 doi:10.20944/preprints202209.0398.v1

Article

Exploring Temporal Dynamics of River Discharge using Univariate Long Short-Term Memory (LSTM) Recurrent Neural Network at East Branch of Delaware River

Md Abdullah Al Mehedi [1], Marzieh Khosravi [2,*], Munshi Md Shafwat Yazdan [3], and Hanieh Shabanian [4]

[1] Graduate Research Assistant, College of Engineering, Villanova University, PA 19085, USA; mmehedi@villanova.edu
[2] Graduate Research Assistant, College of Engineering, Villanova University, PA 19085, USA; mkhosrav@villanova.edu
[3] Graduate Research Assistant, Civil and Environmental Engineering, Idaho State University, Pocatello, ID 83209, USA; yazdmuns@isu.edu
[4] Assistant Professor, Department of Computer Science, Northern Kentucky University, KY 41099, USA; shabanianh1@nku.edu
* Correspondence: mkhosrav@villanova.edu

Abstract: River flow prediction is a pivotal task in the field of water resource management during the era of rapid climate change. The highly dynamic and evolving nature of the climatic variables e.g., precipitation has a significant impact on the temporal distribution of the river discharge in recent days making the discharge forecasting even more complicated for diversified water-related issues e.g., flood prediction and irrigation planning. To predict the discharge, various physics-based numerical models are used using numerous hydrologic parameters. Extensive lab-based investigation and calibration are required to reduce the uncertainty involved in those parameters. However, in the age of data-driven predictions, several deep learning algorithms showed satisfactory performance in dealing with sequential data. In this research, Long Short-term Memory (LSTM) neural network regression model is trained using over 80 years of daily data to forecast the discharge time series up to 3 days ahead of time. The performance of the model is found satisfactory through the comparison of the predicted data with the observed data, visualization of the distribution of the errors and Root Mean Squared Error (RMSE) value of 0.09. Higher performance is achieved through the increase in the number of epochs and hyper parameter tuning. This model can be transferred to other locations with proper feature engineering and optimization to perform univariate predictive analysis and potentially be used to perform real-time river discharge prediction.

Keywords: river discharge; hydro informatics; water resource; data-driven; deep learning; LSTM

1. Introduction

River discharge forecasting is considered a pivotal task in various fields of water resource management i.e., flood control, irrigation planning, and hydropower production [1–7]. River discharge has a significant impact on the physical, chemical, and biological activities in the river contributing high correlation to the fluvial ecosystem [8–11]. Forecast is established on the probability of the river flow and its strong historical data or records. As a result of highly efficient decision-making capability, both the short-term and long-term (e.g., hourly, and daily) forecasting models have a significant interest into the research and scientist community [12–15]. Due to the complexity, climate changeability and the effects on anthropology, hydrological data holds a strong constraint to the advancement of short-term forecasting models [16–19]. Methodologies for river discharge modeling and forecasting can be divided into three groups: conceptual, physics-based, and data-driven models [20]. A lot of research has already been conducted with the focus on the use of conceptual and physics-based models [21,22]. Due to the high degree of spatial and

 doi:10.20944/preprints202209.0398.v1

temporal variability, many hydrological processes, have become very difficult to formulate and model [23]. In addition, several research demonstrate that conceptual and physics-based models have a limited competence to provide short-term forecasts and require long-term datasets to perform the computationally expensive model calibration [24]. Furthermore, some studies show that conceptual and physics-based models, in some circumstances, showed limited capability in flood forecasting in higher dimensions [25]. In the last decade, many scientists and researchers have examined the pros and cons of various conceptual and physics-based models and have compare their prediction performance with that of emerging data-driven techniques e.g., Deep Learning (DL) methods [11,26]. Recently, the practice of data-driven techniques of DL methods, have drawn substantial consideration for the sequential data e.g., precipitation, river flow prediction applications. Scientists from all over the world have concluded that data-driven techniques of DL methods are qualified in capturing non-linear processes numerically without the knowledge of the underlying conceptual or physical processes involved [27–30].

This study is a significant step towards river discharge prediction using the DL method which contributes to understand the dynamics of river flow and offer a framework to predict the impact of it on agriculture, fluvial ecology, flooding, irrigation, and water supply planning in the nearby areas. The proposed methodology is applied to East Branch of Delaware River, in Delaware County, NY. A dam was built to create a reservoir named Peacton for maintaining drinking water supply to the community in Downsville [31]. Due to the construction of the dam thus the withheld of the water upstream creates significant variation in the downstream portion of the river. Rapid climate change is the main trigger to increase the intensity of rainfall over the area. Consequently, fluctuation of the river flow at the area has increased through the years. Fish and Fluvial ecology are one of those sectors which are getting significant impact due to these rapidly changing dynamics of river discharge [32,33]. Therefore, a robust data-driven predictive approach utilizing only the previous observed discharge data could contribute substantially to the nearby community in addressing the issues with less computational efforts.

Data-driven techniques can be divided into two methods, such as statistical methods and black-box models. Statistical models deal with autoregressive Moving Average (MA) techniques [34,35], on the other hand the black box models follow various techniques of Artificial Intelligence (AI) [36]. The Machine Learning (ML) methods to do the river discharge prediction consist of Artificial Neural Networks (ANN) [37–43], Support Vector Regression (SVR) [44–48], Decision Tree (DT) model [49,50], fuzzy inference system [51], Bayesian particle filter [52], expert system [53], hybrid model [54], and Multiple Linear Regression (MLR) [55]. The spatial and temporal hidden patterns in historical data are explored by these ML/DL methods without using a conceptual or even physical model as it requires a large number of physical parameters and broad understanding of the physical processes in the domain of the model [56]. In many cases, streamflow analysis using ANN can accomplish a better and accurate predictive performance than conceptual or physically based models [57,58]. However, the traditional ANN algorithms e.g., Feed Forward Neural Network (FFNN) do not have the ability to learn sequential data as they cannot recall previous information flow, resulting in a constrained prediction capability for long-term time series, e.g., temporal distribution of the discharge/water depth. The FFNN algorithms need complex procedures in the data pre-processing stage to obtain good performance in predicting the target variables. While the comprehensive data pre-processing can bolster the ability of a FFNN model to learn from the observed data, it requires a significant amount of user intervention, e.g., selecting the number of reconstructed components38. In addition, the pre-processing requires substantial amount of time as many reconstructed components need to be calculated.

In contrast, an updated version of Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) has achieved a significant attention among the water scientist community specifically for time series prediction. LSTM was first introduced by Hochreiter and Schmidhuber in 1997, which has later been established as a powerful tool for addressing forecasting problems [59–63]. High computational effort and time is needed by

 doi:10.20944/preprints202209.0398.v1

recurrent backpropagation to learn to store long-term information because of the decaying error backflow. Hence, the concept of the vanishing gradient problem in recognizing long-term dependency of RNN was introduced [64]. LSTM has become a very popular algorithm to deal with time series data in the DL forecasting where variables are dependent on the previous information along the series. LSTM can capture the long-term dependencies and linkage among the variables. Compared to the traditional neural networks e.g., FFNN, LSTMs are capable of capturing both the Chaotic and periodic behaviors of time series data and determine the dependencies with higher accuracy [65]. Kratzert et al. (2018) mentioned the LSTM model as a successful adaptation [42] while describing the rainfall-runoff pattern of a large complex catchments at daily scale [66]. On the other hand, Ni et al. (2019) developed multiple hybrid models, using LSTM model, to analyze monthly streamflow and rainfall prediction [67]. Hu et al. (2018) found that the LSTM model outperformed other traditional ANN model while forecasting the flood frequency and found the result was up to 6 hours ahead [68]. Similar analysis was done by Le et al. (2019), who compared the ANN and LSTM models for predicting the daily scale, two-day and three-day scale ahead streamflow rate at Hoa Binh. The outcome of using LSTMs model revealed that it could determine both the long-term and short-term dependencies between sequential complex data series and perform good results in river discharge forecasting.

