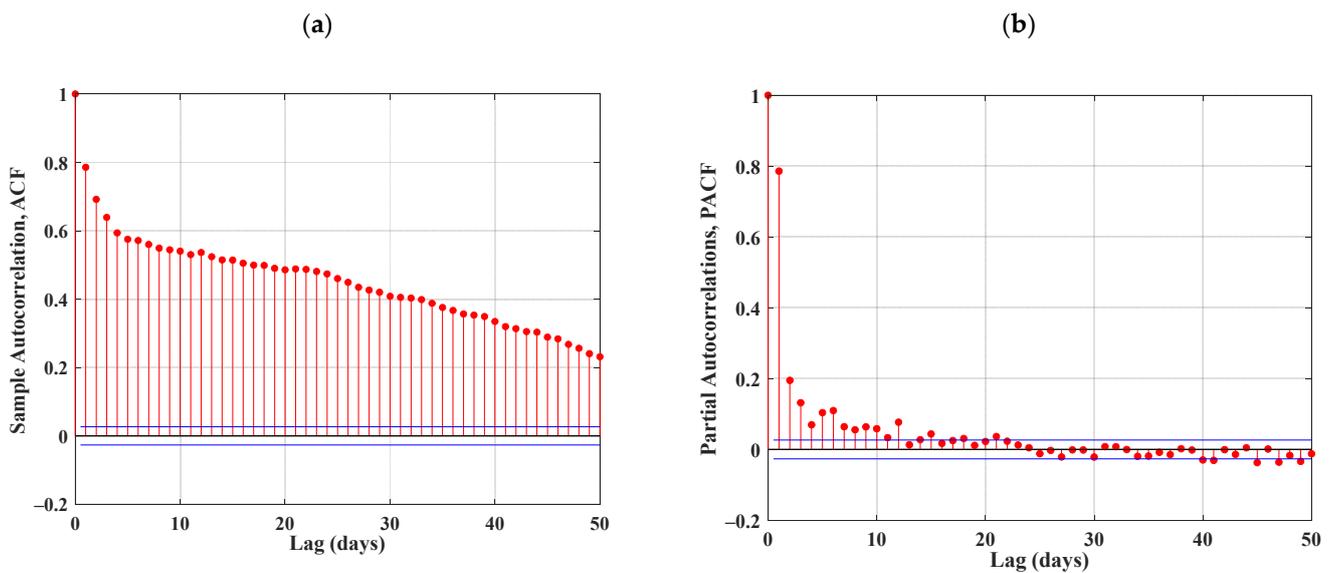
This was performed by resetting the network state in order to prevent previous predictions from affecting the predictions on the new dataset. The forecasting results obtained from the updated network state of the SSR-LSTM model are presented in Figure 5.



**Figure 5.** SSR-LSTM performance after network updating: (a) estimated (FAO-56 PM-computed) and SSR-LSTM-forecasted  $ET_0$  values for the test dataset; (b) future projections beyond the observed  $ET_0$  values.

### 3.2.2. One-Step-Ahead Forecast Using ANFIS, LSTM, and Bi-LSTM Models

For developing ANFIS, LSTM, and Bi-LSTM models to provide one-step-ahead forecasts, we computed PACF functions to obtain time-lagged information from the daily  $ET_0$  time series. This information obtained from the PACF functions was employed to assess the time-based dependences between  $ET_0$  for a present day ( $ET_t$ ) and the  $ET_0$  values at a particular day in a prior period (e.g., at a lag time of  $ET_{t-1}$ ,  $ET_{t-2}$ ,  $ET_{t-3}$ ,  $ET_{t-4}$ , and  $ET_{t-5}$ ). These time-based dependences in the  $ET_0$  time series were assessed for 50 time lags (e.g.,  $ET_{t-1}$  to  $ET_{t-50}$ ), as shown in Figure 6. In Figure 6, the blue lines indicate the 95% confidence band, whereas the red vertical lines represent the corresponding values of ACF and PACF. Time-lagged  $ET_0$  values serve as the inputs to the ANFIS, LSTM, and Bi-LSTM models to forecast one-day-ahead  $ET_0$  values (outputs from the models). The optimal sets of time-lagged  $ET_0$  inputs for model development were selected carefully after observing the PACF functions.



**Figure 6.** ACF (a) and PACF (b) plots of the  $ET_0$  time series for 50 lags at Gazipur station.

A careful observation of the PACF plot shown in Figure 6 determines the following time-lagged  $ET_0$  values as inputs to the developed models:

$$ET_t, ET_{t-1}, ET_{t-2}, ET_{t-3}, ET_{t-4}, ET_{t-5}, ET_{t-6}, ET_{t-7}, ET_{t-8}, ET_{t-9}, ET_{t-10}, ET_{t-11}$$

The outputs from the developed models were  $ET_{t+1}$  (one-day-ahead  $ET_0$  values).

**ANFIS outputs:** The results of the one-step-ahead forecast using the ANFIS model are presented in Figure 7 and Table 9. Figure 7 presents ANFIS forecasts through scatter plots and hydrographs, whereas Table 8 shows model prediction capabilities based on several statistical performance evaluation indices. Hydrographs and scatterplots presented in Figure 7 demonstrate the reasonable precision of the one-day-ahead  $ET_0$  forecasts by the ANFIS model. It is observed from Figure 7 that the training and test RMSE (0.759 and 0.789 mm/d, respectively, for the training and testing phases) did not vary considerably, which indicates a better model fit without model over- or under-fitting. Figure 7 also indicates acceptable values of training and test R-values (0.825 and 0.755, respectively, for the training and testing phases). As far as other performance evaluation indices are considered, the ANFIS model produced the following values of performance measures computed on the test dataset: NS = 0.567, IOA = 0.858, NRMSE = 0.207 mm/d, MAE = 2.710 mm/d, and MAD = 0.308 mm/d.

**LSTM and Bi-LSTM outputs:** Comparison of FAO-56 PM-calculated and model-predicted  $ET_0$  values, error plots, and projected (one-step-ahead)  $ET_0$  values produced by the LSTM and Bi-LSTM models are presented in Figures 8 and 9, respectively. It is noticed from Figures 8 and 9 that both LSTM and Bi-LSTM models captured the trend of the  $ET_0$  time series precisely and that Bi-LSTM model forecasts were superior to those of the LSTM model. The performance evaluation results based on several statistical performance evaluation indices are presented in Table 9. The LSTM model produced the following values of performance measures computed on the test dataset: R = 0.698, NS = 0.698, IOA = 0.429, NRMSE = 0.237 mm/d, MAE = 3.047 mm/d, and MAD = 0.334 mm/d. On the other hand, the Bi-LSTM model produced the following values of performance measures computed on the test dataset: R = 0.999, NS = 0.998, IOA = 0.999, NRMSE = 0.014 mm/d, MAE = 0.491 mm/d, and MAD = 0.017 mm/d.

