*Knowledge-based Engineering and Sciences* (2023), **4**, (1), 37–54
doi:10.51526/kbes.2023.4.1.37-54

RESEARCH PAPER

# Dynamic Sliding Window-Based Long Short-Term Memory Model Development for Pan Evaporation Forecasting

Aayush Bhattarai,[1] Deeba Qadir,[2] Aliyu Muhammad Sunusi,[3] Birhan Getachew,[4] and Abdul Rahman Mallah[*,5]

[1]Department of Mechanical and Aerospace Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, 44700, Nepal
[2]Building and Architectural Engineering at Politecnico di Milano, Milan, Lombardy, Italy
[3]Department of Architecture, Khon kaen University Thailand
[4]Department of Geography and Environmental Studies, Debre Tabor University, Debre Tabor, Ethiopia
[5]Department of Engineering, Reykjavik University, Menntavegur 1, Reykjavík, 102, Iceland
[*]Corresponding author. Email: AR_86_M@hotmail.com

(Received 27 February 2023; revised 11 April 2023; accepted 12 April 2023; first published online 30 April 2023)

**Abstract**

Accurate and precise forecasting for pan evaporation (EPm) is one of the essential processes for multiple purposes of watershed decision making and sustainability. This is because EPm is highly influenced by climate change and thus it is a highly stochastic weather process. In this research long short-term memory (LSTM) and Gaussian process regression (GPR) models were used to forecast monthly EPm data. The EPm data series were used to cover the period (1980 to 2020) and belong to two meteorological stations located in the United States of America. One station has a tropical climate in Hialeah, Florida (station 51) and the other has a Mediterranean climate in Markley Cove, California (station 9). The analysis was carried out using the optimum window sizes for the LSTM and GPR methods, which were 4 and 7, respectively. The modeling results indicated that the LSTM model reported maximum correlation value (R = 0.965) and (R = 0.701) for Station 9 and 51 respectively whereas minimum root mean square error (RMSE = 0.611) and (RMSE = 0.916) for station 9 and 51 respectively which outperformed the GPR model. Based on the relative error diagram, 78% of the LSTM model's results ranged between ±25%. In addition, the LSTM model's normal distribution of errors showed that the mean error and standard deviations for station 9 are -0.0104 and 0.614 whereas for station 51 it was -0.026 and 0.921. For the same stations, the GPR model mean error and standard deviation for station 9 are -0.016 and 0.753 whereas for station 51 were -0.026 and 0.921 indicating that the LSTM model outperforms the GPR model in forecasting monthly EPm. In general, the development of the LSTM model as a dynamic soft computing model reported a robust and reliable tool for EPm forecasting at different climate characteristics.

**Keywords:** Long short-term memory; Gaussian process regression; evaporation forecasting; Hialeah station; Markley Cove station

## 1. Introduction

### *1.1 Research background*

A type of natural process that is a part of the hydrological cycle that influences agriculture productivity and water availability for human purposes is evaporation [1]. Various researchers have identified EPm

as one of the complex elements in the water cycle process [2]–[4]. Despite this fact, accurate and reliable methodologies for forecasting EPm rates are critical for effective water resource management and reliably describing the water balance in natural and artificial water bodies [5]. EPm can be measured and predicted in a variety of ways. The majority of researchers have conducted reviews on these methods [6]–[8]. Over the literature, a plethora of authors uses empirical methods to predict evaporation rate [9]. Furthermore, these models are widely used for predicting and calculating EPm in various geographical locations [10]–[12]. Besides most of these methods require a huge meteorological data as input. Access to meteorological data in all regions may not be possible therefore it makes these methods hard to apply [13]. On the contrary, due to the complexity, non-linearity and unsteady nature of evaporation providing a comprehensive and straightforward formula that involves all the physical processes is difficult [14]. As a result, these problems highlight the need for more adaptable and easily-operated evaporation prediction methods. Several researchers have reported on the determination of EPm using empirical and semi-empirical formulas based on various meteorological data [15]–[17]. Nonetheless, due to the use of various climatic factors, these models possess limitations [4]. As a result, different approaches for forecasting EPm that require less meteorological data are required [18].

### 1.2 Literature review

Data mining approaches that do not need prior knowledge of evaporation's fundamental physical processes or a large amount of data have recently received much attention and they are most certainly effective in real-time decision support systems when made the comparison to conventional parametric methods [4], [19], [20]. Artificial Intelligence (AI) techniques have been widely used in water resource fields over the last few decades, and it is well known for predicting evaporation as well [21]–[24]. The main advantage of the AI techniques is being completely non-parametric, requiring no prior knowledge of the relationships amidst input variables and output data [25], [26]. In addition to this, various AI models and methods were used for forecasting monthly streamflow [5], [27], [28]. Machine learning (ML) model applications for evaporation process modeling have received much attention by several researchers. In addition, they have shown remarkable successful implications within ecological and water engineering issues in recent years [29]–[33]. AI systems are simpler, more robust and capable of dealing with complex non-linear procedures with ease [34].

Even though AI-based models have good prediction performance, the majority of these studies employed several traditional AI algorithms that lacked a data prepossessing scheme optimizer algorithm [4]. Furthermore, the effects of parameter interaction on model performance are not adequately substantiated by current methods. Due to the previously mentioned concern, there is a growing trend toward using evolutionary algorithms to adjust the internal parameters of AI models and as a result, various scholars have attempted to solve the ANN training problem using a mixture of various model optimization techniques. Some researchers combined advanced neural methods with the firefly algorithm to obtain more precise evaporation measurements. Since then, the firefly algorithm has been widely used to train ANN models in a variety of optimization problems [35].

### 1.3 Research motivation and contribution

Deep learning (DL) has received a lot of attention in recent years, thanks to neural networks. As with hydrological modeling, the effectiveness of DL methods is greatly aided by advances in computer technology, especially in the area of Graphic Processing Units (GPU) [36]; and ready accessibility of large data sets [36]–[38]. Few DL algorithms are in the computer vision areas such as [39], speech recognition [40], and Natural Language Processing (NLP) [41], however, there have been few attempts of DL algorithms to hydrological problems. Many deep learning algorithms, such as recurrent neural networks (RNN) and convolutional neural networks (CNN), have proven themselves in addressing real-world issues [42]–[45]. Since it has loops that allow information from

the previous time slice to be transmitted to the following time slice, the RNN has an excellent ability to learn especially for time series Data. The RNN, on the other hand, has an issue with the gradient disappearing and exploding, and this defect causes the RNN's ability to learn long-distance information to decrease over time. [46]. To compensate for this problem, the LSTM network, which is able to learn long-term dependencies in data series was established for long-sequence data learning and solves the problem of vanishing gradients [47]. LSTM approaches have lately become popular amongst researchers in the hydrology and water management field. The robustness of various neural network topologies for simulating and predicting water levels of the combined structure was explored by [48], they found that LSTM (together with a cell memory recurrent neural network design) is preferable for multi-step-ahead predictions when compared with conventional architectures that lack explicit cell memory. LSTM was used by [49] to predict water tables in agricultural areas. Research analyzed the effectiveness of AI models such as LSTM for flow prediction of the river basin in India when compared to the conventional and naive method [50]. To predict the discharge at the Hoa Binn Station in Vietnam a deep LSTM model was used [41]. In their study, in addition to evaluating the model's ability to predict flood flows and flood peaks, they look into the impact of dataset features on model performance. Although there are many studies, to our knowledge no prior studies have examined the use of long-short term methods for forecasting EPm. In the current study Gaussian Processes (GPs) are used to construct the model and the LSTM approach is applied to forecast monthly EPm data for two sites in the United States of America, one is Hialeah in Florida state with a tropical climate, and the other one Markley cove in California with a Mediterranean climate.