The objective of this study is to develop an efficacious and concrete predictive framework with LSTM neural network to forecast the river discharge trained by previous data untangling the temporal dynamics of large range of data. To accomplish the goal, an extensive Exploratory Data Analysis (EDA), Feature Engineering (FE) and hyperparameter optimization are conducted to obtain the best possible performance and a set of learned parameters from the LSTM model. The proposed framework can be used to help researchers, engineers, and decision-makers to perceive the temporal dynamics of discharge and make the accurate engineering/managerial decisions. Engineers and managers will be able to observe both the short-term and long-term behavior and trend of discharge which will eventually help making the precautionary measures for various water-related issues in the surrounding area using the previous observational discharge values. As the LSTM-based approach showed in this paper requires the observed data only, the burden of high computational effort needed for the physics-based numerical models can be reduced significantly. The rest of the paper is organized as follows. In Section 2, the study area and observational data are introduced. Section 3 illustrates the exploratory analysis and FE to prepare the dataset for LSTM model training, the proposed model's architecture, and model evaluation criteria. Section 4 presents the experiment results and discussion. Finally, Section 5 concludes the paper with future recommendations.

2. Materials and Methods

2.1. Study Location and Data Source

The river flow measuring site considered in this study is situated along the East Branch Delaware River, in Delaware County, NY (**Figure 1**). The latitude and longitude of the study site are 42°04'30" N, 74°58'35" W with the coordinate system "North American Datum of 1983". The identification number of the hydrologic unit in the USGS Water database of it is 02040102, on left bank half miles downstream from Downsville Dam, at the downstream end of the outlet channel of Pepacton Reservoir, and one mile east of Downsville. The contributing drainage area of the measuring point is 372 mi^2. The river flow recording gage datum is 1,101 ft. The study location is the situation at the hamlet, Downsville, a census-designated place which was a village in the Colchester town, Delaware County, NY.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

Figure 1. Aerial photo of the study location (a-b) with flow measuring station c) at East Bank Delaware river (USGS 2022).

The period of discharge record is from July 1941 to present. A range of the discharge time series considered in this research is 07/01/1941 to 12/31/2021 with 29393 observations. Maximum yearly flow recorded within the entire period after the construction of Pepacton reservoir is 17,700 ft^3/s on 09/18/1946 with the water level, 12.08 ft and the minimum flow, 0.6 ft^3/s on 10/11/1991 with the minimum water height, 1.39 ft on 01/17/1964. Extreme water level recorded outside the period of record is 16 ft on 10/09/1903 during a flooding event in the surrounding area. Extreme river flow recorded outside the period of record before the construction of Pepacton reservoir is 23,900 ft^3/s on 11/26/1950 with the water level 14.52 ft [USGS 2022].

2.2. Univariate Exploratory Data Analysis and Feature Engineering

To obtain the characteristics and attributes of the dataset with discharge series, EDA is performed. EDA is a process where the internal distribution of a dataset is extracted through various graphical visualization and summarizing techniques. EDA involves a critical process of conducting initial exploration on the variables to investigate the anomalies and hidden patterns. It is the first step towards data preprocessing for ML/DL algorithms. In this study steps in EDA can be further categorized into two parts. It involves descriptive statistics, outlier detection, and probability distribution to determine the skewness. As the scope of this research is univariate analysis i.e., only one variable (discharge) (**Figure 2**), a brief study is performed in the descriptive statistics (Table 1). They are counting the number of observations, obtaining the central values (e.g., mean, median, mode), spread (e.g., standard deviation), range (e.g., minimum, maximum), percentile distribution, interquartile range, quantifying missing data. In probability distribution, graphical representation using histogram and the coefficient of skewness is used to analyze the normality of the discharge series. In this study, the discharge series consists of a significant number of extreme values/outliers. Storm events with a significant amount of rainfall have the greatest impact on the high values of discharge volume. Detecting the outliers revealed both extreme events and the erroneous measurements of discharge.

Univariate outlier detection for the discharge series is performed using the interquantile-range (IQR) rule. According to the IQR proximity rule, a value of the continuous numeric variable is considered an outlier if it stays outside the upper boundary i.e., 75th quantile + (IQR * 1.5) or lower boundary i.e., 25th quantile - (IQR * 1.5) where the IQR is expressed by the difference between 75th quantile and 25th quantile.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

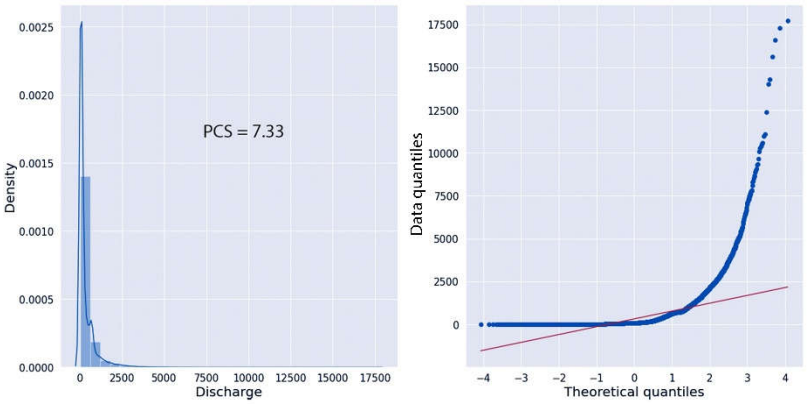

Figure 2. Probability Distribution and Q-Q plot of the discharge series.

Table 1. Descriptive Statistics of the entire discharge series

Count	29393
Mean	323.73
Standard Deviation	681.61
Minimum	0.6
25th percentile	42
50th percentile	88.4
75th percentile	332
Maximum	17700
Inter Quantile Range (IQR)	290.00

After the EDA step, FE is performed. FE is the most crucial step to obtain the appropriate dataset for training/testing the LSTM algorithm. In FE, imputation to make the data set consistent, normality check, necessary data transformation if applicable and data standardization is performed. Without a successful FE, any data-driven method may not yield to a satisfactory performance with minimum error. An adequate optimization through the iterative gradient descent cannot be reached without a successful scrutiny of the dataset. Therefore, a comprehensive FE is performed to transform the dataset most suitable for the learning algorithm of LSTM. As the discharge variable has some null values, imputation with the median of the series is performed to make the series consistent. Direct discarding the observations with null values can also be considered. However, this technique is not recommended as it reduced the shape of the dataset by reducing the observed data point.

After the imputation task, normality of the discharge series is checked using visualization technique accompanied by coefficient of normality measure e.g., Pearson Coefficient of Skewness (PCR). Normal distribution is the most crucial factor in the field of data-driven predictive analysis e.g., deep neural network regression. Smooth progress towards minima in gradient descent required to reach the objective function that the step sizes be updated at the same rate for the values of each feature used in the analysis, which can be achieved by increasing the normality of the variable. Logarithmic transformation of the discharge series conveyed significant increase in the normality. In addition, processing and recalling long-term information in time series data is a unique feature of LSTM model which makes it different from the traditional feedforward neural network. Therefore, a combination of appropriate feature engineering to increase the normality with the robustness of LSTM algorithm is applied in this study to obtain satisfactory performance in discharge prediction. As the distribution of the values of river discharge series is highly skewed to the left indicating non-normally distributed data, the traditional ML and neural

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

network regression algorithms without appropriate data transformation do not offer satisfactory performance even with good optimization. Therefore, the LSTM which is a special type of RNN is used to forecast the river discharge values in this study. In the next section, a detailed explanation of how the LSTM is utilized for the river discharge time series is presented.

As the distribution of the discharge series is found to be highly skewed, data transformation is performed to decrease the non-normality of the series. In this study, three methods of data transformation are considered e.g., logarithmic, square-root, cubic transformation to transform the distribution of the features more to the normal distribution. Pearson's coefficient is used as a numerical indicator of normality. In the Figure 3**Error! Reference source not found.**, distributions of the transformed and observed discharge series can be seen. Data transformation with all three functions mentioned above helps to increase the normality thus decrease the skewness. The Pearson's Coefficient of Skewness (PCS) are also added to the individual figures to get the idea of the degree of the skewness. Logarithmic transformation with the lowest PCS of 0.55 outperforms other transformation functions it reduced the right-skewness more than other functions which can be observed in the Figure 3. As the PCS value from the logarithmic transformation is promising compared all others, the data series from this transformation is considered in this research. Logarithmic transformation outperforms other transformation functions, e.g., reduced the non-normality significantly by changing the value of Pearson's coefficient of skewness (PCS).