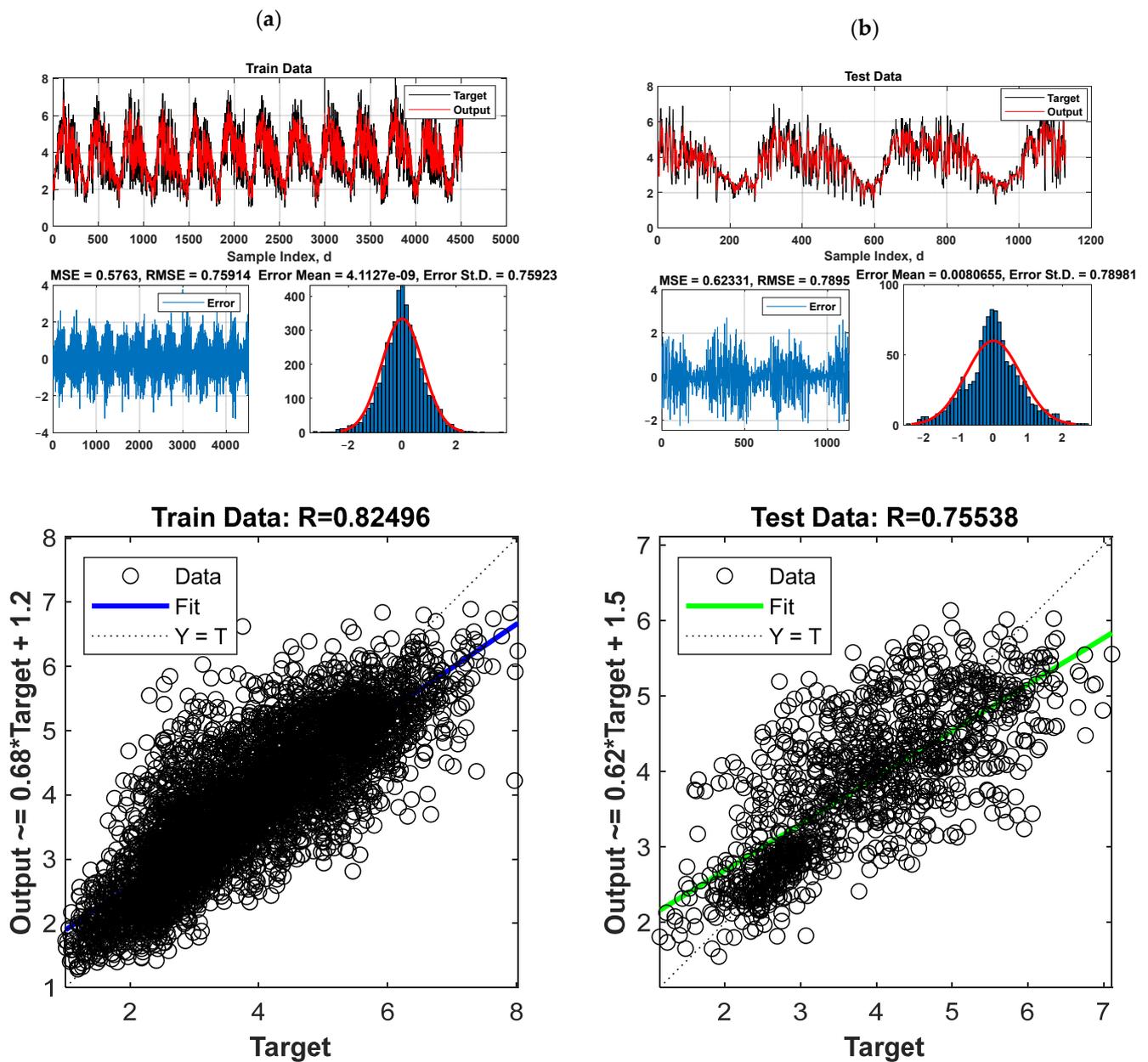


Figure 7. Scatter plots and regression plots for the values of FAO-56 PM-calculated  $ET_0$  and ANFIS-forecasted  $ET_0$  for the training (a) and testing (b) phases.

Table 9. Performance indices of the one-day-ahead  $ET_0$  prediction models for the testing dataset.

Model	Performance Evaluation Indices					
	R	NS	IOA	NRMSE	MAE, mm/d	MAD, mm/d
ANFIS	0.755	0.567	0.858	0.207	2.710	0.308
Bi-LSTM	0.999	0.998	0.999	0.014	0.491	0.017
LSTM	0.698	0.429	0.833	0.237	3.047	0.334
SSR-LSTM	0.818	0.666	0.898	0.184	2.687	0.279

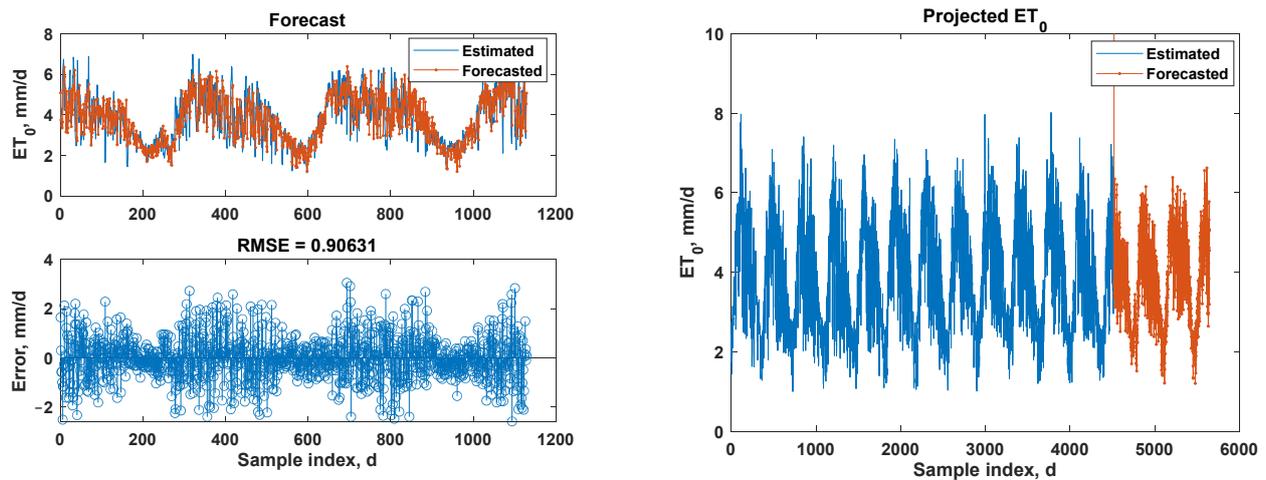


Figure 8. FAO-56 PM-calculated and LSTM-projected  $ET_0$  values with error plots computed on the test dataset.

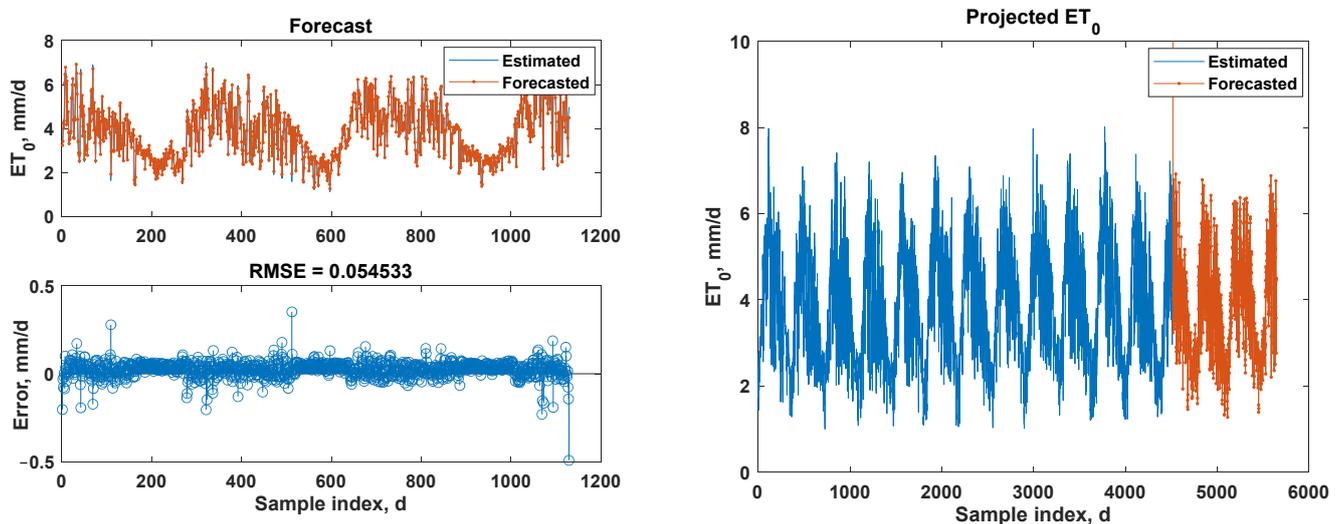


Figure 9. FAO-56 PM-calculated and Bi-LSTM-projected  $ET_0$  values with error plots computed on the test dataset.

It is observed from Table 9 that the Bi-LSTM model provided a superior performance compared to the other models (SSR-LSTM, ANFIS, and LSTM) according to the statistical indices computed on the test dataset. It is noted that the prediction results with respect to the calculated performance indices did not demonstrate a considerable inconsistency. However, to reach a solid conclusion regarding the best-performing model, we applied the concept of Shannon’s entropy to provide a performance ranking (Table 10). It is observed from Table 10 that Bi-LSTM appeared to be the best performer, while SSR-LSTM, ANFIS, and LSTM held the second, third, and fourth positions, respectively. Therefore, according to the performance results for one-step-ahead forecasting, the best-performing Bi-LSTM model was employed to provide multi-step (5 day)-ahead forecasting.

Table 10. Shannon’s entropy-based model ranking for one-day-ahead  $ET_0$  forecasts.

Model	Shannon’s Entropy Value	Rank
Bi-LSTM	1.00	1
SSR-LSTM	0.30	2
ANFIS	0.27	3
LSTM	0.24	4

### 3.3. Multi-Step (5 Day-Ahead) Forecasting Using the Bi-LSTM Model

The forecasting performance of the developed  $ET_0$  prediction model using the Bi-LSTM algorithm was evaluated using several statistical performance indices on the test dataset. However, to ascertain that no model over- or under-fitting occurred, we quantitatively evaluated the results obtained from both the training and validation phases. Five Bi-LSTM models were developed to forecast 1, 2, 3, 4, and 5 day-ahead  $ET_0$  values. For all models, the selected time-lagged variables were served as inputs to the Bi-LSTM models. Table 11 presents the performances of the developed Bi-LSTM models on the training and validation datasets. It is evident from Table 11 that the absolute variances between the training and validation performances increased with the increase in the forecasting horizon. Overall, the training performances were satisfactory for all forecasting horizons.

**Table 11.** Training and validation performances of the developed Bi-LSTM models at Gazipur station.

Forecasting Horizon	Training RMSE, mm/d	Validation RMSE, mm/d
1 day	0.08	0.11
2 days	0.12	0.17
3 days	0.09	0.18
4 days	0.10	0.22
5 days	0.10	0.28

The trained and validated Bi-LSTM models were then used to forecast  $ET_0$  values on the test dataset, which were selected from the entire dataset. Testing performances were assessed using several evaluation indices, as shown in Table 12. It is perceived from Table 12 that the forecasting horizon greatly influenced the forecasting accuracies. The accuracy decreased with the increase in the forecasting horizon as in the case of the training and validation performances. However, the overall performances of the Bi-LSTM model for all forecasting horizons showed particularly good performance, as indicated by the computed statistical performance evaluation indices. The performance of the developed models was also assessed using line graphs and error plots as shown in Figure 10.