Based on the reported literature, there was a substantial amount of research was adopted on the utilization of the LSTM model in solving hydrological processes [50]–[52]. Hence, the inspiration of the current research was initiated. LSTM model was developed to forecast the EPm process at two different climate characteristics in the USA. The feasibility of the LSTM was compared with the GPR model. As a matter of fact, the identification of the appropriate correlated lag time series is one of the popular obstacles in hydrology [53]. Hence, in the current research, a dynamic LSTM model was developed to identify the proper and essential lag time for the development one month ahead EPm process. As a matter of fact, the identification of the appropriate correlated lag time series is one of the popular obstacles in hydrology. Hence, in the current research, a dynamic LSTM model was developed to identify the proper and essential lag time for the development one month ahead EPm. The dynamic RNN (LSTM) takes advantage of the vanishing gradient issue to increase speed in a way that conventional LSTM does not [54]. Variable sequence length is possible with dynamic LSTM unlike static LSTM, which anticipates running for the entire length of the LSTM, dynamic LSTM expects to run for the entire length of the LSTM. The modeling process is conducted using a univariate modeling procedure where only historical data of evaporation are used to build the forecasting model. This is giving the potential to use minimal information for the modeling process. This gives credit to the catchments where there is a serious limitation in the data measurement.

### *1.4 Research objectives*

The research objectives of the current research are: i. Develop a relatively new machine learning model called the LSTM model for EPm forecasting.

ii. Validate the proposed model with the GPR model and assess both models statistically and graphically.

iii. Investigate the capacity of the proposed model for two different climate characteristics at the USA.

iv. Perform univariate modeling procedure where only memorial historical EPm data are used for the modeling development.

v. The input features for the forecasting matrix "lag times" are identified using the feasibility of

the LSTM model.

## 2. Materials and Methods

The schematic of the Dynamic LSTM is shown in Figure 1. The data set obtained has been processed as per the given flow chart diagram.

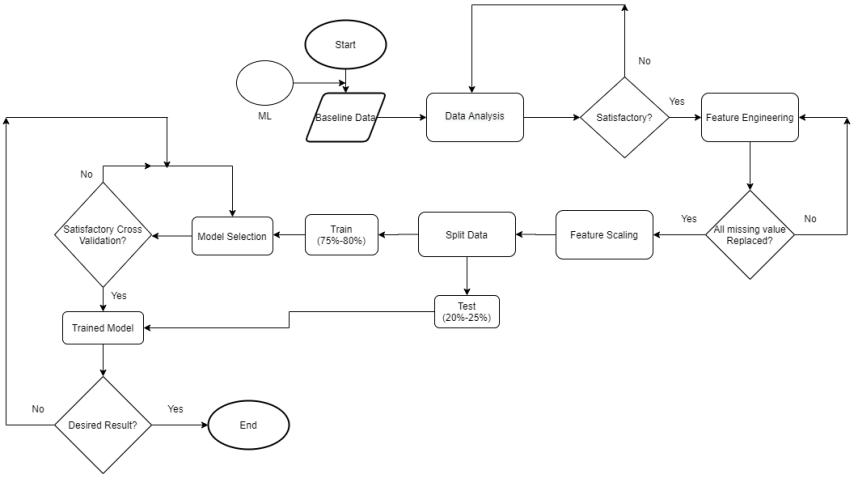


**Figure 1:** *The schematic framework of the proposed methodology.*

### 2.1 LSTM

Deep learning (DL) is well known among ML algorithms as it confers the ability to think and see like humans. DL is a group of frameworks whose main objective is to represent abstract ideas in data sets. A deep graph made up of a range of transformations (linear & nonlinear) and several processing layers are used to implement deep learning models. Long-term dependencies can be recognized using the LSTM, a particular kind of RNN [47]. This model has the ability to generate errors without declining gradients by introducing the cell modes with static errors. In contrast to other RNNs, the LSTM comprises four distinct layers that co-interact by a predetermined topology. The parameter and computational costs of the LSTM are 4x those of the RNN. LSTM has a network for the computation of the memory input. The input or update gate is responsible for determining the relevance of the received data from the current moment for storage in long-term memory, while the job of the output gate is to just forward the needed data to the output; the role of the forget gate is determining the level of available long in the memory cells, and or wiping out (forget) unnecessary previous information [55], [56]. The memory of present and prior data is stored and passed through the hidden layer activation; this allows subsequent activations to have the patterns and sequences of the earlier data [57].

In ML algorithms, the data set is mostly divided into training and testing sets. The parameters that must be determined in these ML algorithms include the number of hidden layers, the batch size, the maximum epoch, the learning rate, and the number of neurons. These parameters could be chosen algorithmically or at random. The selection and calculation of these functions in the gate are based on the maximum epoch and expected error level [55], [56], [58], [59].

Figure 2 depicts the LSTM design, where X represents the time series of C-channels and S-length; Ct is the state of the cell at time-step t, while ht represents the hidden/output state [60]. The first LSTM is utilized for the series' first step and the network's early state to calculate the first input &

output states. The input cell state and output (ct) in this block is calculated at time t using the current network state (ct-1, ht-1) and the following time step in the sequence.

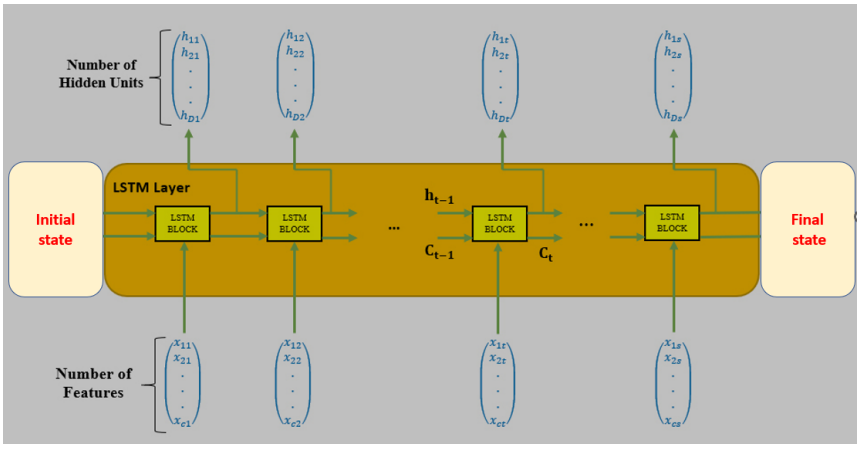


**Figure 2:** *LSTM layer architecture [60].*

The cell state and the hidden (output) state make up the state of the layer. The layer is in charge of changing the information in the cell state at every given time step. These upgrades are managed by the gates. Several LSTM architecture components are used to regulate hidden and cell state. Input or update gate (i) controls the cell state update level, while the level of cell state added to the output state (o) is controlled by the output gate (i). Cell candidate (g) exasperates the cell state with informational knowledge, while the forget gate checks the cell state level reset (f).

Figure 3 displays the data as it was at time-step t. This figure graphically illustrates the processes involved in the output, update, and forget of the hidden and cell states by the gates. For an LSTM layer, the learnable weights include recurrent weights R, input weights W, and bias b. R and W matrices are insertions of recurrent weights and input weights, respectively. b matric is the bias of any part.