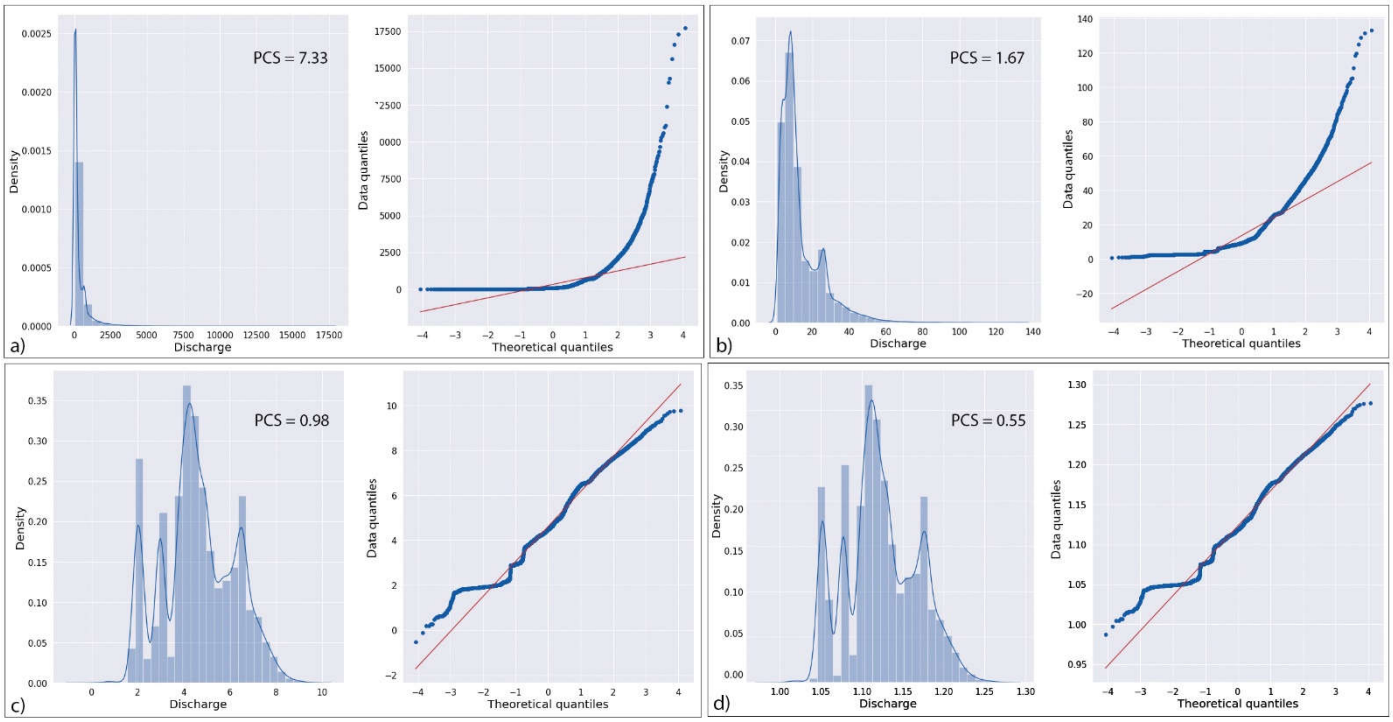

Figure 3. Multiple data transformation techniques are applied to observe the shift in the skewness i.e., decrease in the normality. Logarithmic transformation outperforms other transformation functions. (a) original data (b) square-root-transformed; (c) cubic-transformed; (d) log-transformed data.

The autoregressive integrated MA model is the most extensively used parametric time series method. Although determining trend significance can be challenging, among the analyses and different methods the most direct approaches for detecting discharge trends in a time series is the MA method. The MA method as one of the most fundamental ways for analyzing meteorological and hydrological data smooths and clarifies trend lines by screening out frequent random variations in hydrological data. The focus of the analysis is more on the accentuate longer-term patterns rather than the short-term fluctuation for a better representation of MA. It can be calculated for a different duration and number

 doi:10.20944/preprints202209.0398.v1

of years by shifting the average value of specified variable year by year and including all the data points and conclude the entire data set at the end of the process for target duration range. For this study the MA method used to analyze discharge variation as a yearly average for the duration of 81 years (from 1941 to 2021). The average discharge investigated by considering two types of MA models: the Simple Moving Average (SMA) and Exponential Moving Average (EMA). SMA is an accounting MA that is calculated by averaging recent discharge and adding recent values. Then dividing the outcome by the number of periods in the data range (Eq. 1).

$$SMA = \frac{Q_1 + Q_3 + Q_3 + \cdots + Q_N}{N} \quad (1)$$

Where N is the total number of periods, and Q_N is the discharge value at period N.

An exponential moving average (EMA) is a type of MA that assigns higher weight to the recent data and measures direction of the discharge trend over a period of time (Eq. 2). The EMA benefits from the recent discharge changes and capable of capturing the possible recent year trend and climate change impact.

$$EMA = Q_t \times k + EMA_{t-1} \times (1 - k) \quad (2)$$

$$k = \frac{2}{(N + 1)} \quad (3)$$

Where k is the weighted multiplier, t is represented as the present values and $(t - 1)$ is a symbol for the previous period. The time interval is dependent on the time-series dataset and the interval in the actual dataset (i.e., day, hour, or minute).

Both 10-year SMA and EMA, MA analysis result in the **Figure 4** illustrated the average discharge to be highest in 1945, before decreasing to the minimum in the late 1955. However, the minimum range of average discharge continued for over a decay, compared to the second rise in its values in 1975, which was still about half of the maximum discharge average experienced in 1945. The decadal decrease was defined for more than two decays after 1975 along with the same increasing trend and more variance until 2012.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

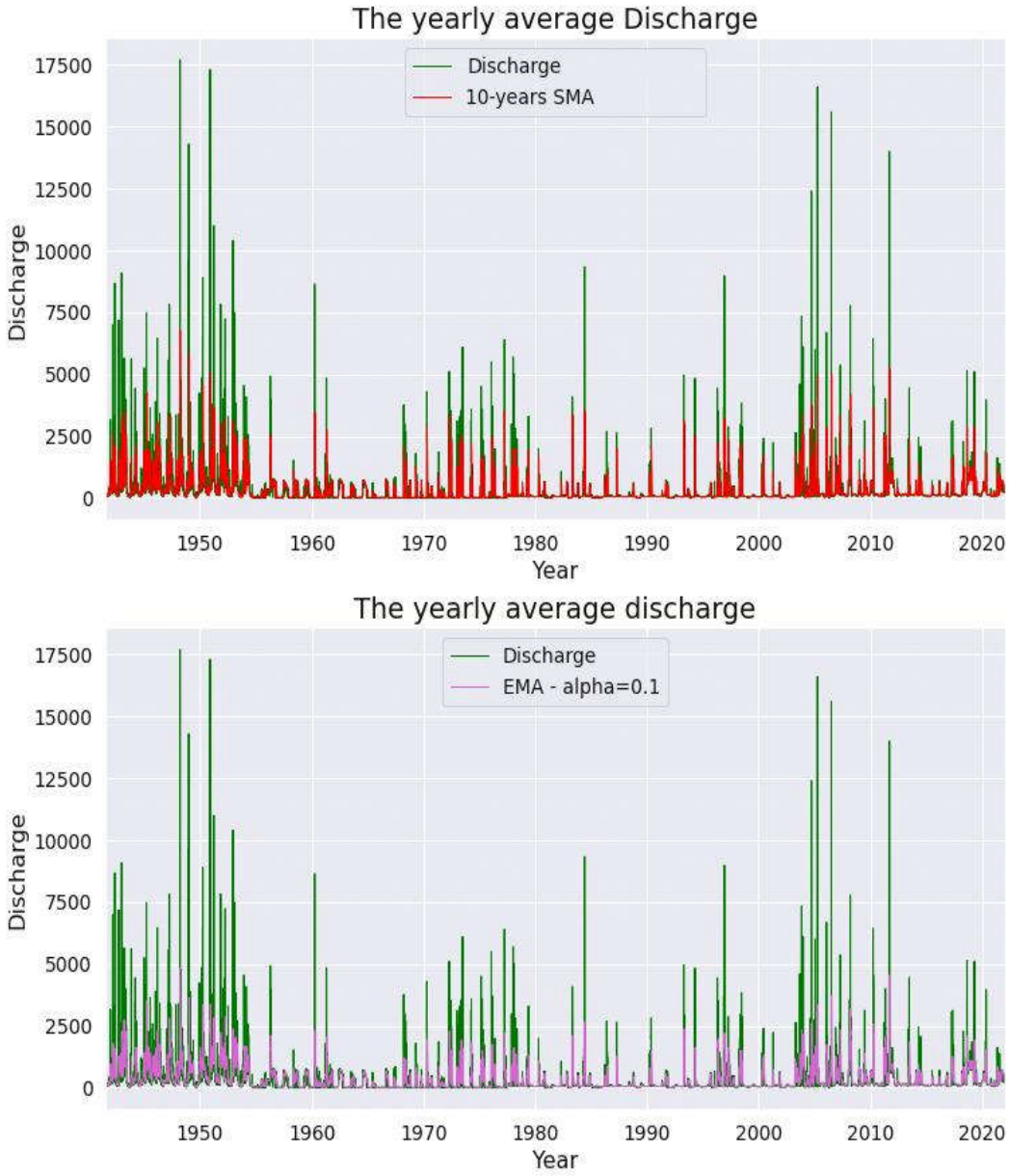

Figure 4. The yearly Moving Average (MA) with 10 years Simple Moving Average (MA) and Exponential Moving Average (EMA).