**Table 12.** Multi-day-ahead forecasting performance of the Bi-LSTM model on the test dataset at Gazipur station.

Indices	Forecasting Horizon				
	1 Day	2 Days	3 Days	4 Days	5 Days
RMSE, mm/d	0.11	0.17	0.18	0.22	0.28
NRMSE	0.03	0.04	0.05	0.06	0.07
R	1.00	0.99	0.99	0.98	0.97
MAD, mm/d	0.03	0.04	0.04	0.06	0.08
MAE, mm/d	0.07	0.08	0.10	0.13	0.17
NS	0.99	0.98	0.98	0.97	0.95
IOA	1.00	0.99	0.99	0.99	0.99

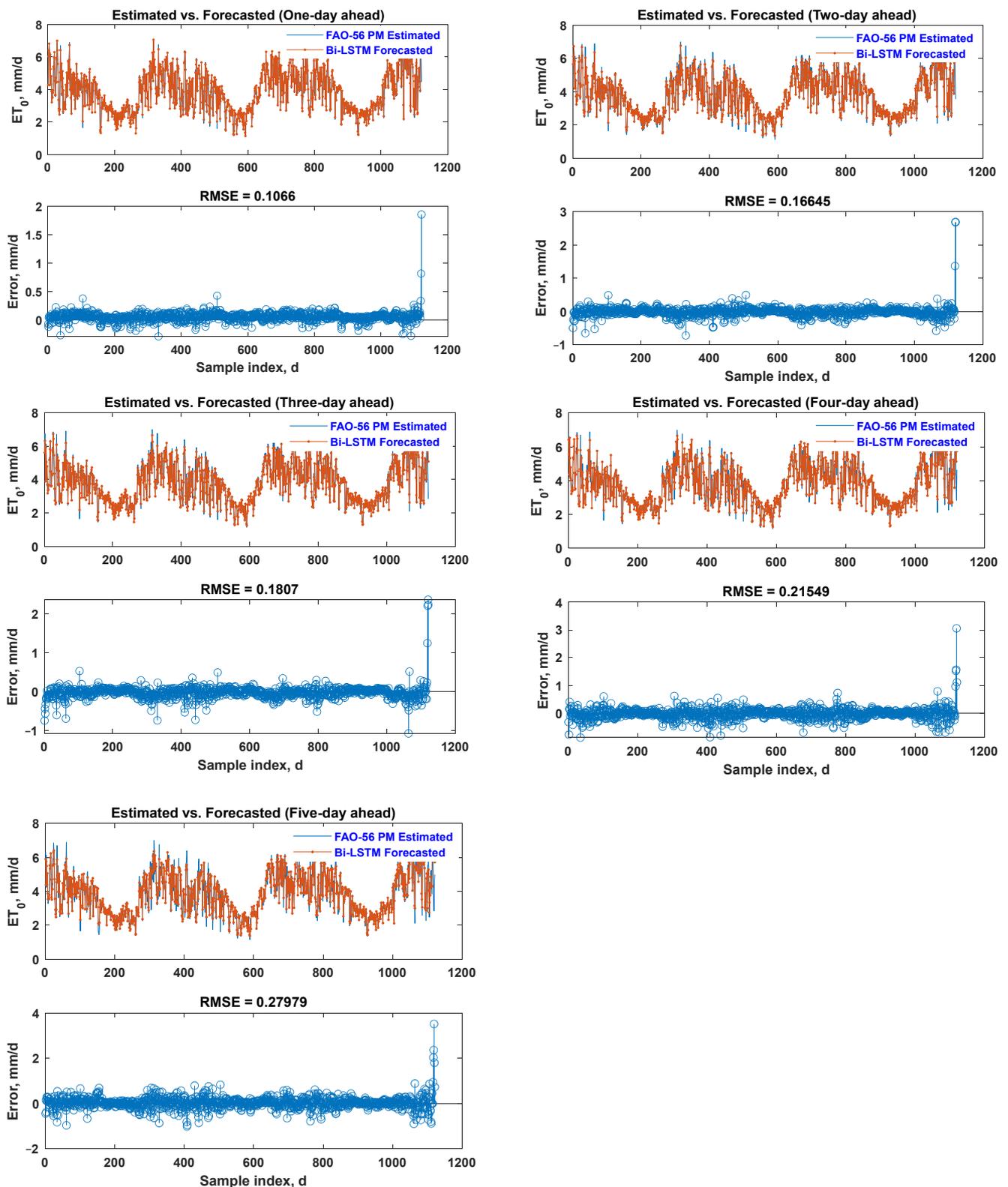


Figure 10. Line graph and error plots for 1, 2, 3, 4, and 5 day-ahead forecasting at Gazipur station.

It is observed from Table 12 that the Bi-LSTM model showed reasonably good performance, as evidenced by the computed performance indices. It produced lower values of cost indices (RMSE, NRMSE, MAD, and MAE) as well as higher values of benefit indices (R, NS, IOA). However, it is noted that the forecasting accuracy largely depended on the forecasting horizon, i.e., the sequence of forecasting accuracies are as follows:

1 day > 2 days > 3 days > 4 days > 5 days. This finding is in good agreement with the work of Yin et al. [128], who also stated that forecasting accuracy decreased with the increased forecasting horizon. Nevertheless, the forecasting accuracy of the Bi-LSTM model at 5 days ahead was also found acceptable for deep learning-based modeling of  $ET_0$ . Ferreira and da Cunha [82] also reported better deep learning model performance (CNN-LSTM) on the first and second forecasting days. Our findings using the Bi-LSTM model (RMSE = 0.11–0.28 mm/d) outperformed the CNN-LSTM model proposed by Ferreira and da Cunha [82] (mean RMSE values of 0.87 to 0.88 mm/d) with respect to RMSE criterion. Our proposed Bi-LSTM model performed better than the Bi-LSTM model proposed by Yin et al. [128] with respect to RMSE, R, and NS criteria. For instance, for 1 day-ahead forecasting, Yin et al. [128] obtained RMSE, R, and NS values of 0.159 mm/d, 0.992, and 0.988, respectively, whereas the values of RMSE, R, and NS in our study were found to be 0.11 mm/d, 1.00, and 0.99, respectively. Similarly, our proposed Bi-LSTM model outperformed the Bi-LSTM model presented by Yin et al. [128] for 4 day-ahead  $ET_0$  forecasting. Moreover, our results also showed superior performance than the Bi-LSTM model results presented by Roy [81] in terms of R and IOA criteria for 1 day-ahead  $ET_0$  forecasting. Roy [81] reported R and IOA values of 0.698 to 0.999 and 0.833 to 0.999, respectively, while the present study provided R and IOA values of 1.00 and 1.00, respectively. Therefore, it can be inferred that our proposed Bi-LSTM model is suitable for forecasting multi-step-ahead  $ET_0$  values quite efficiently and precisely. It is noted that the Bi-LSTM model produced a slightly higher forecast error, especially at the end of the  $ET_0$  time series. This comparatively big error at the end of the dataset may have arisen from higher values of  $ET_0$  (outliers), which was not smoothed in order to evaluate the performance of the proposed modeling approaches for datasets containing outliers. Nevertheless, these values are still acceptable in the context of modeling  $ET_0$  using machine learning approaches.