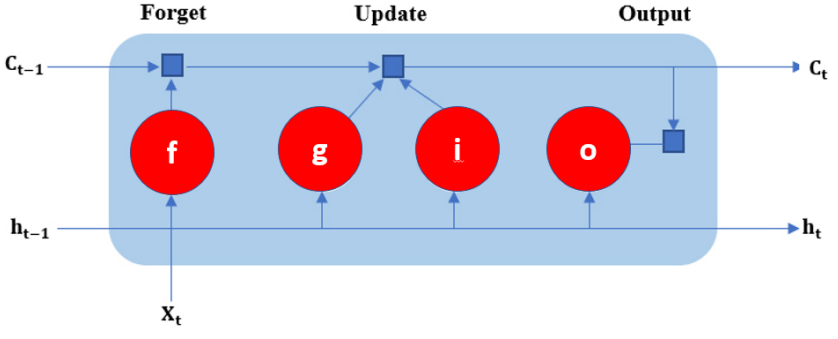


**Figure 3:** *The procedure of the data flow at time step t [58].*

These three matrices can be represented as follows:

$$\mathrm{W} = \begin{bmatrix} Wi \\ Wf \\ Wg \\ Wo \end{bmatrix}$$
$$\mathrm{R} = \begin{bmatrix} Ri \\ Rf \\ Rg \\ Ro \end{bmatrix} \quad (1)$$
$$\mathrm{b} = \begin{bmatrix} bi \\ bf \\ bg \\ bo \end{bmatrix}$$

where $o$ = output gate, $g$ = cell candidate, $f$ = forget gate, and $I$ = input gate. The cell state (Ct) and hidden state (ht) at time step t can be represented thus:

$$C_t = f_t \odot C_{t-1} + i_t \odot g_t \quad (2)$$

$$h_t = o_t \odot \sigma_c (C_t) \quad (3)$$

where  represents the multiplication of the element-wise vector (Hadamard product), and c depicts the state activation function. The hyperbolic tangent function is employed by default in the model to represent the state activation function. The LSTM components at time step t can be represented as [58]:

$$i_t = \sigma_g (W_i x_t + R_i h_{t-1} + b_i) \quad (4)$$

$$f_g = \sigma_g \left( W_f x_t + R_f h_{t-1} + b_f \right) \quad (5)$$

$$g_t = c_{\mathrm{H}} \left( W_g x_t + R_g h_{t-1} + b_g \right) \quad (6)$$

$$o_t = \sigma_g (W_o x_t + R_o h_{t-1} + b_o) \quad (7)$$

where $g$ is the gate activation function. For the determination of the gate activation function in LSTM, the sigmoid function is used as follows:

$$\sigma(\mathrm{x}) = (1 + e^{-x})^{-1} \quad (8)$$

### 2.2 Gaussian Process Regression (GPR)

Gaussian Process Regression (GPR), which employs the idea of a comprehensive Bayesian learning process, is used to tackle supervised regression & classification problems [61]. The GPR model has several advantages over other ML methods, including ease of use, hyperparameter prediction, effectiveness on small data, self-adaptive to enhance parameter prediction, and recognition for potential values rather than absolute values for each parameter defined inside the function in the probability distribution. The output value of y can be represented as:

$$y = f(X(k)) + \varepsilon \quad (9)$$

where X is the input variable measurement,  is the noise with a Gaussian distribution, and f stands for the function that needs to be modeled. A Gaussian process with a mean of m(x) and a covariance function of cov(x,x') determines the earlier probability of the space of functions [62].

$$f(x) \sim GP\left(m(x), \operatorname{cov}\left(x, x'\right)\right) \tag{10}$$

For a specific example of predictor variables $x_*$, the mean and variance $p(y|X, y, x)$ of the prediction probability distribution can be represented as:

$$\hat{y}_* = m\left(x_*\right) + k_*^T \left(K + \sigma_n^2 I\right)^{-1} \left(y - m\left(x_*\right)\right) \tag{11}$$

$$\sigma_{y_*}^2 = \mathrm{k}_* + \sigma_n^2 - \mathrm{k}_*^T \left(\mathrm{K} + \sigma_n^2 I\right)^{-1} \mathrm{k}_* \tag{12}$$

where K represents a covariance matrix with elements $[\mathrm{K}]_{i,j} = \operatorname{cov}\left(x_i, x_j\right)$ ,$\mathrm{k}_*$ is a vector that can be introduced by $[k_*]_i = \operatorname{cov}\left(x_i, x_*\right)$ ,$\mathrm{k}_* = \operatorname{cov}\left(x_i, x_*\right)$ ,and I*represent the identity matrix*.

GPR, unlike conventional regression techniques, uses the training datasets X, y to decide the model's output. The selection of the hyperparameter values is based on the maximization of the log-likelihood function of the training dataset [63]:

$$\log p(y \mid X) = -\frac{1}{2} y^T \left(\mathrm{K} + \sigma_n^2 I\right)^{-1} \mathrm{y} - \frac{1}{2} \log\left(\left|\mathrm{K} + \sigma_n^2 I\right|\right) - \frac{n}{2} \log(2\pi) \tag{13}$$

### 2.3 Study area and used data

The monthly EPm data from two meteorological stations in the United States of America are used. The location of these two stations and the geographical coordinates are shown in Figure 4. One station has a tropical climate in Hialeah, Florida (station 51), and the other has a Mediterranean climate in Markley Cove, California (station 9). The range of the used data is thirty years ranging from 1980–2020.


**Figure 4:** *The location of meteorological stations used in this study.*

### 2.4 Performance metrics

Different metrics were used to evaluate the efficiency and effectiveness of the model [35]. In equations (14–16), the p(i), o(i) and p are forecasted, observed and mean forecasted values, respectively. The applied forecasting models were evaluated statistically using the mean absolute error, root mean square error and determination coefficient [64]. The mathematical expression is presented as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|(p(i) - o(i))| \tag{14}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|(p(i) - o(i))|} \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{i=N}(p(i) - o(i))^2}{\sum_{i=1}^{i=N}(p(i) - \bar{p})^2} \tag{16}$$

Various metrics are used to evaluate the proposed and comparable forecasting models. The minimal value of RMSE is an indication of high prediction accuracy. Whereas, the higher value of correlation exhibited the better model's results. Besides the statistical metrics presented in the equations. 14–16, the Taylor diagram [65] is also utilized to evaluate models' accuracy. Taylor diagram was utilized to visualize models forecasting performance using three metrics including RMSE, correlation and standard deviation [35].

## 3. Application results and analysis

The main goal of the current research establishment is the motivation of using an advanced computer aid model called based on deep learning to investigate the forecastability performance against the GPR model. This research was conducted based on numerical statistic evaluation and the capacity of the models to simulate the EPm in two different chitchatted climate coordinates. The statistical properties of the investigated dataset were reported in Table 1 and based on the presentation of the numeric, both stations comprise total records of 492 monthly scale data. In comparison with the adopted literature review, this data span is sufficient to construct such an ML model and it is adequate to mimic the actual trend of the EPm dataset. The maximum and minimum records of EPm at MARKLEY COVE (station 9) and HIALEAH (station 51) are (10.119 & 0.237) and (11.979 & 3.408), respectively. Apparently, the stochasticity of the dataset at the MARKLEY COVE station is more complex as the variation between the maximum and minimum records is very high.

**Table 1:** *Monthly EPm (mm) statistics characteristics of the study stations, USA.*

| Stations | Dataset | No.records | Mean | Maximum | Minimum | Variance | Skewness |
|---|---|---|---|---|---|---|---|
| No. 9 | Training | 400 | 3.875 | 10.119 | 0.237 | 6.952 | 0.348 |
| USC00045360 | Testing | 92 | 3.786 | 7.824 | 0.327 | 5.473 | 0.105 |
| MARKLEY COVE, CA, USA | All data | 492 | 3.843 | 10.119 | 0.237 | 6.68 | 0.331 |
| No.51 | Training | 400 | 6.11 | 11.979 | 3.408 | 1.919 | 0.34 |
| USC00083909 | Testing | 92 | 6.05 | 8.881 | 3.744 | 1.661 | -0.001 |
| HIALEAH, FL, USA | All data | 492 | 6.113 | 11.979 | 3.408 | 1.865 | 0.279 |

Based on the sliding window initiated for identifying the appropriate lags for both stations 9 and 51, the efficient lag time for the EPm forecasting was four months lead time for station 9 and station 51, the efficient lag time for the EPm forecasting was seven months lead times.