Through the data standardization process, the values of a variable are rescaled so that the variable has the mean 0 and variance 1 (or Z-score normalization) which is identical to the bell-shaped normal distribution curve. As the variable considered in this study is the continuous independent variable, the standardization of the variable is crucial for training/testing the neural network algorithm. Standardization is an important step for the optimization problem. The LSTM RNN model uses the gradient descent technique where the feature value (discharge) affects the step size of the technique. Smooth progress towards minima in gradient descent requires the update of the steps at the same rate for all the feature values. A standardized variable is a prerequisite of reaching the minima in the gradient descends process.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

All the values in the discharge series are standardized to prepare the training dataset for the LSTM model.

Equation 1 shows the formula of standardization of the discharge series. The difference between the discharge value and the minimum of the entire discharge series is divided by the range of the series provides the standardized data which is further used in the training/testing process of the LSTM. The entire standardized discharge series is split into two portions i.e., a training set that is used to train the model and a testing set that is used to test/evaluate the model. Seventy (70) percent of the dataset is used for training and thirty (30) percent is used for testing. In a nutshell, EDA, and FE are pivotal steps for the satisfactory performance of the predictive model.

2.3. Long Short-term Memory (LSTM) Recurrent Neural

LSTM is a special type of RNN which is frequently applied specially in the sequential forecasting. LSTM feedback connections is the principal component of processing and recalling long-term information and a unique feature which makes it different from the traditional feedforward neural network. This unique property of LSTM is utilized in processing the sequence of datasets e.g., discharge time series and treating all the datapoints independently. LSTM RNN is suitable for various water related variables with time series e.g., river flow, groundwater table, precipitation, etc. [32,64,69–72].

Both long-term memory (c[t-1]) and short-term memory (h[t-1]) are processed in a typical LSTM algorithm through the utilization of multiple gates to filter the information showed in the **Figure 5**. For an unchanged flow of gradients, forget and update gates update the memory cell state [73,74]. Three gates i.e., input gate i_g, forgot gate f_g and output gate o_g handle the information flow by writing, deleting, and reading respectively. Hence, LSTM is capable in memorizing information at different time tags and intervals making it suitable in time series prediction within a certain interval [75]. In forget gate, long-term information enters and passes through a filtration where unnecessary information is discarded. The forget gate filter out unnecessary data by using the sigmoid activation function where the range of the function is 0 (gate close) and 1 (gate open). Input gate filter and quantify the significance of a new data coming as input to the cell. Like the forget fate, input gate filters out information by using binary activation functions and controls the flow of both long-term and short-term information. The output gates regulate the value of the upcoming hidden state which is a function of the information on previous inputs. Entire schematic of the information flow through LSTM cells can be seen in the **Figure 6**.

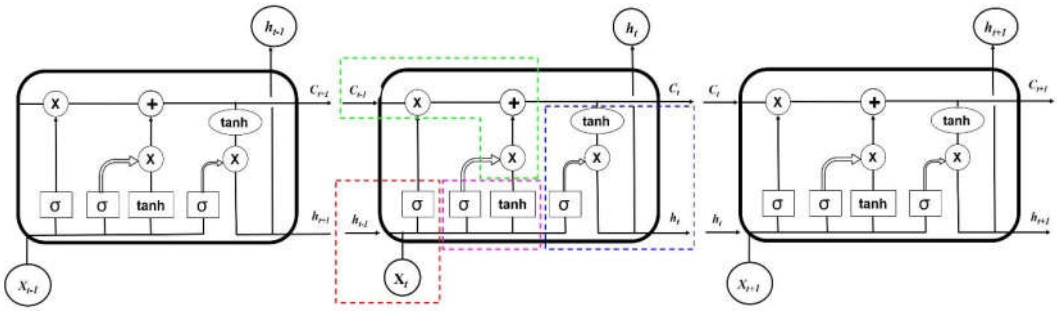

Figure 5. Schematic representation of a LSTM architecture

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

Table 2. All cell and gate operations are presented in the following equations for all the gates and cells

LSTM component	Equations
Forget gate	$f_g = sigmoid(X_t V_f + h_{t-1} W_f + b_f)$
Input gate	$i_g = sigmoid(X_t V_i + h_{t-1} W_i + b_i)$
Output gate	$o_g = sigmoid(X_t V_o + h_{t-1} W_o + b_o)$
Cell state	$C_t = i_g \odot \tilde{C}_t + f_g \odot C_{t-1}$
Candidate for cell state	$\tilde{C}_t = tanh\ tanh\ (X_t V_c + h_{t-1} W_c + b_c)$
Hidden state	$h_t = o_g \circ tanh\ tanh\ (C_t)$

In the above equations h_t represents a vector for the hidden state which is linked to the short-term memory. The C_t is the cell state linked to the long-term memory, and $\tilde{C}_t$ is the candidate for cell state at time tag t, which is used to filter important data to store over time. Several weight matrices are used in the input gate, forget gate and output gate and a cell state. They are indicated as W_i, W_f, W_o, W_c. For current input X_t, several weight matrices and biases i.e., V_i, V_f, V_o, V_o, and b_i, b_f, b_o, b_c are used. The operator $\odot$ denotes the Hadamard product (element-wise product)

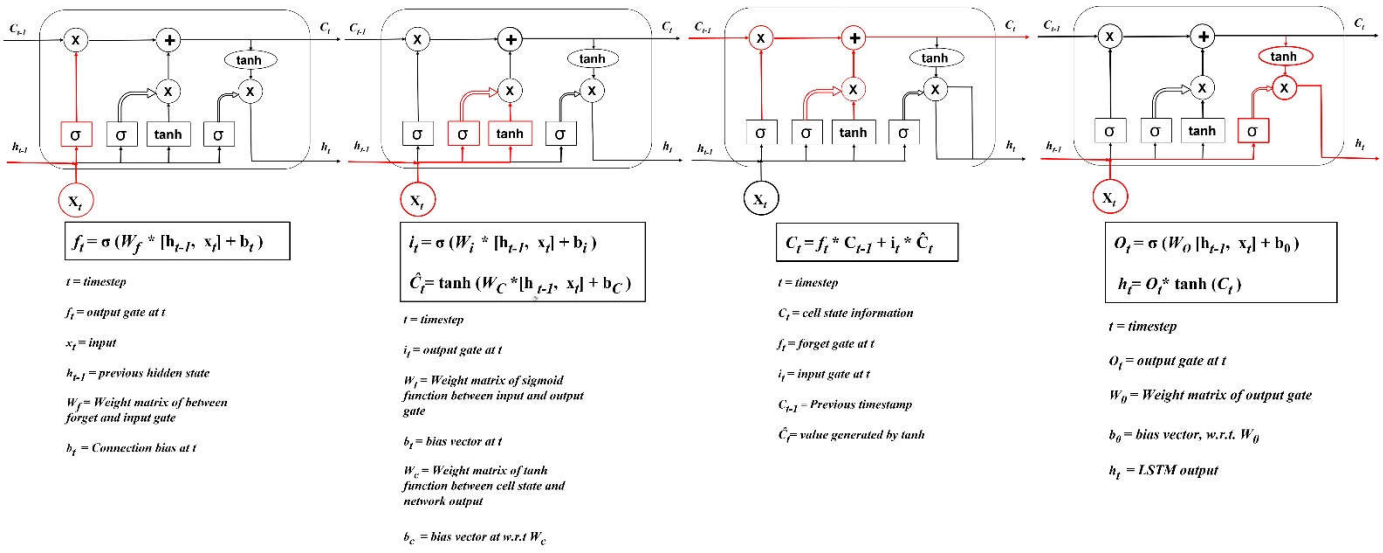

Figure 6. LSTM cells with gates (forget, input, cell, and output) showing the flow of information.