### 3.4. Generalization Capability of the Proposed Best $ET_0$ Prediction Models

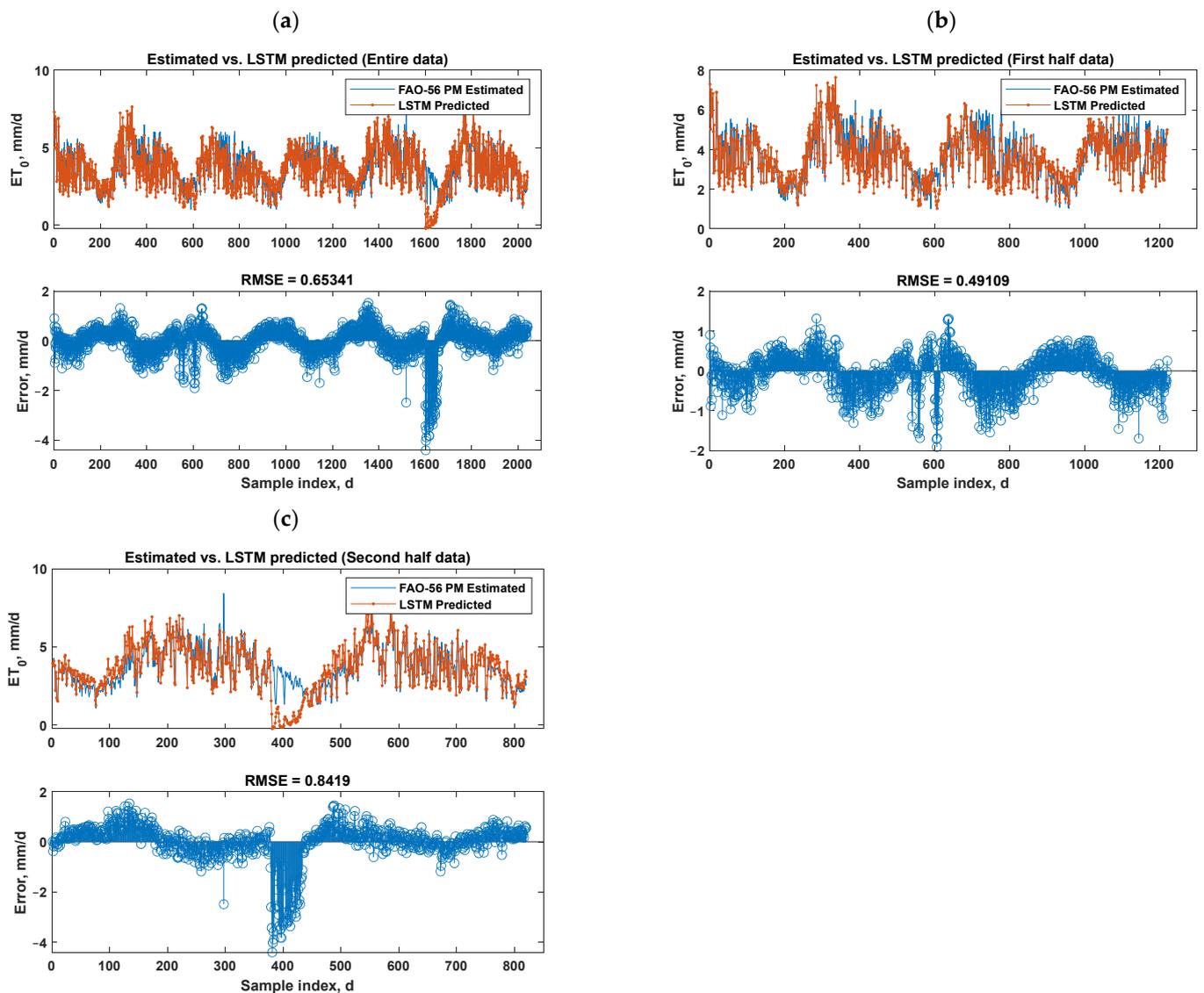
The validation of the proposed best models (LSTM for daily predictions and Bi-LSTM for multi-step-ahead forecasts) was performed using data obtained from a new test station at which the models were not developed. The entire dataset of the test station (Ishurdi station) was split into three separate sets, each of which was employed to validate the models developed at the training station (Gazipur Sadar station). These three standalone datasets were fed into the LSTM and Bi-LSTM models to predict daily  $ET_0$  values and forecast multi-day-ahead  $ET_0$ , respectively. The outputs from the models were weighed against the FAO-56 PM-computed  $ET_0$  values using numerous statistical performance evaluation indices.

#### 3.4.1. Generalization Capability of Proposed Best LSTM Model: Daily Prediction of $ET_0$

Table 13 summarizes the evaluation results for a variety of performance indices. The LSTM model exhibited a reasonably good performance at the test station data's three different sets (entire, first half, and second half). The computed performance indices indicated a satisfactory performance of the proposed LSTM model. It produced reasonably higher values of benefit indices (R, NS, and IOA) and lower values of the cost indices (RMSE, NRMSE, MAD, and MAE) for the entire, the first half, and the second half of the test station data. It is also observed that the first half of the dataset produced relatively better performance when compared to that of the second half and the entire dataset. Overall, the performance is satisfactory. On this basis, it is arguably concluded that the proposed LSTM model at Gazipur Sadar station can predict daily  $ET_0$  values at Ishurdi station without developing a model at Ishurdi station. Additionally, performance data were presented using scatter and error plots, as illustrated in Figure 11, which depict the distribution of errors at individual data points.

**Table 13.** Performance of the LSTM model for predicting daily  $ET_0$  values on the Ishurdi dataset.

Performance Indices	Entire Dataset	First Half Data	Second Half Data
RMSE, mm/d	0.65	0.49	0.84
NRMSE	0.18	0.13	0.23
R	0.87	0.92	0.83
MAD, mm/d	0.18	0.18	0.20
MAE, mm/d	0.44	0.39	0.52
NS	0.72	0.84	0.57
IOA	0.97	0.98	0.96



**Figure 11.** Line graph and error plots of FAO-56 PM-computed and LSTM-predicted daily  $ET_0$  at Ishurdi station: (a) entire dataset, (b) first half of the dataset, and (c) second half of the dataset.

### 3.4.2. Generalization Capability of Proposed Best Bi-LSTM Model: Multi-Step (Multi-Day)-Ahead $ET_0$ Forecasting

For multi-step (multi-day)-ahead  $ET_0$  forecasting, new Bi-LSTM models were developed because the nature of data was different. However, a similar model structure and parameters as in the case of Gazipur station were used. As a Bi-LSTM model performed better for one-step-ahead prediction at Gazipur station, the Bi-LSTM model was used to develop models for forecasting 1, 2, 3, and 5 day-ahead  $ET_0$  values at the Ishurdi station.

For this, time-lagged information from the  $ET_0$  time series was collected for 50 lags. The most significant input variables were determined by observing partial autocorrelation functions of the lagged time series, as shown in Figure 12.

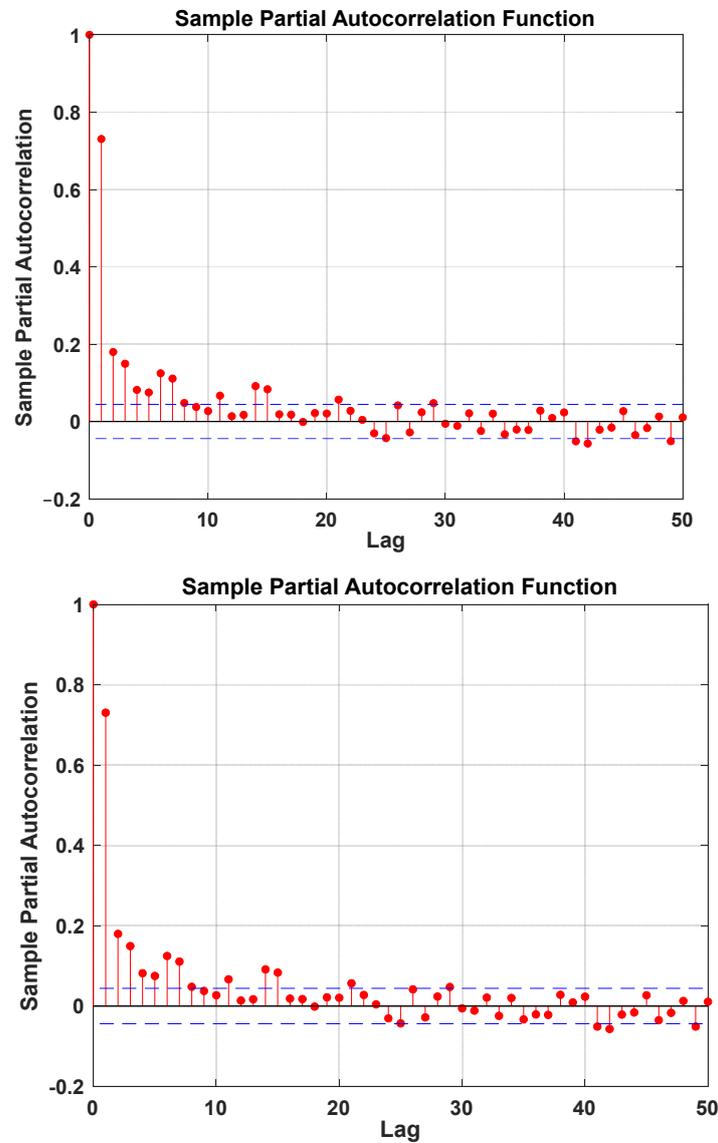


Figure 12. Sample partial autocorrelation functions of the lagged  $ET_0$  time series at Ishurdi station.

Five Bi-LSTM models were developed to forecast 1, 2, 3, 4, and 5 day-ahead  $ET_0$  forecasting. For all models, the selected time-lagged variables were served as inputs to the Bi-LSTM models. Table 14 presents the performances of the proposed Bi-LSTM models at the training and validation datasets. The absolute variances between the training and validation performances increased with the increase in the forecasting horizon. Overall, the training performances were satisfactory for all forecasting horizons.

Table 14. Training and validation performances of the developed Bi-LSTM models at Ishurdi station.

Forecasting Horizon	Training RMSE, mm/d	Validation RMSE, mm/d
1 day	0.09	0.12
2 days	0.10	0.17
3 days	0.11	0.29
4 days	0.12	0.56
5 days	0.10	0.73

The trained and validated Bi-LSTM models were then used to forecast  $ET_0$  values on the test dataset, which were selected from the entire dataset. Testing performances were assessed using several statistical index values, as shown in Table 15. The forecasting horizon greatly influenced the forecasting accuracies. The accuracy decreased with the increase in the forecasting horizon, as in the case of the training and validation performances. However, the overall performances of the Bi-LSTM model for all forecasting horizons showed particularly good performance, as indicated by the computed statistical performance evaluation indices. Performance evaluation results of the developed models were also assessed with the aid of line graphs and error plots, as shown in Figure 13. The performance results illustrated in Figure 13 were in good agreement with the statistical index values presented in Table 15. As observed in the line graphs and error plots, forecasting accuracy largely depended on the forecasting horizon: forecasting accuracy decreased with increases in the forecasting horizon.