### *3.1 Station 9 modeling accuracy*

Figure 5 was presented the time series of station 9 over the period (1980–2020), it can be observed that the dataset within the recent years trended to decrease the EPm with a range of limitation (1–8 mm), whereas in the 1980s the EPm was ranged between (1–10 mm). This could be due to climate change or some other human interactions such as deforestation. It is also clearly can be understood from the heat map of the time series data that the highest rate of EPm was experienced between May–September. For the modeling development of the forecasting matrix, the LSTM model used an auto generated technique for identifying the essential features "highly correlated lags" based on the minimal error. The efficient lag time for the EPm forecasting was four months lead time.

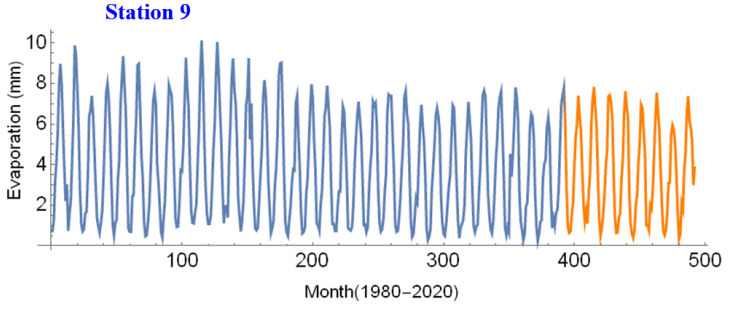


**Figure 5:** *Station 9: The time series dataset used for the modeling development over the training and validation.*

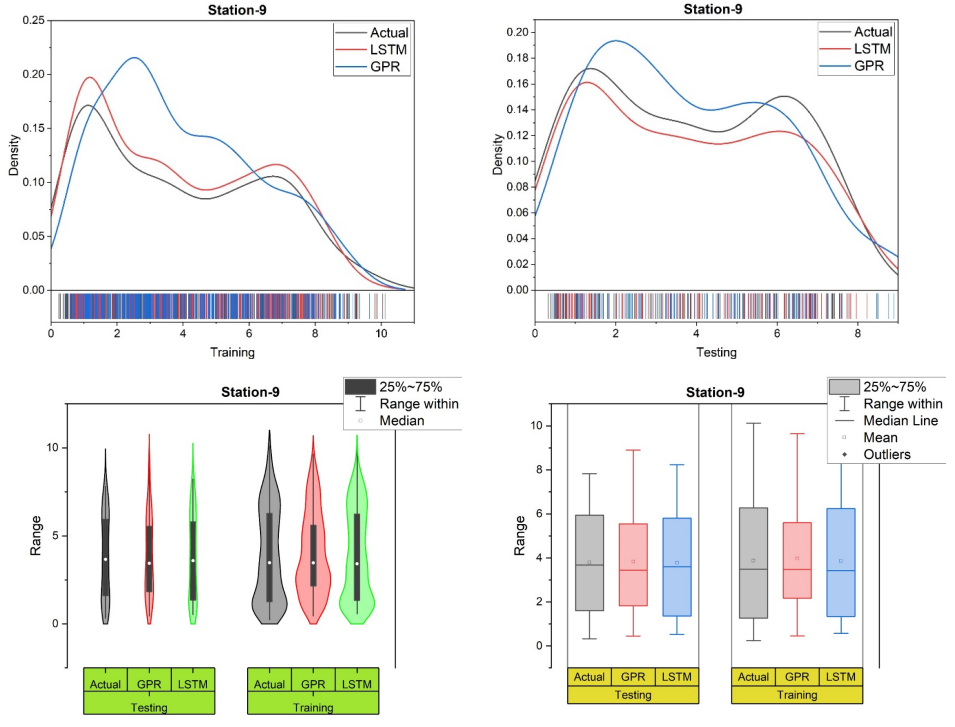


**Figure 6:** *Station 9: a) LSTM modeling process (Training and Testing), b) Relative error metric for the LSTM model, c) Forecasting performance for the LSTM.*

Graphically, Figure 6 reveals the forecast modeling accuracy of the developed LSTM model and the GPR "benchmark model" for station 9. For a better understanding of the LSTM model running process, different types of graphical evaluations were generated including density, villion, boxplot, scatter plot and Taylor diagram for both the training and testing phases. The density graph displays that the LSTM model attained a closer match with the actual observations of EPm. The villion and boxplot present almost alike visualization for the range of the error values. In general, the graphics indicated an acceptable forecasting capability of the LSTM model with determination coefficient magnitude (R = 0.95), mean absolute percentage error (MAPE= 19.3%) and mean absolute error (MAE = 0.446). Time series actual and forecasted values were mostly identical with marginal error and a single observation of overforecasted at the highest rate of EPm. Scatterplot shows an acceptable match between the actual and predicted values except some observations are diverted from the identical 45o line. This best can be explained due to the shortcoming of the effective lag times needed for the modeling simulation, or some dramatic changes in the climate of this particular month need extra climatic information to comprehend the actual mechanism of the EPm. The root mean square of the LSTM model is lower for station-9 and so, the LSTM model has better fitting as depicted in the Taylor diagram (Figure 6). Moreover, the LSTM model has a significantly high Pearson correlation coefficient greater than 0.95. Hence, the LSTM model is strongly correlated to the measured data. On the other hand, the GPR model still has a high Pearson correlation coefficient that exceeds 0.9.

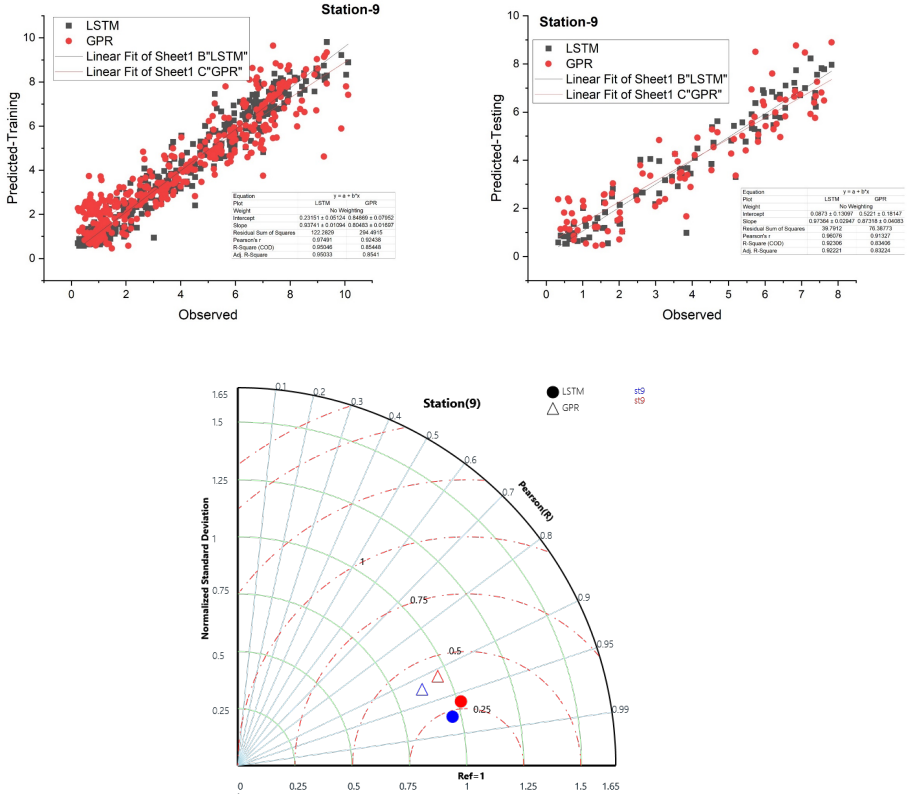


**Figure 7:** *Station 9: a) Relative error metric for the GPR model, b) Forecasting performance for the GPR, c) Taylor diagram presentation.*

Additionally, the GPR model has a slightly lower normalized standard deviation which indicates a closer distribution of the predicted points by the GPR model around the mean compared to those for the LSTM model.