Hyperparameters are required to be tuned to maximize the performance of every ML model. Other parameters involved in the stochastic process of any ML model are learned through the iteration. However, the hyperparameters are decided manually. Therefore, they must be tuned to achieve satisfactory performance. As there is no systematic approach to select the hyperparameters, iterative trial and error method is applied to find the most appropriate values where the model performs best. In this study, Keras, python library which offers a space search for ML algorithms is used to find the best combination of the hyperparameters. Hyperparameter of the LSTM algorithm considered in this study are the size of epoch and batch and number of neurons [76,77].

2.4. Comparative Study

A comparative study is performed to investigate the performance of the LSTM algorithm with other DNN time series prediction algorithm. In this study, Multilayer Perceptron (MLP), RNN and Convolution Neural Network (CNN) is used to predict the discharge. A multilayer perceptron (MLP) is a fully connected type of feed-forward neural network (FFNN). An MLP is constructed with three layers i.e., an input, hidden and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

output layer. Except for the nodes in the input layer where the inputs are embedded, each node in the hidden layer is a neuron uses a nonlinear activation function. MLP is based on a supervised learning and backpropagation techniques for training. All the layers and non-linear activations distinct MLP algorithm from a linear perceptron-based prediction. Convolutional neural network (CNN) is a special type of neural network mathematical convolution in general matrix multiplication is at least one layer. CNN is specifically designed to preprocess the pixel data applied in image processing. CNN substitutes the mathematical operation known as convolution for generic matrix multiplication in at least one of its layers. A convolutional neural network is built of an input layer, hidden layer(s) and an output layer to generate the outcome of the model. Unlike the feed-forward neural network, the hidden layers of CNN include layers to perform convolutions on the input data. Convolutional layers are among the hidden layers in a CNN. This typically contains a layer that does a dot product of the input matrix of the layer with the convolution kernel. The convolution procedure develops a feature map as the convolution kernel moves across the input matrix for the layer, adding to the input of the following layer. Following this are further layers like normalizing, pooling, and fully connected layers.

2.5. Error Analysis

The literature provides a wide range of evaluation metrices that has been mostly used in hydrologic modelling to compare the prediction's accuracy. The predictions' error defined via various methods represents the difference between data points as the observed real values and predicted values as the measured ones. Multiple model variability settings were used in this study. The top four standard performance assessment methods to evaluate the analytical output and draw conclusion were Root Mean Square Error (RMSE), correlation coefficient (r), relative error (RE), and the Nash Sutcliffe model efficiency coefficient (E). Multiple metrics should indeed be employed to assess model accuracy rather than one error metric which can also effectively capture the high streamflow time series values. The term "norms" refers to the varied forms of multi-dimensional error measures. The norm normalizations lead to a dimensionless relative metric and reduce the error measures' sensitivity to the dimensions of the data frame. The most popular evaluation metric is the Root Mean Square Error (RMSE) as the function is more sensitive about significant errors. That's because the squared term multiplies greater errors exponentially more than smaller ones. RMSE is the mean of the absolute value of the errors and is normalized by the number of data points, N:

$$RMSE = \frac{1}{N}\sum_{t=1}^{N}\left|Q_{t(obs)} - Q_{t(com)}\right|^2 \tag{5}$$

where $Q_{t(obs)}$ = observed discharge, $Q_{t(com)}$ = computed discharge, so $(Q_{t(obs)} - Q_{t(com)})$ represents the error term between the real and measured value for each data point which is normalized by dividing by the total number of observations after summation of all terms. The lowest RMSE score corresponds to the best predictive accuracy.

The coefficient of determination (R^2) is a popular error metric for the accuracy of the model and the model fitness to the data points' values depicted by this metric. The better the model fits the data, the higher the R^2 is. The correlation coefficient (r) which is the second error function implemented in this study represented in Eq. 6.

$$r = \frac{\sum_{t=1}^{N}\left(Q_{t(com)} - \bar{Q}_{(com)}\right)\left(Q_{t(obs)} - \bar{Q}_{(obs)}\right)}{\sqrt{\left[\sum_{t=1}^{N}\left(Q_{t(com)} - \bar{Q}_{(com)}\right)^2\right]\left[\sum_{t=1}^{N}\left(Q_{t(obs)} - \bar{Q}_{(obs)}\right)^2\right]}} \tag{6}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

Where $\bar{Q}_{(com)}$ = average of computed discharge, $\bar{Q}_{(obs)}$ = average of observed discharge. The R^2 range is 0 to 1, with 0 indicating no correlation and 1 signifying perfect correlation between observed and computed values.

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (8)$$

The above equation shows the third error matrix, the Mean Absolute Percentage Error (MAPE) used in this study. MAPE is a measure of prediction accuracy of a time series forecasting method. . The fourth evaluation metric refers the Nash Sutcliffe model efficiency coefficient (E) and is one of the most widely used metrics for evaluating a hydrologic model's performance. E can be classified as one of the scaled forecasts that the compare the predicted error to the observed error.

$$E = 1 - \frac{\sum_{t=1}^{T}\left(Q_{t(obs)} - Q_{t(com)}\right)^2}{\sum_{t=1}^{T}\left(Q_{t(obs)} - \bar{Q}_{t(obs)}\right)^2} \quad (9)$$

where $\bar{Q}_{t(obs)}$ = average of observed discharge and E is dimensionless with a range of $\infty \leq E \leq 1$, with 1 as the largest value that E can obtained, representing the best model accuracy. Having a positive value for E (greater than zero), shows that the prediction and computed discharge value is better than simply selecting the average observed value.

3. Results

LSTM RNN are used to predict the time series of discharge data based on multiple lead times. Several lead time durations e.g., 15 mins, 30 mins, 1 hour, 5 hours, 10 hours, 24hours, 36 hours and 72 hours are used to forecast the future values of discharge. Model performances are recorded for the lead times based on several error indicators to show the efficiency of LSTM model in predicting the discharge values. Predicted values are compared to the observed dataset and the difference of them are computed to quantify the error matrices. Root mean square error (RMSE), correlation coefficient (R-squared), and Nash Sutcliffe (E) error are used to estimate error from the predicted discharge data in the comparative study of LSTM, CNN and MLP based predictions. Model performance is improved by the increase in iteration i.e., increase in the number of epochs. Error matrices are obtained through multiple models runs to demonstrate the linkage between the model performance and the lead times in LSTM model. Hyperparameters of the LSTM model are adjusted to optimize the model performance considering a set of batch size, epoch size and total number of neurons.

3.1. Predicted and Observed Discharge Series

The output from the LSTM algorithm is compared to the observed discharge data from USGS database through visualization in the **Error! Reference source not found.**. A comparative study with the prediction from CNN and MLP is also presented to investigate the LSTM performance. Both the observed and predicted discharge time series is plotted in cubic feet per second (cfs) against the number of observations. The overall distribution of the predicted discharge values from LSTM, CNN and MLP are approximately identical to the observed data providing a satisfactory performance of all the algorithm. However, the LSTM approach outperforms the CNN and MLP based approaches in predicting discharge based on the past values. The error matrix, RMSE is for LSTM algorithm is 0.042 where 0.193 and 0.211 for CNN and MLP respectively where the MAPE values are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

0.92, 2.17 and 2.95. Model performance improvement through continuous iterations i.e., with the increase of the number of epochs is presented in the **Error! Reference source not found.** for all algorithms. However, LSTM ended up with the best performance after 100 epochs among all other algorithms with minimum error. After the LSTM model is trained with the training portion of the dataset which is highlighted in the **Error! Reference source not found.** (b), the entire observed dataset is fed to predict the outcome. The entire dataset is divided into training and testing sets with the proportion of 70% and 30%. Training dataset is used to train the model and testing dataset is used to evaluate the model performance. Observed data is showed in the **Error! Reference source not found.** (a) with green color. In the **Error! Reference source not found.** (b), deep cyan portion of the plot illustrate training portion of the dataset whereas deep blue portion shows the testing portion. The RMSE values of the training and testing portion are 0.097 and 0.045 respectively. The lowest RMSE score corresponds to the best predictive accuracy and shows the better satisfactory performance of LSTM algorithm