**Table 15.** Multi-day-ahead forecasting performance of the Bi-LSTM model on the test dataset at Ishurdi station.

Indices	Forecasting Horizon				
	1 Day	2 Days	3 Days	4 Days	5 Days
RMSE, mm/d	0.12	0.17	0.29	0.56	0.73
NRMSE	0.03	0.05	0.08	0.16	0.20
R	1.00	0.99	0.98	0.90	0.86
MAD, mm/d	0.04	0.05	0.08	0.14	0.24
MAE, mm/d	0.09	0.12	0.19	0.37	0.56
NS	0.99	0.98	0.95	0.81	0.69
IOA	1.00	1.00	0.99	0.95	0.91

It is observed from Figure that 1 day- and 2 day-ahead forecasting results were relatively better when compared to the results produced in three, four, and five day-ahead forecasts with respect to the RMSE criterion. A closer look at the line graphs also revealed the superiority of one day- and two day-ahead forecasts over the other three forecasting horizons and that Bi-LSTM models captured the lower values of the  $ET_0$  time series quite accurately in comparison with the higher values for one day-, two day-, and three day-ahead forecasts. While producing acceptable results, the Bi-LSTM models followed similar trends for both the lower and higher values in the  $ET_0$  time series in the case of the four day- and five day-ahead forecasts. It is also perceived from the line graphs that errors were relatively smaller at the end of the time series for the one day- and two day-ahead forecasts, while the Bi-LSTM models produced relatively higher errors at the end of the dataset for the three, four, and five day-ahead forecasts. Although performed differently at different forecast horizons, the Bi-LSTM model forecasts were quite accurate and closer to the FAO-56 PM-estimated  $ET_0$  values. This is also evident from the statistical performance evaluation indices presented in Table 15. In particular, the NRMSE values of 0.03, 0.05, 0.08, 0.16, and 0.20 for the one, two, three, four, and five day-ahead forecasts, respectively, revealed the reasonable accurate forecasts of the proposed Bi-LSTM model. A model's performance is said to be excellent when the NRMSE value is lower than 0.1, good when the NRMSE value is between 0.1 and 0.2, fair when the NRMSE value is between 0.2 and 0.3, poor when the NRMSE is greater than 0.3 [129,130].

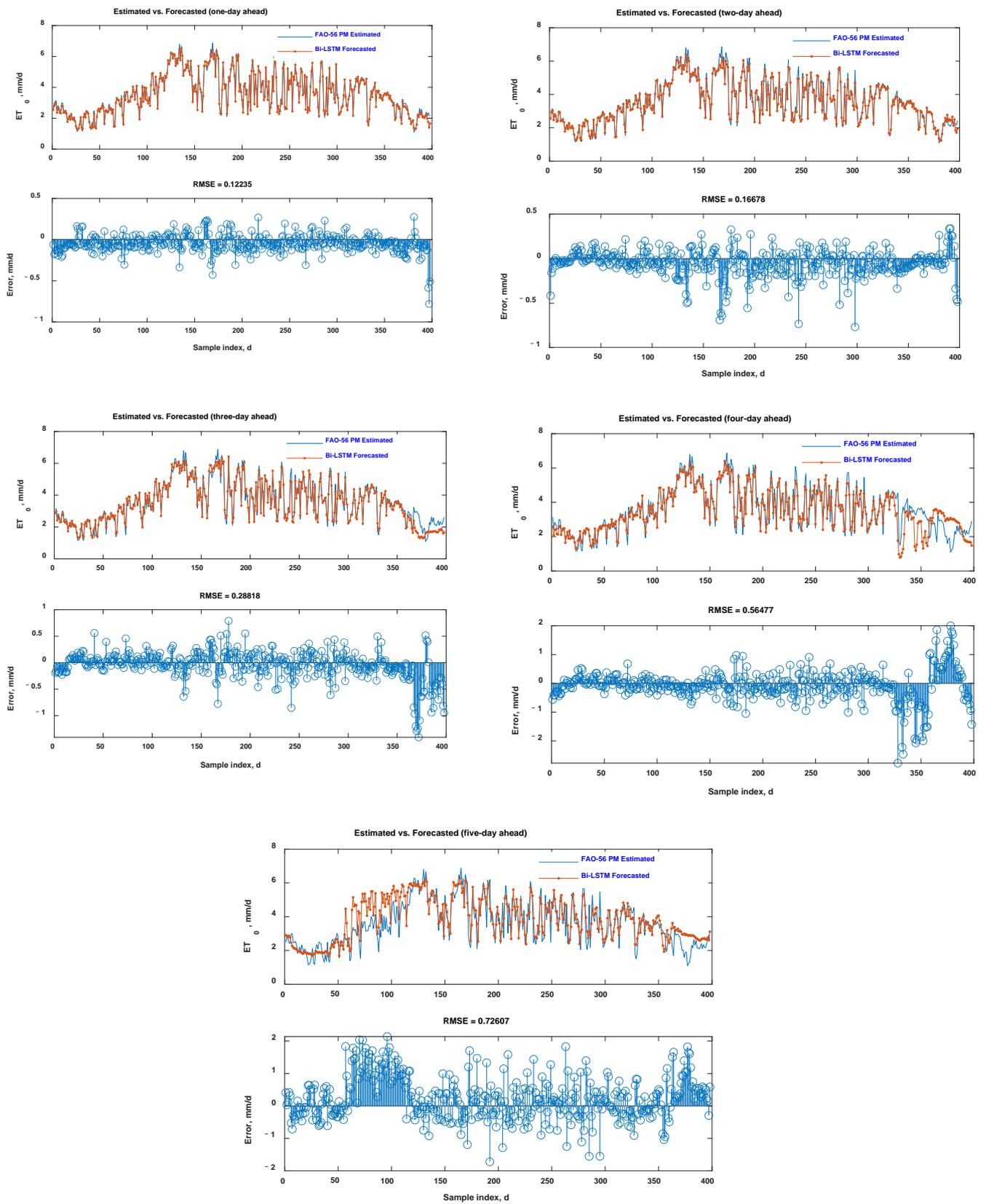


Figure 13. Line graph and error plots for 1, 2, 3, 4, and 5 day-ahead forecasting at Ishurdi station.

#### 4. Conclusions

Precise prediction and forecasting of  $ET_0$  have been a critical and emerging first step for developing a justifiable and effective irrigation scheduling plan. This research provided a selection of the best machine and deep learning algorithms to develop robust prediction and forecasting tools for daily and multi-step (5 day)-ahead  $ET_0$  prediction and forecasting, respectively. The selection results indicated the superiority of the LSTM model for daily  $ET_0$  predictions, whereas for multi-step-ahead forecasting, the Bi-LSTM model provided superior performance. For daily  $ET_0$  prediction, a number of meteorological variables were used as inputs to the model, whereas the computed  $ET_0$  values were used as outputs from the model. For multi-step (5 day)-ahead forecasting, the appropriate daily time-lagged  $ET_0$  values were used as inputs to the Bi-LSTM model, and the outputs from the Bi-LSTM model were the one, two, three, four, and five step-ahead  $ET_0$  values. On the basis of the results of the one-step-ahead prediction performed previously for model selection, we found that the Bi-LSTM model was further employed to provide multi-step (5 day)-ahead forecasting. Results revealed the suitability of the Bi-LSTM model in predicting multi-step-ahead  $ET_0$  values.

In a further step, best models for daily prediction (LSTM) and multi-step-ahead forecasting (Bi-LSTM) were used to generalize the  $ET_0$  values for the data obtained from a different weather station, for which the models were neither trained nor validated. More specifically, the LSTM network was used to generalize the daily  $ET_0$  predictions in a nearby meteorological station without developing a model for that station. On the other hand, the Bi-LSTM model was developed for the Ishurdi station to forecast 1, 2, 3, 4, and 5 day-ahead  $ET_0$  forecasting. The relatively low errors obtained by the LSTM and Bi-LSTM approaches led to a good fit of the models in predicting daily  $ET_0$  values and forecasting multi-step-ahead  $ET_0$  values. This can be expected to be very useful in the practice of irrigation water management, for which  $ET_0$  is an important parameter.

**Author Contributions:** Conceptualization, supervision, methodology, formal analysis, writing—original draft preparation, writing—review and editing, D.K.R., T.K.S., S.S.A.K.; data curation, project administration, investigation, T.G., M.A.M. (Md Abdul Muktedir); conceptualization, funding acquisition, writing—review and editing, writing—original draft preparation, H.M.A.-G., A.A., A.Z.D., A.A.E.-S., M.A.M. (Mohamed A. Mattar). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was financially supported by the Vice Deanship of Research Chairs at King Saud University.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** This work was financially supported by the Vice Deanship of Research Chairs at King Saud University.