The modeling evaluation of the benchmark model (GPR) at Station 9 was presented in Figures 7 in parallel with the LSTM. The results are in general compared with the proposed LSTM model. This can be revealed the limitation of the GPR model to professionally mimic the EPm stochasticity. The scatter plot indicates a lower determination coefficient magnitude ($R^2$ = 0.85 for the training phase and 0.83 for the testing phase), with much higher value of MAPE = 45.3% and MAE = 0.729 over the training phase and MAPE = 45.8% and MAE = 0.78for the testing phase. Based on the coordination map of the Taylor diagram, the LSTM model has reported closer coordination to the observed EPm data in comparison with the GPR model.

**Table 2:** *Performance indicators for Station (9)*

| Algorithm | Training Phase | | | | | |
|---|---|---|---|---|---|---|
| | $R^2$ | KGE | MAPE | MAE | RMSD | NSE |
| LSTM | 0.95 | 0.954 | 19.30% | 0.446 | 0.587 | 0.95 |
| GPR | 0.854 | 0.848 | 45.30% | 0.729 | 1.019 | 0.85 |
| Algorithm | Testing Phase | | | | | |
| | $R^2$ | KGE | MAPE | MAE | RMSD | NSE |
| LSTM | 0.923 | 0.958 | 20.90% | 0.493 | 0.657 | 0.92 |
| GPR | 0.834 | 0.902 | 45.80% | 0.78 | 0.954 | 0.832 |

### 3.2 Station 51 modeling accuracy

The time series dataset of station 51 was presented in Figure 8 for the period (1980–2020). In comparison with station 9 and station 51, the dataset was reported a totally different pattern in which the data ranges between (4–8 mm); however, several jumps in the EPm exceed 10 mm. The jumps' values are distributed all along the studied period and did not follow particular seasonality or distinguished periodical phenomena. For station 51, the efficient lag time for the EPm forecasting was seven months lead times. However, eight- and nine months lead times reveal almost similar potential. Nevertheless, the modeling of station 51 was adopted based on seven leg times as forecasting matrix predictors.

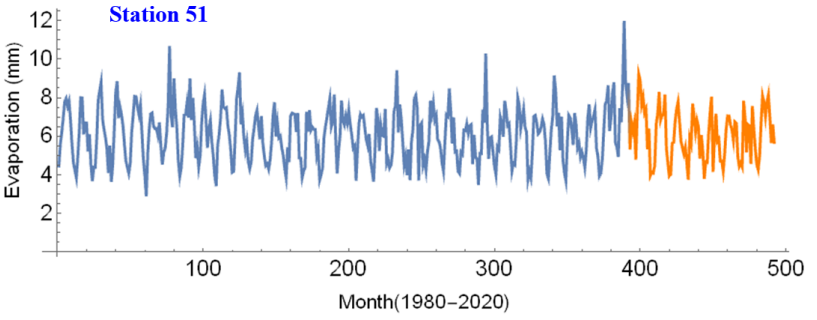


**Figure 8:** *The time series dataset used for the modeling development over the training and validation.*

The forecasting modeling accuracy of the developed LSTM model and the GPR "benchmark model" at station 51, was stated in Figure 9. For this station, both the training and testing phases'

results are considered for the evaluation of the results. The graphical presentation was demonstrated the relative error percentage of the LSTM model over the testing phase relatively enhanced in comparison with the GPR model (Table 2). The graphics indicated that $R^2$ = 0.58 and 0.50 for LSTM and GPR models, respectively over the training phase. The results can be observed much worse than those in station 9. This is clearly can be explained due to the nature of the evaporation at this station, which is associated with much complexity and non-stationarity. The time series of the actual and forecasted values were not in harmony perfectly and an obvious mismatching at certain observations where the associated sudden records of the EPm were incorporated. The GPR model at station 51 attains substantial forecasting accuracy. However, this is just one single evaluation metric and hence other evaluation metrics should be tested on the modeling performance. The scatter plot indicates a much lower correlation magnitude. Time series actual and forecasted values were implied as almost totally mismatching. Based on the multiple metrics graphical presentation (correlation, RMSE and standard deviation) in the form of a Taylor diagram, the LSTM model was described as closer coordination based on the three metrics to the observed EPm data in comparison with the GPR model. The root mean square of the LSTM model is slightly lower for station-51 and so, the LSTM and GPR models have similar fitting characteristics as depicted in the Taylor diagram (Figure 10). Moreover, the LSTM model has a higher Pearson correlation coefficient greater than 0.75, whereas it is 0.7 to 0.75 for the GPR model. Hence, the LSTM model is better correlated to the measured data compared to the GPR (Table 3). In contrast, the GPR model has a significantly lower normalized standard deviation below 0.75 which indicates a closer distribution of the predicted points by the GPR model around the mean compared to those for the LSTM model.

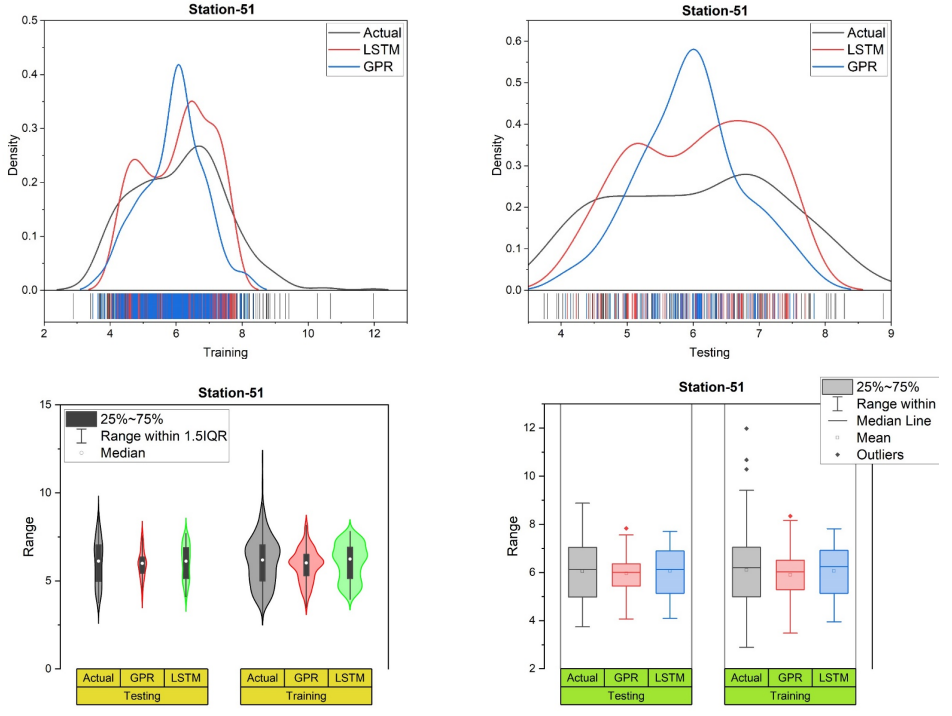


**Figure 9:** *Station 51: a) LSTM modeling process (Training and Testing, b) Relative error metric for the LSTM model, c) Forecasting performance for the LSTM*

These models can be used in planning the water budget of river basins and reservoirs where

other models fail due to the linear nature and ineffectiveness of reducing errors between the actual and forecasted value. These accurate ML techniques can be used to plan and implement effective irrigation policies. The primary limitation of this research is that the forecasting accuracy is very much affected by the predictors, the research did not consider a different combination of input parameters for models.