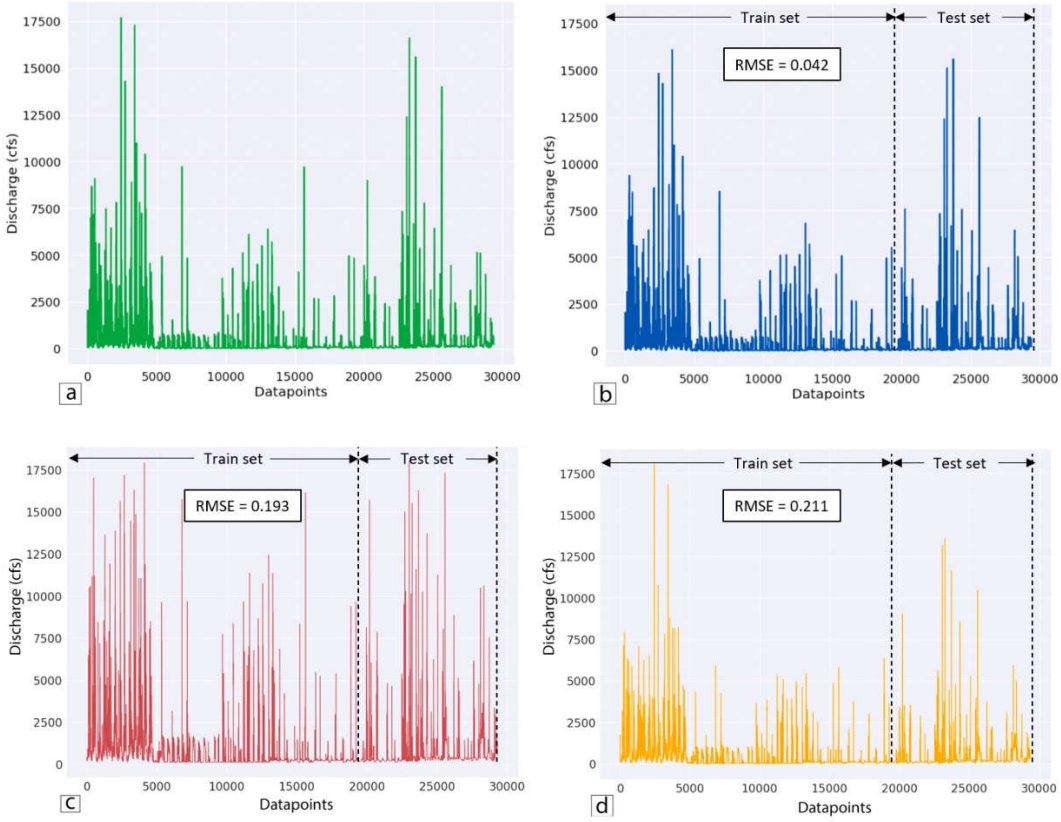

Figure 7. Distribution of observed (a), predicted discharge values from LSTM (b), CNN (c) and MLP (d) for the entire discharge time series used in this study.

Error matrices e.g., RMSE, R^2 and E are documented for several lead times and illustrated in the Figure 8. Time lags are pivotal parameters of LSTM algorithm towards model performance. Time lag values are Q(t1) = 15 mins, Q(t2) = 30 mins, Q(t3) =1 hour, Q(t4) = 5 hours, Q(t5) = 10 hours, Q(t6) = 24 hours, Q(t7) = 36 hours and Q(t8) = 72 hours. The values of RMSE decreases with the increase in the number of time lags whereas of R^2 and E increases showing the improvement in the model performance in the **Error! Reference source not found.**. The increase in the time lags conveys increase in the duration of prediction e.g., discharge prediction after 48 hours from present. The RMSE value 0.96 of for the time lag of 72 hours is the lowest among all the time lags.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

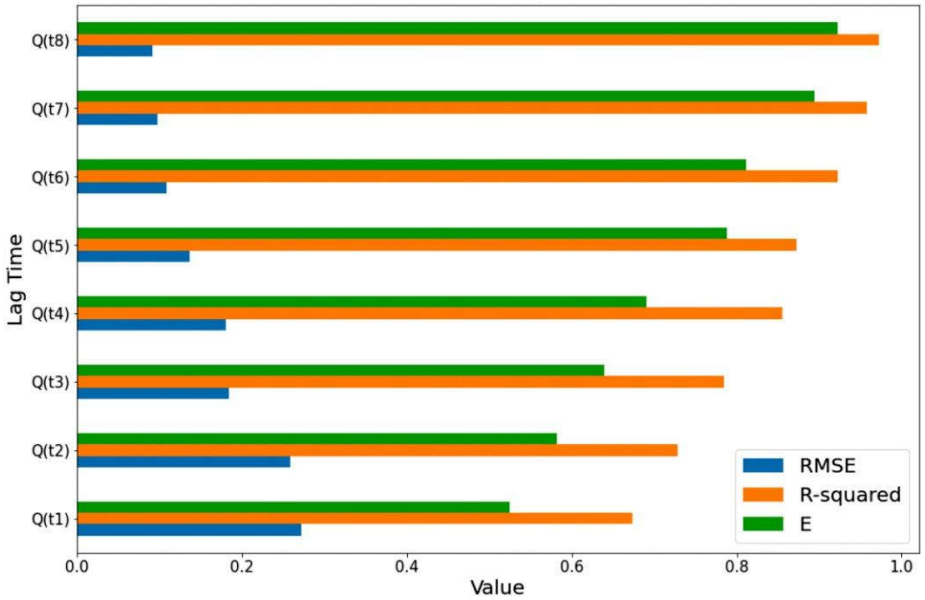

Figure 8. Error matrices for various lead times in the LSTM neural network model.

3.2. Model Evaluation Matrices and Improvement

The performance of the LSTM neural network is evaluated using three error matrices e.g., Root Mean Square Error (RMSE), the coefficient of determination (R^2) and Nash Sutcliffe model efficiency coefficient (E). The performance variation with the change in the lag time is represented in the **Error! Reference source not found.**. Further, the performance of the model also evaluated and improved through increasing the number of iterations i.e., epoch in the neural network. The value of RMSE is observed with the increase in the number of iterations/epochs in the LSTM neural network in the **Error! Reference source not found.**. The number of epochs is increased up to 100 to increase the performance and converge to a more stable narrow range of RMSE values. The RMSE value is found to decrease from 0.01 to 0.0025 which indicates satisfactory performance in the LSTM algorithm. The model performance increases significantly from the very beginning of the iteration for both the train and test scenarios. Changes in the RMSE values are obtained for both the train and test dataset marked in blue and orange color. Comparatively mild decrease in the RMSE value i.e., increase in the model performance can be observed between 20 and 100 epochs. Several local abrupt variations in the performance can also be identified after approximately 5 epochs. The trend of change in the decrease in the RMSE values for the models shows that approximately 100 epochs may yield the best performance.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