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

#### References

1. Liu, S.M.; Xu, Z.W.; Zhu, Z.L.; Jia, Z.Z.; Zhu, M.J. Measurements of evapotranspiration from eddy-covariance systems and large aperture scintillometers in the Hai River Basin, China. *J. Hydrol.* **2013**, *487*, 24–38. [[CrossRef](#)]
2. Kisi, O. Modeling reference evapotranspiration using three different heuristic regression approaches. *Agric. Water Manag.* **2016**, *169*, 162–172.
3. Kool, D.; Agam, N.; Lazarovitch, N.; Heitman, J.L.; Sauer, T.J.; Ben-Gal, A. A review of approaches for evapotranspiration partitioning. *Agric. For. Meteorol.* **2014**, *184*, 56–70. [[CrossRef](#)]
4. Martí, P.; González-Altozano, P.; López-Urrea, R.; Mancha, L.A.; Shiri, J. Modeling reference evapotranspiration with calculated targets. Assessment and implications. *Agric. Water Manag.* **2015**, *149*, 81–90. [[CrossRef](#)]

5. Zhang, B.; Liu, Y.; Xu, D.; Zhao, N.; Lei, B.; Rosa, R.D.; Paredes, P.; Paço, T.A.; Pereira, L.S. The dual crop coefficient approach to estimate and partitioning evapotranspiration of the winter wheat–summer maize crop sequence in North China Plain. *Irrig. Sci.* **2013**, *31*, 1303–1316. [[CrossRef](#)]
6. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration-Guidelines for Computing crop Water Requirements*; FAO: Rome, Italy, 1998.
7. Ding, R.; Kang, S.; Zhang, Y.; Hao, X.; Tong, L.; Du, T. Partitioning evapotranspiration into soil evaporation and transpiration using a modified dual crop coefficient model in irrigated maize field with ground-mulching. *Agric. Water Manag.* **2013**, *127*, 85–96. [[CrossRef](#)]
8. Landeras, G.; Ortiz-Barredo, A.; López, J.J. 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.* **2008**, *95*, 553–565. [[CrossRef](#)]
9. Ferreira, L.B.; da Cunha, F.F.; Fernandes Filho, E.I. Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil. *Agric. Water Manag.* **2022**, *259*, 107281. [[CrossRef](#)]
10. Kelley, J.; Pardyjak, E.R. Using neural networks to estimate site-specific crop evapotranspiration with low-cost sensors. *Agronomy* **2019**, *9*, 108. [[CrossRef](#)]
11. Yassin, M.A.; Alazba, A.A.; Mattar, M.A. Modelling daily evapotranspiration using artificial neural networks under hyper arid conditions. *Pak. J. Agric. Sci.* **2016**, *53*, 695–712.
12. Yassin, M.A.; Alazba, A.A.; Mattar, M.A. Artificial neural networks versus gene expression programming forestimating reference evapotranspiration in arid climate. *Agric. Water Manag.* **2016**, *163*, 110–124. [[CrossRef](#)]
13. Kumar, M.; Raghuwanshi, N.S.; Singh, R.; Wallender, W.W.; Pruitt, W.O. Estimating evapotranspiration using artificial neural network. *J. Irrig. Drain. Eng.* **2002**, *128*, 224–233. [[CrossRef](#)]
14. Feng, Y.; Cui, N.; Gong, D.; Zhang, Q.; Zhao, L. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agric. Water Manag.* **2017**, *193*, 163–173. [[CrossRef](#)]
15. Feng, Y.; Peng, Y.; Cui, N.; Gong, D.; Zhang, K. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput. Electron. Agric.* **2017**, *136*, 71–78. [[CrossRef](#)]
16. Gocić, M.; Arab Amiri, M. Reference evapotranspiration prediction using neural networks and optimum time lags. *Water Resour. Manag.* **2021**, *35*, 1913–1926. [[CrossRef](#)]
17. Doğan, E. Reference evapotranspiration estimation using adaptive neuro-fuzzy inference systems. *Irrig. Drain.* **2009**, *58*, 617–628. [[CrossRef](#)]
18. Dou, X.; Yang, Y. Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Comput. Electron. Agric.* **2018**, *148*, 95–106. [[CrossRef](#)]
19. Gavili, S.; Sanikhani, H.; Kisi, O.; Mahmoudi, M.H. Evaluation of several soft computing methods in monthly evapotranspiration modelling. *Meteorol. Appl.* **2018**, *25*, 128–138. [[CrossRef](#)]
20. Roy, D.K.; Lal, A.; Sarker, K.K.; Saha, K.K.; Datta, B. Optimization algorithms as training approaches for prediction of reference evapotranspiration using adaptive neuro fuzzy inference system. *Agric. Water Manag.* **2021**, *255*, 107003. [[CrossRef](#)]
21. Roy, D.K.; Barzegar, R.; Quilty, J.; Adamowski, J. Using ensembles of adaptive neuro-fuzzy inference system and optimization algorithms to predict reference evapotranspiration in subtropical climatic zones. *J. Hydrol.* **2020**, *591*, 125509. [[CrossRef](#)]
22. Shiri, J.; Marti, P.; Nazemi, A.H.; Sadraddini, A.A.; Kisi, O.; Landeras, G.; Fakheri Fard, A. Local vs. external training of neuro-fuzzy and neural networks models for estimating reference evapotranspiration assessed through k-fold testing. *Hydrol. Res.* **2013**, *46*, 72–88. [[CrossRef](#)]
23. Tabari, H.; Kisi, O.; Ezani, A.; Hosseinzadeh, T.P. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *J. Hydrol.* **2012**, *444–445*, 78–89. [[CrossRef](#)]
24. Huang, G.; Wu, L.; Ma, X.; Zhang, W.; Fan, J.; Yu, X.; Zeng, W.; Zhou, H. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* **2019**, *574*, 1029–1041. [[CrossRef](#)]
25. Wu, T.; Zhang, W.; Jiao, X.; Guo, W.; Alhaj Hamoud, Y. Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. *Comput. Electron. Agric.* **2021**, *184*, 106039. [[CrossRef](#)]
26. Zhang, Y.; Zhao, Z.; Zheng, J. CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. *J. Hydrol.* **2020**, *588*, 125087. [[CrossRef](#)]
27. Lu, X.; Fan, J.; Wu, L.; Dong, J. Forecasting multi-step ahead monthly reference evapotranspiration using hybrid extreme gradient boosting with grey wolf optimization algorithm. *Comput. Model. Eng. Sci.* **2020**, *125*, 699–723.
28. Abdullah, S.S.; Malek, M.A.; Abdullah, N.S.; Kisi, O.; Yap, K.S. Extreme Learning Machines: A new approach for prediction of reference evapotranspiration. *J. Hydrol.* **2015**, *527*, 184–195. [[CrossRef](#)]
29. Feng, Y.; Cui, N.; Zhao, L.; Hu, X.; Gong, D. Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southwest China. *J. Hydrol.* **2016**, *536*, 376–383. [[CrossRef](#)]
30. Wu, L.; Zhou, H.; Ma, X.; Fan, J.; Zhang, F. Daily reference evapotranspiration prediction based on hybridized extreme learning machine model with bio-inspired optimization algorithms: Application in contrasting climates of China. *J. Hydrol.* **2019**, *577*, 123960. [[CrossRef](#)]