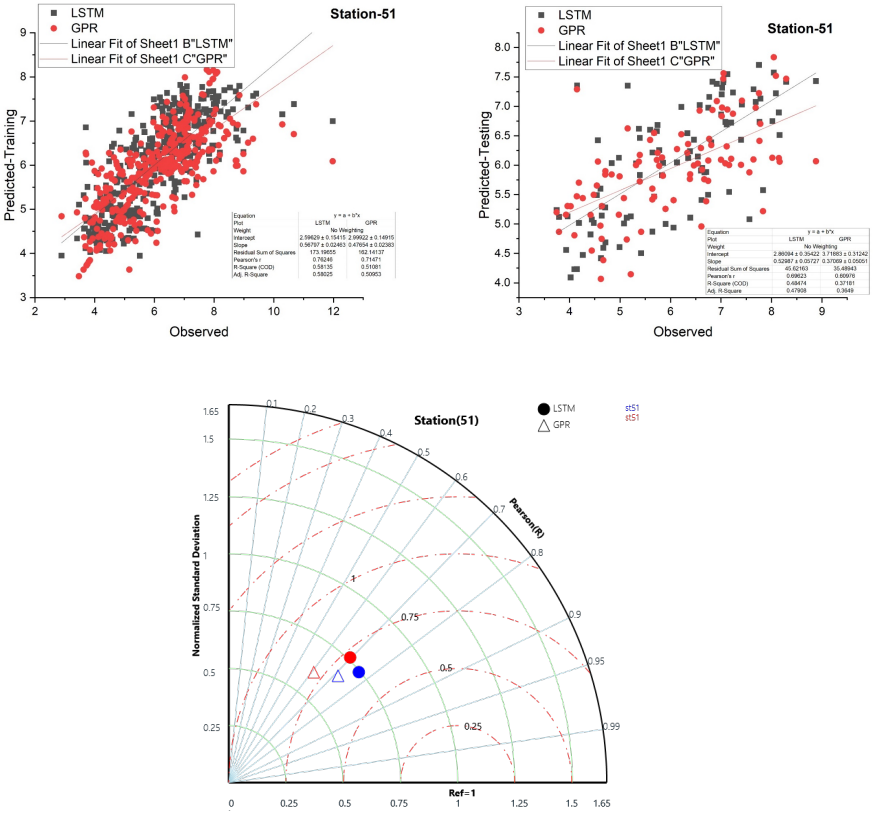


**Figure 10:** *Station 51: a) Relative error metric for the GPR model, b) Forecasting performance for the GPR, c) Taylor diagram presentation.*

## 4. Validation with the literature

Various ML models such as MLP, ANN, Adaptive Neuro-Fuzzy Interference Systems (ANFIS), Graph Neural Network (GNN), SVM, ELM and many more have been used in forecasting of EPm. One such study was done that compared Functional Link Artificial Neural Network (FLANN) with Multi-Layer Artificial Neural Network (MLANN) which used the aforementioned models at India in Chattisgarh State [66]. The FLANN attained minimum RMSE and MAE of 0.85-1.27 and 0.63-0.95 respectively when compared to MLANN with RMSE of 0.94-1.58 and MAE of 0.73-1.14. However, a daily EPm forecast was also carried out in a Deep LSTM where its comparison was made with empirical models with an average $R^2$ of 0.79 for three climatic regions in Chattisgarh, India [67]. Furthermore, a hybrid model of Slap Swarm Algorithm (SSA) and Kernel-Based Non Linear Arps Decline (KNEA) outperformed the M5 model tree (M5) and Multivariate Adaptive Regression Splines (MARS) in research conducted in north China [68]. The RMSE and MAE of the hybrid

**Table 3:** *Performance indicators for Station (51)*

| Algorithm | Training Phase | | | | | |
|---|---|---|---|---|---|---|
| | $R^2$ | KGE | MAPE | MAE | RMSD | NSE |
| LSTM | 0.581 | 0.651 | 11.40% | 0.676 | 0.902 | 0.58 |
| GPR | 0.511 | 0.56 | 11.40% | 0.691 | 0.995 | 0.489 |
| Algorithm | Testing Phase | | | | | |
| | $R^2$ | KGE | MAPE | MAE | RMSD | NSE |
| LSTM | 0.485 | 0.614 | 12.40% | 0.714 | 0.924 | 0.48 |
| GPR | 0.372 | 0.447 | 14.20% | 0.81 | 1.02 | 0.367 |

model were 1.067 and 0.867. Three ML models such as M5, Random Forests (RF) and Gradient Boosting Decision Tree (GBDT) were compared for EPm prediction in local and cross-station data in China with various empirical models [69]. With an RMSE of 0.86 and $R^2$= 0.73, the GBDT excelled over all other models. Even though, it is seen that hybrid models such as those mentioned above give an effective prediction of EPm since the input variables are much less when compared to stand alone models. This is shown in research where Support Vector Regression (SVR) a standalone model is compared with SVR combined with Krill Herd Algorithm (SVR-KHA) for which R2 were 0.71 and 0.66 respectively [70]. Similarly, three ML models Extreme Gradient Boosting (EGB), Elastic Net Linear Regression and LSTM were used to forecast the monthly EPm rate in Malaysia in which LSTM showed the most precise prediction of EPm with R2 of 0.97 and 0.98 for Alor Setar and Kota Bharu respectively [71].

## 5. Conclusion and possible future research

The objective of this research was to investigate the feasibility of an over popular version Deep-learning method called (i.e., LSTM model) to forecast monthly EPm and compare its validity to the GPR model. The data set is based on two main stations in the United States with varying climates and coordinates. Station 9 in Markley Cove, California, has a Mediterranean climate, while Station 51 in Hialeah, Florida, has a tropical climate. The LSTM model used for modeling EPm data is evaluated using statistical metrics and graphical presentation. The results of the simulated data with both the LSTM and GPR methods show that the LSTM method outperforms the GPR method in simulating EPm data. Checking the evaluating indicators such as RMSE and $R^2$ in both stations 9 and 51 demonstrates this. The results indicate the limitations of the GPR method in simulating the stochasticity of EPm. In addition, based on the Taylor diagram's coordination map, the LSTM model reported closer coordination to the observed EPm data than the GPR model. For possible future research, only data from one study area was used in this study, and further work is recommended with additional data from two or more stations to support these conclusions. Since AI models perform differently depending on the properties of the simulated matrix in different situations so the selection of input variable is one of the important factors, further study can be done using different combinations of climatological variables as inputs. These variables can be configured using various techniques such as Gamma Test.

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

## References

[1] J. L. Monteith, "Evaporation and environment," in Symposia of the society for experimental biology, 1965, vol. 19, pp. 205–234.