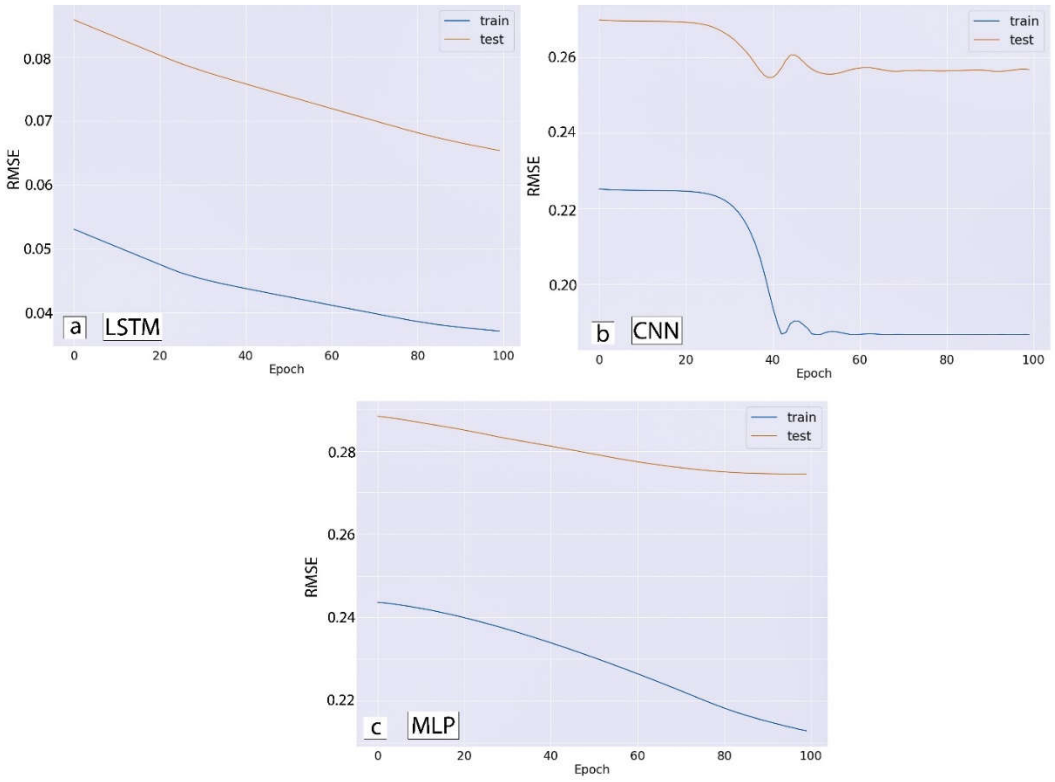

Figure 9. Model performance improvement with increase in the number of Epoch for both the train and test dataset. RMSE values decrease with inclusion of more epoch or iterations.

Observed and predicted values of the discharge from LSTM model and the distribution of the RMSE values are illustrated in the **Error! Reference source not found.** using a scatter plot (a) and histogram (b). The scatterplot shows that the points follow an approximately 45° trendline originated at zero. Some points are discovered to be outside of the main cluster of points, indicating a possible erroneous prediction in the procedure.. The R^2 value of the best fitted straight line is +0.79 which gives a good indication of the performance. In the **Error! Reference source not found.** (b), the distribution of the differences between the observed and predicted values is showed using a histogram. Majority of the difference values is located near zero i.e., lower prediction errors. The maximum and minimum of the difference values are -0.04 and 0.06 delineating an overall satisfactory performance of the LSTM model.

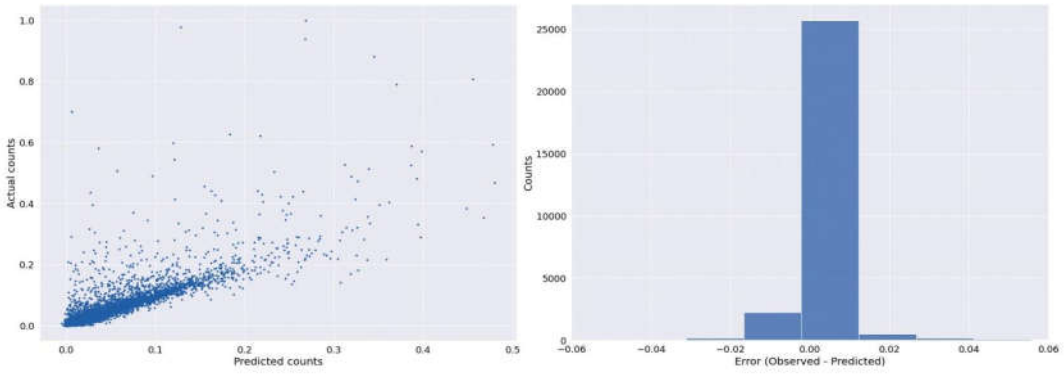

Figure 10. Model accuracies are presented through the scatterplot of the observed and predicted discharge values and the histogram of the distribution of the RMSE values.

3.3. Tuning Hyperparameters

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2022 doi:10.20944/preprints202209.0398.v1

Tuning the hyperparameters of LSTM can be intimidating as there is no simple and robust hypothesis to perform the task to optimize the model [78]. To conduct hyper parameter tuning for LSTM algorithms, a systematic approach should be undertaken to perceive the dynamical and stochastic characteristics of the process [79]. In this study, LSTM neural network is applied on a discharge time series data. The performance of the model is further improved through tuning the size of epoch and batch and the number of neurons for the neural network stochastic procedure. The tuned parameters are only applicable for the discharge time series used in this study as the stochastic nature of the neural network optimization.

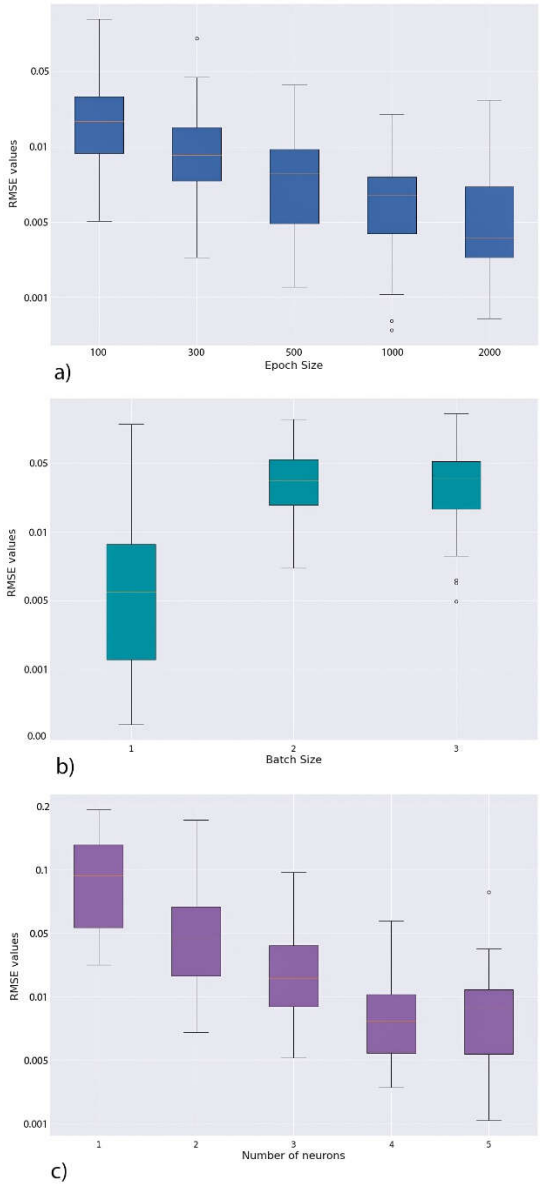

Figure 11. Change in the RMSE values with the increase in the number of batch size, epoch size and neurons.

Epoch size is tuned first for constant batch sizes of 4 and a single neuron. The number of epochs is considered to observe the variation in the performance are 100, 300, 500, 1000 and 2000. Similarly, the number of batch sizes of 1, 2 and 4 and neurons of 1-5 are considered keeping the epoch size constant in 2000 to see the model improvement. Epoch size of 2000 is selected to further the tuning task with the batch size and the number of neurons as it contributes to the lowest RMSE value. A batch size of 1 and 4 neurons provides further lower RMSE values which can be seen in the ***Error! Reference source not found.***. Therefore, an epoch size of 2000, batch size of 1 and number of neurons of 4 are selected

 doi:10.20944/preprints202209.0398.v1

as the best hyperparameter combination for LSTM algorithm to maximize the model performance.

4. Conclusions

Time series prediction for river flow is a pivotal task in the field of water resource management. Discharge is the most important parameter in various aspects of water resource engineering and management e.g., flood and irrigation warning system. Several approaches through physics-based numerical modelling techniques are proven inefficient in terms of real-time forecasting and assumptions. On the contrary, the application of the data-driven prediction models is highly efficacious in predicting various hydrological variables without taking complicated equations and assumptions into consideration. In this study, river discharge is predicted using the most powerful neural network in predicting sequential data i.e., LSTM RNN. LSTM algorithm can recall both the short- and long-term pattern of the time series to forecast. The range of the discharge time series considered in this research is quite large containing multiple seasonal dynamics and climatic variations. Traditional physics-based numerical modelling tool requires assumptions, other correlated variables, and expensive calibration of the parameters. Compared to the other neural network regression models, LSTM are proved to show good performance especially the time series prediction. As the river flow provides a sequential data which has high temporal dynamics, LSTM is used to quantify future values based on the past data. As the shape of the discharge dataset is comparatively large containing 29,392 observations datapoints from July 1941 to December 2021, LSTM algorithms showed highly satisfactory performance with longer lead periods.