31. Yin, Z.; Feng, Q.; Yang, L.; Deo, R.C.; Wen, X.; Si, J.; Xiao, S. Future projection with an extreme-learning machine and support vector regression of reference evapotranspiration in a mountainous inland watershed in north-west China. *Water* **2017**, *9*, 880. [[CrossRef](#)]
32. Ahmadi, F.; Mehdizadeh, S.; Mohammadi, B.; Pham, Q.B.; Doan, T.N.C.; Vo, N.D. Application of an artificial intelligence technique enhanced with intelligent water drops for monthly reference evapotranspiration estimation. *Agric. Water Manag.* **2021**, *244*, 106622. [[CrossRef](#)]
33. Ferreira, L.B.; da Cunha, F.F.; de Oliveira, R.A.; Fernandes Filho, E.I. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM—A new approach. *J. Hydrol.* **2019**, *572*, 556–570. [[CrossRef](#)]
34. Torres, A.F.; Walker, W.R.; McKee, M. Forecasting daily potential evapotranspiration using machine learning and limited climatic data. *Agric. Water Manag.* **2011**, *98*, 553–562. [[CrossRef](#)]
35. Gocić, M.; Motamedi, S.; Shamsirband, S.; Petković, D.; Ch, S.; Hashim, R.; Arif, M. Soft computing approaches for forecasting reference evapotranspiration. *Comput. Electron. Agric.* **2015**, *113*, 164–173. [[CrossRef](#)]
36. Karbasi, M. Forecasting of Multi-Step Ahead Reference Evapotranspiration Using Wavelet- Gaussian Process Regression Model. *Water Resour. Manag.* **2018**, *32*, 1035–1052. [[CrossRef](#)]
37. Kisi, O.; Keshtegar, B.; Zounemat-Kermani, M.; Heddam, S.; Trung, N.T. Modeling reference evapotranspiration using a novel regression-based method: Radial basis M5 model tree. *Theor. Appl. Climatol.* **2021**, *145*, 639–659. [[CrossRef](#)]
38. Mattar, M.A. Using gene expression programming in monthly reference evapotranspiration modeling: A case study in Egypt. *Agric. Water Manag.* **2018**, *198*, 28–38. [[CrossRef](#)]
39. Mattar, M.A.; Alazba, A.A. GEP and MLR approaches for the prediction of reference evapotranspiration. *Neural Comput. Appl.* **2019**, *31*, 5843–5855. [[CrossRef](#)]
40. Alazba, A.A.; Yassin, M.A.; Mattar, M.A. Modeling daily evapotranspiration in hyper-arid environment using gene expression programming. *Arab J. Geosci.* **2016**, *9*, 202. [[CrossRef](#)]
41. Yassin, M.A.; Alazba, A.A.; Mattar, M.A. Comparison Between Gene Expression Programming and Traditional Models for Estimating Evapotranspiration under Hyper Arid Conditions. *Water Resour.* **2016**, *43*, 412–427. [[CrossRef](#)]
42. Shiri, J.; Sadraddini, A.A.; Nazemi, A.H.; Kişi, Ö.; Landeras, G.; Fakheri Fard, A.; Marti, P. Generalizability of Gene Expression Programming-based approaches for estimating daily reference evapotranspiration in coastal stations of Iran. *J. Hydrol.* **2014**, *508*, 1–11. [[CrossRef](#)]
43. Shiri, J.; Kişi, Ö.; Landeras, G.; López, J.J.; Nazemi, A.H.; Stuyt, L.C.P.M. Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). *J. Hydrol.* **2012**, *414–415*, 302–316. [[CrossRef](#)]
44. Wang, S.; Lian, J.; Peng, Y.; Hu, B.; Chen, H. Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China. *Agric. Water Manag.* **2019**, *221*, 220–230. [[CrossRef](#)]
45. Wang, S.; Fu, Z.-Y.; Chen, H.-S.; Nie, Y.-P.; Wang, K.L. Modeling daily reference ET in the karst area of northwest Guangxi (China) using gene expression programming (GEP) and artificial neural network (ANN). *Theor. Appl. Climatol.* **2016**, *126*, 493–504. [[CrossRef](#)]
46. Roy, D.K.; Saha, K.K.; Kamruzzaman, M.; Biswas, S.K.; Hossain, M.A. Hierarchical fuzzy systems integrated with particle swarm optimization for daily reference evapotranspiration prediction: A novel approach. *Water Resour. Manag.* **2021**, *35*, 5383–5407. [[CrossRef](#)]
47. Yan, S.; Wu, L.; Fan, J.; Zhang, F.; Zou, Y.; Wu, Y. A novel hybrid WOA-XGB model for estimating daily reference evapotranspiration using local and external meteorological data: Applications in arid and humid regions of China. *Agric. Water Manag.* **2021**, *244*, 106594. [[CrossRef](#)]
48. Başağaoğlu, H.; Chakraborty, D.; Winterle, J. Reliable evapotranspiration predictions with a probabilistic machine learning framework. *Water* **2021**, *13*, 557. [[CrossRef](#)]
49. Fu, T.; Li, X.; Jia, R.; Feng, L. A novel integrated method based on a machine learning model for estimating evapotranspiration in dryland. *J. Hydrol.* **2021**, *603*, 126881. [[CrossRef](#)]
50. Chia, M.Y.; Huang, Y.F.; Koo, C.H. Resolving data-hungry nature of machine learning reference evapotranspiration estimating models using inter-model ensembles with various data management schemes. *Agric. Water Manag.* **2022**, *261*, 107343. [[CrossRef](#)]
51. Pasqualotto, N.; D'Urso, G.; Bolognesi, S.F.; Belfiore, O.R.; Van Wittenberghe, S.; Delegido, J.; Pezzola, A.; Winschel, C.; Moreno, J. Retrieval of evapotranspiration from sentinel-2: Comparison of vegetation indices, semi-empirical models and SNAP biophysical processor approach. *Agronomy* **2019**, *9*, 663. [[CrossRef](#)]
52. Rodrigues, G.C.; Braga, R.P. A Simple application for computing reference evapotranspiration with various levels of data availability—ETo tool. *Agronomy* **2021**, *11*, 2203. [[CrossRef](#)]
53. Rodrigues, G.C.; Braga, R.P. Estimation of daily reference evapotranspiration from NASA POWER reanalysis products in a hot summer mediterranean climate. *Agronomy* **2021**, *11*, 2077. [[CrossRef](#)]
54. Zheng, S.; Ni, K.; Ji, L.; Zhao, C.; Chai, H.; Yi, X.; He, W.; Ruan, J. Estimation of evapotranspiration and crop coefficient of rain-fed tea plants under a subtropical climate. *Agronomy* **2021**, *11*, 2332. [[CrossRef](#)]
55. 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](#)]

56. Vásquez, C.; Célleri, R.; Córdova, M.; Carrillo-Rojas, G. Improving reference evapotranspiration (ET<sub>o</sub>) calculation under limited data conditions in the high Tropical Andes. *Agric. Water Manag.* **2022**, *262*, 107439. [[CrossRef](#)]
57. Nourani, V.; Elkiran, G.; Abdullahi, J. Multi-step ahead modeling of reference evapotranspiration using a multi-model approach. *J. Hydrol.* **2020**, *581*, 124434. [[CrossRef](#)]
58. Tien Bui, D.; Hoang, N.-D.; Martínez-Álvarez, F.; Ngo, P.-T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Sci. Total Environ.* **2020**, *701*, 134413. [[CrossRef](#)]
59. Xu, H.; Zhou, J.; Asteris, P.G.; Jahed Armaghani, D.; Tahir, M.M. Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate. *Appl. Sci.* **2019**, *9*, 3715. [[CrossRef](#)]
60. Yang, H.-F.; Chen, Y.-P.P. Hybrid deep learning and empirical mode decomposition model for time series applications. *Expert Syst. Appl.* **2019**, *120*, 128–138. [[CrossRef](#)]
61. Fang, W.; Zhong, B.; Zhao, N.; Love, P.E.D.; Luo, H.; Xue, J.; Xu, S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [[CrossRef](#)]
62. Fan, L.; Zhang, T.; Zhao, X.; Wang, H.; Zheng, M. Deep topology network: A framework based on feedback adjustment learning rate for image classification. *Adv. Eng. Inform.* **2019**, *42*, 100935. [[CrossRef](#)]
63. Cummins, N.; Baird, A.; Schuller, B.W. Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning. *Methods* **2018**, *151*, 41–54. [[CrossRef](#)] [[PubMed](#)]
64. Plappert, M.; Mandery, C.; Asfour, T. Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks. *Rob. Auton. Syst.* **2018**, *109*, 13–26. [[CrossRef](#)]
65. Bowes, B.D.; Sadler, J.M.; Morsy, M.M.; Behl, M.; Goodal, J.L. Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks. *Water* **2019**, *11*, 1–38. [[CrossRef](#)]
66. Supreetha, B.S.; Shenoy, N.; Nayak, P. Lion algorithm-optimized long short-term memory network for groundwater level forecasting in Udipi District, India. *Appl. Comput. Intell. Soft Comput.* **2020**, *2020*, 8685724. [[CrossRef](#)]
67. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [[CrossRef](#)]
68. Chang, F.-J.; Chang, L.-C.; Huang, C.-W.; Kao, I.-F. Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. *J. Hydrol.* **2016**, *541*, 965–976. [[CrossRef](#)]
69. Daliakopoulos, I.N.; Coulibaly, P.; Tsanis, I.K. Groundwater level forecasting using artificial neural networks. *J. Hydrol.* **2005**, *309*, 229–240. [[CrossRef](#)]
70. Guzman, S.M.; Paz, J.O.; Tagert, M.L.M. The use of NARX neural networks to forecast daily groundwater levels. *Water Resour. Manag.* **2017**, *31*, 1591–1603. [[CrossRef](#)]
71. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw. Learn. Syst.* **1994**, *5*, 157–166. [[CrossRef](#)]
72. Fischer, T.; Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* **2018**, *270*, 654–669. [[CrossRef](#)]
73. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.Y.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [[CrossRef](#)]
74. Hu, C.; Wu, Q.; Li, H.; Jian, S.; Li, N.; Lou, Z. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water* **2018**, *10*, 1543. [[CrossRef](#)]
75. Liang, C.; Li, H.; Lei, M.; Du, Q. Dongting lake water level forecast and its relationship with the three Gorges dam based on a long short-term memory network. *Water* **2018**, *10*, 1389. [[CrossRef](#)]
76. Tian, Y.; Xu, Y.-P.; Yang, Z.; Wang, G.; Zhu, Q. Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. *Water* **2018**, *10*, 1655. [[CrossRef](#)]
77. Zhang, J.; Zhu, Y.; Zhang, X.; Ye, M.; Yang, J. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* **2018**, *561*, 918–929. [[CrossRef](#)]
78. Majhi, B.; Naidu, D.; Mishra, A.P.; Satapathy, S.C. Improved prediction of daily pan evaporation using Deep-LSTM model. *Neural Comput. Appl.* **2020**, *32*, 7823–7838. [[CrossRef](#)]
79. Hu, X.; Shi, L.; Lin, G.; Lin, L. Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation. *J. Hydrol.* **2021**, *601*, 126592. [[CrossRef](#)]
80. Ferreira, L.B.; da Cunha, F.F. New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. *Agric. Water Manag.* **2020**, *234*, 106113. [[CrossRef](#)]
81. Roy, D.K. Long short-term memory networks to predict one-step ahead reference evapotranspiration in a subtropical climatic zone. *Environ. Process.* **2021**, *8*, 911–941. [[CrossRef](#)]
82. Ferreira, L.B.; da Cunha, F.F. Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. *Comput. Electron. Agric.* **2020**, *178*, 105728. [[CrossRef](#)]
83. Ahmed, A.A.; Deo, R.C.; Feng, Q.; Ghahramani, A.; Raj, N.; Yin, Z.; Yang, L. Hybrid deep learning method for a week-ahead evapotranspiration forecasting. *Stoch. Environ. Res. Risk Assess.* **2021**, *36*, 831–849. [[CrossRef](#)]
84. Saggi, M.K.; Jain, S. Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning. *Comput. Electron. Agric.* **2019**, *156*, 387–398. [[CrossRef](#)]