[2] M. F. Fingas, "A literature review of the physics and predictive modelling of oil spill evaporation," Journal of hazardous materials, vol. 42, no. 2, pp. 157–175, 1995.
[3] P. S. Shirgure and G. S. Rajput, "Evaporation modeling with neural networks – A Research review," International Journal of Research and Reviews in Soft and Intelligent Computing (IJRRSIC), 2011.
[4] W. Jing et al., "Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions," Engineering Applications of Computational Fluid Mechanics, vol. 13, no. 1, pp. 811–823, 2019.
[5] A. Ashrafzadeh, M. A. Ghorbani, S. M. Biazar, and Z. M. Yaseen, "Evaporation process modelling over northern Iran: application of an integrative data-intelligence model with the krill herd optimization algorithm," Hydrological Sciences Journal, no. just-accepted, 2019.
[6] W. Brutsaert, "Some exact solutions for nonlinear desorptive diffusion," Zeitschrift für angewandte Mathematik und Physik, vol. 33, no. 4, pp. 540–546, 1982.
[7] R. K. Singh, L. Ganapathi, P. Tiwari, and J. Narayan, "Effect of processing geometry in oxygen incorporation and insitu formation of YBa2Cu3O7 superconducting thin films by pulsed laser evaporation technique," Applied physics letters, vol. 55, no. 22, pp. 2351–2353, 1989.
[8] F. I. Morton, "STUDIES IN EVAPORATION AND THEIR LESSONS FOR THE ENVIRONMENTAL SCIENCES," Canadian Water Resources Journal, vol. 15, no. 3, pp. 261–286, 1990, doi: 10.4296/cwrj1503261.
[9] E. Sartori, "A critical review on equations employed for the calculation of the evaporation rate from free water surfaces," Solar Energy, 2000, doi: 10.1016/S0038-092X(99)00054-7.
[10] F. I. Morton, "Evaporation research—a critical review and its lessons for the environmental sciences," Critical reviews in environmental science and technology, vol. 24, no. 3, pp. 237–280, 1994.
[11] U. S. Panu and T. Nguyen, "Estimation of mean areal evaporation in northwestern Ontario," Canadian Water Resources Journal, vol. 19, no. 1, pp. 69–82, 1994.
[12] H. D. Abeysiriwardana, N. Muttil, and U. Rathnayake, "A Comparative Study of Potential Evapotranspiration Estimation by Three Methods with FAO Penman–Monteith Method across Sri Lanka," Hydrology, vol. 9, no. 11, p. 206, 2022.
[13] K. Khosravi et al., "Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq," Computers and Electronics in Agriculture, vol. 167, p. 105041, 2019.
[14] D. Han and A. Moghaddamnia, "Reply to comments on 'Evaporation estimation using artificial neural networks and adaptive neurofuzzy inference system techniques' by A. Moghaddamnia, M. Ghafari Gousheh, J. Piri, S. Amin and D. Han [Adv. Water Resour. 32 (2009) 88–97]," Advances in Water Resources, vol. 32, no. 6, pp. 967–968, 2009, doi: 10.1016/j.advwatres.2009.02.012.
[15] J. F. Griffiths, "Another evaporation formula," Agricultural Meteorology, vol. 3, no. 3–4, pp. 257–261, 1966.
[16] H. L. Penman, "Natural evaporation from open water, bare soil and grass," Proceedings of the Royal Society of London. Series A: Mathematical and physical sciences, vol. 193, no. 1032, pp. 120–145, 1948, doi: 10.1098/rspa.1948.0037.
[17] C. H. B. Priestley and R. J. Taylor, "On the Assessment of the Surface heat Flux and Evaporation using Large-scale Parameters," Monthly Weather Review, vol. 100, pp. 81–92, 1972.
[18] L. Wu, G. Huang, J. Fan, X. Ma, H. Zhou, and W. Zeng, "Hybrid extreme learning machine with meta-heuristic algorithms for monthly pan evaporation prediction," Computers and Electronics in Agriculture, vol. 168, p. 105115, 2020, doi: 10.1016/j.compag.2019.105115.
[19] R. C. Deo and M. Şahin, "Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia," Atmospheric Research, vol.

153, pp. 512–525, 2015, doi: 10.1016/j.atmosres.2014.10.016.
[20] A. D. Mehr and O. Akdegirmen, "Estimation of Urban Imperviousness and its Impacts on Flashfloods in Gazipaşa, Turkey," Knowledge-Based Engineering and Sciences, vol. 2, no. 1, pp. 9–17, 2021.
[21] A. Malik et al., "Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model," Engineering Applications of Computational Fluid Mechanics, vol. 14, no. 1, pp. 323–338, 2020.
[22] L. Diop et al., "The influence of climatic inputs on stream-flow pattern forecasting: case study of Upper Senegal River," Environmental Earth Sciences, vol. 77, no. 5, p. 182, 2018.
[23] M. Ali, R. Prasad, Y. Xiang, and Z. M. Yaseen, "Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts," Journal of Hydrology, 2020, doi: 10.1016/j.jhydrol.2020.124647.
[24] R. Tur and S. Yontem, "A Comparison of Soft Computing Methods for the Prediction of Wave Height Parameters," Knowledge-Based Engineering and Sciences, vol. 2, no. 1, pp. 31–46, 2021.
[25] O. Kisi, B. Choubin, R. C. Deo, and Z. M. Yaseen, "Incorporating synoptic-scale climate signals for streamflow modelling over the Mediterranean region using machine learning models," Hydrological Sciences Journal, no. just-accepted, 2019.
[26] S. R. Naganna, B. H. Beyaztas, N. Bokde, and A. M. Armanuos, "ON THE EVALUATION OF THE GRADIENT TREE BOOSTING MODEL FOR GROUNDWATER LEVEL FORECASTING," Knowledge-Based Engineering and Sciences, vol. 1, no. 1, pp. 48–57, 2020.
[27] M. Gocić et al., "Soft computing approaches for forecasting reference evapotranspiration," Computers and Electronics in Agriculture, vol. 113, pp. 164–173, 2015, doi: 10.1016/j.compag.2015.02.010.
[28] S. Q. Salih et al., "Viability of the advanced adaptive neuro-fuzzy inference system model on reservoir evaporation process simulation: case study of Nasser Lake in Egypt," Engineering Applications of Computational Fluid Mechanics, vol. 13, no. 1, pp. 878–891, 2019, doi: 10.1080/19942060.2019.1647879.
[29] M. A. Ghorbani, M. A. Jabehdar, Z. M. Yaseen, and S. Inyurt, "Solving the pan evaporation process complexity using the development of multiple mode of neurocomputing models," Theoretical and Applied Climatology, 2021, doi: 10.1007/s00704-021-03724-8.
[30] H. Moeeni and H. Bonakdari, "Impact of Normalization and Input on ARMAX-ANN Model Performance in Suspended Sediment Load Prediction," Water Resources Management, vol. 32, no. 3, pp. 845–863, 2018, doi: 10.1007/s11269-017-1842-z.
[31] A. Sharafati, S. B. H. S. Asadollah, and A. Neshat, "A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran," Journal of Hydrology, 2020, doi: 10.1016/j.jhydrol.2020.125468.
[32] A. Sharafati, R. Yasa, and H. M. Azamathulla, "Assessment of stochastic approaches in prediction of wave-induced pipeline scour depth," Journal of Pipeline Systems Engineering and Practice, vol. 9, no. 4, 2018, doi: 10.1061/(ASCE)PS.1949-1204.0000347.
[33] D. Myronidis, K. Ioannou, D. Fotakis, and G. Dörflinger, "Streamflow and Hydrological Drought Trend Analysis and Forecasting in Cyprus," Water Resources Management, 2018, doi: 10.1007/s11269-018-1902-z.
[34] K. P. Sudheer, A. K. Gosain, D. Mohana Rangan, and S. M. Saheb, "Modelling evaporation using an artificial neural network algorithm," Hydrological Processes, vol. 16, no. 16, pp. 3189–3202, 2002.
[35] M. A. Ghorbani, R. C. Deo, Z. M. Yaseen, M. H. Kashani, and B. Mohammadi, "Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran," Theoretical and Applied Climatology, vol. 133, no. 3–4, pp. 1119–1131, 2017, doi: 10.1007/s00704-017-2244-0.