85. Granata, F.; Di Nunno, F. Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. *Agric. Water Manag.* **2021**, *255*, 107040. [CrossRef]
86. Maroufpoor, S.; Bozorg-Haddad, O.; Maroufpoor, E. Reference evapotranspiration estimating based on optimal input combination and hybrid artificial intelligent model: Hybridization of artificial neural network with grey wolf optimizer algorithm. *J. Hydrol.* **2020**, *588*, 125060. [CrossRef]
87. Yu, H.; Wen, X.; Li, B.; Yang, Z.; Wu, M.; Ma, Y. Uncertainty analysis of artificial intelligence modeling daily reference evapotranspiration in the northwest end of China. *Comput. Electron. Agric.* **2020**, *176*, 105653. [CrossRef]
88. Estévez, J.; García-Marín, A.P.; Morábito, J.A.; Cavagnaro, M. Quality assurance procedures for validating meteorological input variables of reference evapotranspiration in mendoza province (Argentina). *Agric. Water Manag.* **2016**, *172*, 96–109. [CrossRef]
89. Allen, R.G.; Pereira, L.S.; Raes, D. *Evapotranspiracion Del Cultivo. Guías Para la Determinacion de Los Requerimientos de Agua de Los Cultivos (Technical Report)*; FAO: Roma, Italia, 2006.
90. Shiri, J.; Nazemi, A.H.; Sadraddini, A.A.; Landeras, G.; Kişi, O.; Fakheri Fard, A.; Marti, P. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Comput. Electron. Agric.* **2014**, *108*, 230–241. [CrossRef]
91. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
92. Yuan, X.; Chen, C.; Lei, X.; Yuan, Y.; Muhammad Adnan, R. Monthly runoff forecasting based on LSTM–ALO model. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2199–2212. [CrossRef]
93. Jang, J.-S.R.; Sun, C.T.; Mizutani, E. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*; Prentice Hall: Upper Saddle River, NJ, USA, 1997.
94. Sugeno, M.; Yasukawa, T. A fuzzy-logic-based approach to qualitative modeling. *IEEE Trans. Fuzzy Syst.* **1993**, *1*, 7. [CrossRef]
95. Takagi, T.; Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man. Cybern.* **1985**, *SMC-15*, 116–132. [CrossRef]
96. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [CrossRef]
97. *MATLAB Version R2019b*; The MathWorks, Inc.: Natick, MA, USA, 2019.
98. Jang, J.-S.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man. Cybern.* **1993**, *23*, 665–685. [CrossRef]
99. Rasmussen, C.E.; Williams, C.K. *Gaussian Process for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006.
100. Bishop, C. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
101. Quinlan, J.R. Learning with continuous classes. In Proceedings of the Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November 1992; pp. 343–348.
102. Wang, Y.; Witten, I. Induction of model trees for predicting continuous classes. *Work. Pap.* **1996**, *96*, 23.
103. Bhattacharya, B.; Solomatine, D.P. Neural networks and M5 model trees in modelling water level–discharge relationship. *Neurocomputing* **2005**, *63*, 381–396. [CrossRef]
104. Solomatine, D.P.; Dulal, K. Model trees as an alternative to neural networks in rainfall-runoff modelling. *Hydrol. Sci. J.* **2003**, *48*, 399–411. [CrossRef]
105. Solomatine, D.P.; Yunpeng, X. M5 model trees and neural networks: Application to flood forecasting in the upper reach of the Huai River in China. *J. Hydrol. Eng.* **2004**, *9*, 491–501. [CrossRef]
106. Jekabsons, G. *M5PrimeLab: M5' Regression Tree, Model Tree, and Tree Ensemble Toolbox for Matlab/Octave*; The MathWorks, Inc.: Natick, MA, USA, 2020. Available online: <http://www.cs.rtu.lv/jekabsons/regression.html> (accessed on 23 December 2021).
107. Friedman, J.H. Multivariate adaptive regression splines (with discussion). *Ann. Stat.* **1991**, *19*, 1–67.
108. Bera, P.; Prasher, S.O.; Patel, R.M.; Madani, A.; Lacroix, R.; Gaynor, J.D.; Tan, C.S.; Kim, S.H. Application of MARS in simulating pesticide concentrations in soil. *Trans. Asabe* **2006**, *49*, 297–307.
109. Salford-Systems. *SPM Users Guide: Introducing MARS*; Minitab, LLC.: State College, PA, USA, 2019. Available online: <https://www.minitab.com/content/dam/www/en/uploadedfiles/content/products/spm/IntroMARS.pdf> (accessed on 23 December 2021).
110. Roy, D.K.; Datta, B. Multivariate adaptive regression spline ensembles for management of multilayered coastal aquifers. *J. Hydrol. Eng.* **2017**, *22*, 4017031. [CrossRef]
111. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser.* **1977**, *39*, 763–768.
112. MacKay, D.J.C. The evidence framework applied to classification networks. *Neural Comput.* **1992**, *4*, 720–736. [CrossRef]
113. Chen, M. Probabilistic Linear Regression. 2021. Available online: <https://www.mathworks.com/matlabcentral/fileexchange/55832-probabilistic-linear-regression> (accessed on 23 December 2021).
114. Yu, P.-S.; Chen, S.-T.; Chang, I.-F. Support vector regression for real-time flood stage forecasting. *J. Hydrol.* **2006**, *328*, 704–716. [CrossRef]
115. Yoon, H.; Jun, S.-C.; Hyun, Y.; Bae, G.-O.; Lee, K.-K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **2011**, *396*, 128–138. [CrossRef]
116. Basak, D.; Pal, S.; Patranabis, D.C. Support vector regression. *Neural Inf. Process.* **2007**, *11*, 203–224.
117. Chevalier, R.F.; Hoogenboom, G.; McClendon, R.W.; Paz, J.A. Support vector regression with reduced training sets for air temperature prediction: A comparison with artificial neural networks. *Neural Comput. Appl.* **2011**, *20*, 151–159. [CrossRef]

118. Zhang, G.; Ge, H. Prediction of xylanase optimal temperature by support vector regression. *Electron. J. Biotechnol.* **2012**, *15*, 7. [[CrossRef](#)]
119. Wu, J.; Sun, J.; Liang, L.; Zha, Y. Determination of weights for ultimate cross efficiency using Shannon entropy. *Expert Syst. Appl.* **2011**, *38*, 5162–5165. [[CrossRef](#)]
120. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
121. Willmot, C.J. On the validation of models. *Phys. Geogr.* **1981**, *2*, 184–194. [[CrossRef](#)]
122. Legates, D.R.; McCabe Jr, G.J. Evaluating the use of “goodness-of fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [[CrossRef](#)]
123. Tao, H.; Diop, L.; Bodian, A.; Djaman, K.; Ndiaye, P.M.; Yaseen, Z.M. Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. *Agric. Water Manag.* **2018**, *208*, 140–151. [[CrossRef](#)]
124. Chia, M.Y.; Huang, Y.F.; Koo, C.H. Swarm-based optimization as stochastic training strategy for estimation of reference evapotranspiration using extreme learning machine. *Agric. Water Manag.* **2021**, *243*, 106447. [[CrossRef](#)]
125. Mohammadi, B.; Mehdizadeh, S. Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. *Agric. Water Manag.* **2020**, *237*, 106145. [[CrossRef](#)]
126. Elbeltagi, A.; Deng, J.; Wang, K.; Malik, A.; Maroufpoor, S. Modeling long-term dynamics of crop evapotranspiration using deep learning in a semi-arid environment. *Agric. Water Manag.* **2020**, *241*, 106334. [[CrossRef](#)]
127. Gao, L.; Gong, D.; Cui, N.; Lv, M.; Feng, Y. Evaluation of bio-inspired optimization algorithms hybrid with artificial neural network for reference crop evapotranspiration estimation. *Comput. Electron. Agric.* **2021**, *190*, 106466. [[CrossRef](#)]
128. Yin, J.; Deng, Z.; Ines, A.V.; 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](#)]
129. Heinemann, A.B.; Oort, P.A.V.; Fernandes, D.S.; Maia, A. Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. *Bragantia* **2012**, *71*, 572–582. [[CrossRef](#)]
130. Li, M.-F.; Tang, X.-P.; Wu, W.; Liu, H.-B. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Convers. Manag.* **2013**, *70*, 139–148. [[CrossRef](#)]