[36] J. Schmidhuber, "Deep Learning in neural networks: An overview," Neural Networks, vol. 61. pp. 85–117, 2015. doi: 10.1016/j.neunet.2014.09.003.
[37] A. Halevy, P. Norvig, and F. Pereira, "The Unreasonable Effectiveness of Data," IEEE Intelligent Systems, vol. 24, no. 2, pp. 8–12, 2009, doi: 10.1109/mis.2009.36.
[38] T. Tiyasha, T. M. Tung, and Z. M. Yaseen, "Deep Learning for Prediction of Water Quality Index Classification: Tropical Catchment Environmental Assessment," Natural Resources Research, pp. 1–20, 2021.
[39] J. Tompson, M. Stein, Y. Lecun, and K. Perlin, "Real-time continuous pose recovery of human hands using convolutional networks," ACM Transactions on Graphics (ToG), vol. 33, no. 5, pp. 1–10, 2014.
[40] G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Processing Magazine, 2012, doi: 10.1109/MSP.2012.2205597.
[41] I. Sutskever, O. Vinyals, and Q. V Le, "Sequence to sequence learning with neural networks," in Advances in neural information processing systems, 2014, pp. 3104–3112.
[42] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," Mechanical Systems and Signal Processing. 2018. doi: 10.1016/j.ymssp.2017.11.024.
[43] M. Fu et al., "Integration of complete ensemble empirical mode decomposition with deep long short-term memory model for particulate matter concentration prediction," Environmental Science and Pollution Research, pp. 1–12, 2021.
[44] Y. Geng, L. Su, Y. Jia, and C. Han, "Seismic events prediction using deep temporal convolution networks," Journal of Electrical and Computer Engineering, vol. 2019, 2019.
[45] C. Yang and S. Guo, "Inflation Prediction Method Based on Deep Learning," Computational Intelligence and Neuroscience, vol. 2021, 2021.
[46] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 157–166, 1994, doi: 10.1109/72.279181.
[47] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[48] D. Zhang, G. Lindholm, and H. Ratnaweera, "Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring," Journal of Hydrology, vol. 556, pp. 409–418, 2018, doi: 10.1016/j.jhydrol.2017.11.018.
[49] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, "Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas," Journal of Hydrology, vol. 561, pp. 918–929, Jun. 2018, doi: 10.1016/j.jhydrol.2018.04.065.
[50] B. B. Sahoo, R. Jha, A. Singh, and D. Kumar, "Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting," Acta Geophysica, vol. 67, no. 5, pp. 1471–1481, 2019.
[51] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks," Hydrology and Earth System Sciences, vol. 22, no. 11, pp. 6005–6022, 2018, doi: 10.5194/hess-22-6005-2018.
[52] F. Kratzert, D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing, "Benchmarking a Catchment-Aware Long Short-Term Memory Network (LSTM) for Large-Scale Hydrological Modeling," Hydrology and Earth System Sciences Discussions, 2019, doi: 10.5194/hess-2019-368.
[53] P. Sharma and S. Singh, "Artificial neural network approach for hydrologic river flow time series forecasting," Agricultural Research, pp. 1–12, 2021.
[54] Y. Wang, "A new concept using LSTM Neural Networks for dynamic system identification," 2017. doi: 10.23919/ACC.2017.7963782.

[55] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, "Single Layer Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting," Procedia Computer Science, vol. 135, pp. 89–98, 2018, doi: 10.1016/j.procs.2018.08.153.
[56] Q. Cai et al., "Short-term load forecasting method based on deep neural network with sample weights," International Transactions on Electrical Energy Systems, vol. 30, no. 5, 2020, doi: 10.1002/2050-7038.12340.
[57] S. Dhakal, Y. Gautam, and A. Bhattarai, "Exploring a deep LSTM neural network to forecast daily PM2.5 concentration using meteorological parameters in Kathmandu Valley, Nepal," Air Quality, Atmosphere Health, vol. 14, no. 1, pp. 83–96, Jan. 2021, doi: 10.1007/s11869-020-00915-6.
[58] M. Cho, J. Chang, and C. Huang, "Application of parallel Elman neural network to hourly area solar PV plant generation estimation," International Transactions on Electrical Energy Systems, vol. 30, no. 8, 2020, doi: 10.1002/2050-7038.12470.
[59] and A. T. B. Zahroh, S., Y. Hidayat, R. S. Pontoh, A. Santoso, "Modeling and forecasting daily temperature in bandung," in Proceedings of the International Conference on Industrial Engineering and Operations Management, 2019, pp. 406–412.
[60] Mathworks. 2020b, "Long short-term memory networks."
[61] M. Guermoui, F. Melgani, and C. Danilo, "Multi-step ahead forecasting of daily global and direct solar radiation: A review and case study of Ghardaia region," Journal of Cleaner Production, 2018, doi: 10.1016/j.jclepro.2018.08.006.
[62] C. E. Rasmussen and C. K. I. Williams, "Gaussian Processes for Machine Learning." The MIT Press, 2005. doi: 10.7551/mitpress/3206.001.0001.
[63] E. Shabani, B. Hayati, E. Pishbahar, M. A. Ghorbani, and M. Ghahremanzadeh, "A novel approach to predict CO2 emission in the agriculture sector of Iran based on Inclusive Multiple Model," Journal of Cleaner Production, vol. 279, p. 123708, 2021, doi: 10.1016/j.jclepro.2020.123708.
[64] Z. S. Khozani et al., "Determination of compound channel apparent shear stress: application of novel data mining models," Journal of Hydroinformatics, 2019.
[65] K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," Journal of Geophysical Research: Atmospheres, vol. 106, no. D7, pp. 7183–7192, 2001, doi: 10.1029/2000JD900719.
[66] B. Majhi and D. Naidu, "Pan evaporation modeling in different agroclimatic zones using functional link artificial neural network," Information Processing in Agriculture, vol. 8, no. 1, pp. 134–147, Mar. 2021, doi: 10.1016/j.inpa.2020.02.007.
[67] B. Majhi, D. Naidu, A. P. Mishra, and S. C. Satapathy, "Improved prediction of daily pan evaporation using Deep-LSTM model," Neural Computing and Applications, 2019, doi: 10.1007/s00521-019-04127-7.
[68] H. Wang, H. Yan, W. Zeng, G. Lei, C. Ao, and Y. Zha, "A novel nonlinear Arps decline model with salp swarm algorithm for predicting pan evaporation in the arid and semi-arid regions of China," Journal of Hydrology, vol. 582, p. 124545, Mar. 2020, doi: 10.1016/j.jhydrol.2020.124545.
[69] X. Lu, Y. Ju, L. Wu, J. Fan, F. Zhang, and Z. Li, "Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models," Journal of Hydrology, 2018, doi: 10.1016/j.jhydrol.2018.09.055.
[70] Y. Guan et al., "A novel approach for predicting daily pan evaporation in the coastal regions of Iran using support vector regression coupled with krill herd algorithm model," Theoretical and Applied Climatology, vol. 142, no. 1–2, pp. 349–367, Oct. 2020, doi: 10.1007/s00704-020-03283-4.
[71] M. Abed, M. A. Imteaz, A. N. Ahmed, and Y. F. Huang, "Application of long short-term memory neural network technique for predicting monthly pan evaporation," Scientific Reports, vol. 11, no. 1, pp. 1–19, 2021.