This study contributes to a reproducible template to investigate the uniqueness of the temporal dynamics of river discharge through extensive EDA. Hidden pattern of the distribution of discharge values through over 80 years of data is discovered various up-to-date data exploration tools which is a mandatory requirement for the satisfactory training of LSTM algorithm. After a successful training step, LSTM is tuned and optimized through an explicit iterative performance record which can further be transferred to forecast discharge value in identical geographical location. The performance of the LSTM algorithm in predicting the river discharge illustrates the algorithm is highly suitable to the discharge time series. Several error matrices show promising performance of the with minimum error compared to the other DNN approaches.

It should also be considered that there are couple of drawbacks of using LSTM along the advantages which are: 1) the time consumption on training the model, which LSTM analysis required more time and running the model over the large dataset can take longer than other conceptual models. In this study, the computational effort/time required for the LSTM algorithm was found to be considerably high. 2) The low-speed process also resulted in requirement of using more memory and storage potential of the system, which again can cause a challenge for training on a large dataset likewise in this research. 3) Complex process of LSTM for time series data has a high potential of facing overfitting challenge and resulted in inaccurate extremely low error measures. 4) Although, the ability to compare the difference lead time throughout the entire time series dataset is key point of implementation of LSTM, we were not able to get a satisfactory performance for more than three-day period for this dataset. Any lead times within three-day period works perfect and resulted in satisfying low error measures. Future research works should be conducted to incorporate high performance computing and cloud-based operations to obtain the best result. LSTM models with different configurations should also be applied in different geographical and climatic locations to investigate the transferability of the model

Author Contributions: Conceptualization, M.A.A.M. and M.K.; methodology, M.A.A.M. and M.K.; formal analysis, M.A.A.M., M.K. and M.M.S.Y.; investigation, M.A.A.M.; writing—original draft preparation, M.A.A.M., M.K. and M.M.S.Y.; writing—review and editing, M.A.A.M., M.K. and M.M.S.Y.; visualization, M.A.A.M.; supervision, M.A.A.M., M.K. H.S.; project administration,

M.A.A.M., M.K. H.S.; funding acquisition, M.K. H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Computer Science program in the School of Computing and Analytics at Northern Kentucky University.

Data Availability Statement: Data collected for the study can be made available upon request from the corresponding author.

Acknowledgments: This study was supported by the Computer Science program in the School of Computing and Analytics at Northern Kentucky University. The authors would also like to thank Dr. Hanieh Shabanian for her support and guidance immensely.

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

References

1. Md, B.; Hossain, T.A.; Ahmed, T.; Aktar, M.N.; Fida, M.; Khan, A.; Islam, A.K.M.S.; Md, M.; Yazdan, S.; Noor, F.; et al. 5.
2. Hossain, B.M.T.A.; Ahmed, T.; Aktar, N.; Khan, F.; Islam, A.; Yazdan, M.M.S.; Noor, F.; Rahaman, A. *Climate Change Impacts on Water Availability in the Meghna Basin*; 2015;
3. Kao, I.-F.; Zhou, Y.; Chang, L.-C.; Chang, F.-J. Exploring a Long Short-Term Memory Based Encoder-Decoder Framework for Multi-Step-Ahead Flood Forecasting. *J. Hydrol.* **2020**, *583*, 124631, doi:10.1016/j.jhydrol.2020.124631.
4. Song, X.; Liu, Y.; Xue, L.; Wang, J.; Zhang, J.; Wang, J.; Jiang, L.; Cheng, Z. Time-Series Well Performance Prediction Based on Long Short-Term Memory (LSTM) Neural Network Model. **2020**, doi:10.1016/j.petrol.2019.106682.
5. Le, X.-H.; Ho, H.V.; Lee, G.; Jung, S. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* **2019**, *11*, 1387, doi:10.3390/w11071387.
6. Wang, F.; Chen, Y.; Li, Z.; Fang, G.; Li, Y.; Wang, X.; Zhang, X.; Kayumba, P.M. Developing a Long Short-Term Memory (LSTM)-Based Model for Reconstructing Terrestrial Water Storage Variations from 1982 to 2016 in the Tarim River Basin, Northwest China. *Remote Sens.* **2021**, *13*, 889, doi:10.3390/rs13050889.
7. Ma, B.; Pang, W.; Lou, Y.; Mei, X.; Wang, J.; Gu, J.; Dai, Z. Impacts of River Engineering on Multi-Decadal Water Discharge of the Mega-Changjiang River. *Sustainability* **2020**, *12*, 8060, doi:10.3390/su12198060.
8. Bouwer, H. Integrated Water Management: Emerging Issues and Challenges. *Agric. Water Manag.* **2000**, *45*, 217–228, doi:10.1016/S0378-3774(00)00092-5.
9. Evans, R.G.; Sadler, E.J. Methods and Technologies to Improve Efficiency of Water Use. *Water Resour. Res.* **2008**, *44*, doi:10.1029/2007WR006200.
10. Sophocleous, M. Groundwater Recharge and Sustainability in the High Plains Aquifer in Kansas, USA. *Hydrogeol. J.* **2005**, *13*, 351–365, doi:10.1007/s10040-004-0385-6.
11. 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, doi:10.1016/j.jhydrol.2018.04.065.
12. Kisi, O.; Cimen, M. A Wavelet-Support Vector Machine Conjunction Model for Monthly Streamflow Forecasting. *J. Hydrol.* **2011**, *399*, 132–140, doi:10.1016/j.jhydrol.2010.12.041.
13. ZhongMin, L.; ZhangLing, X.; Jun, W.; Long, S.; BinQuan, L.; YiMing, H.; YaQi, W. An Improved Chaos Similarity Model for Hydrological Forecasting. *J. Hydrol. Amst.* **2019**, *577*.
14. Kilsdonk, R.A.H.; Bomers, A.; Wijnberg, K.M. Predicting Urban Flooding Due to Extreme Precipitation Using a Long Short-Term Memory Neural Network. *Hydrology* **2022**, *9*, 105, doi:10.3390/hydrology9060105.
15. Ayzel, G.; Kurochkina, L.; Abramov, D.; Zhuravlev, S. Development of a Regional Gridded Runoff Dataset Using Long Short-Term Memory (LSTM) Networks. *Hydrology* **2021**, *8*, 6, doi:10.3390/hydrology8010006.
16. Bai, Y.; Bezak, N.; Sapač, K.; Klun, M.; Zhang, J. Short-Term Streamflow Forecasting Using the Feature-Enhanced Regression Model. *Water Resour. Manag. Int. J. Publ. Eur. Water Resour. Assoc. EWRA* **2019**, *33*, 4783–4797.
17. Xiao, Z.; Liang, Z.; Li, B.; Hou, B.; Hu, Y.; Wang, J. New Flood Early Warning and Forecasting Method Based on Similarity Theory. *J. Hydrol. Eng.* **2019**, *24*, 04019023, doi:10.1061/(ASCE)HE.1943-5584.0001811.
18. Milly, P.C.D.; Dunne, K.A.; Vecchia, A.V. Global Pattern of Trends in Streamflow and Water Availability in a Changing Climate. *Nature* **2005**, *438*, 347–350, doi:10.1038/nature04312.
19. Chang, L.-C.; Liou, J.-Y.; Chang, F.-J. Spatial-Temporal Flood Inundation Nowcasts by Fusing Machine Learning Methods and Principal Component Analysis. *J. Hydrol.* **2022**, *612*, 128086, doi:10.1016/j.jhydrol.2022.128086.
20. Devia, G.K.; Ganasri, B.P.; Dwarakish, G.S. A Review on Hydrological Models. *Aquat. Procedia* **2015**, *4*, 1001–1007, doi:10.1016/j.aqpro.2015.02.126.
21. Askarizadeh, A.; Rippy, M.A.; Fletcher, T.D.; Feldman, D.L.; Peng, J.; Bowler, P.; Mehring, A.S.; Winfrey, B.K.; Vrugt, J.A.; AghaKouchak, A.; et al. From Rain Tanks to Catchments: Use of Low-Impact Development To Address Hydrologic Symptoms of the Urban Stream Syndrome. *Environ. Sci. Technol.* **2015**, *49*, 11264–11280, doi:10.1021/acs.est.5b01635.
22. Zhao, J.; Xu, J.; Xie, X.; Lu, H. Drought Monitoring Based on TIGGE and Distributed Hydrological Model in Huaihe River Basin, China. *Sci. Total Environ.* **2016**, *553*, 358–365, doi:10.1016/j.scitotenv.2016.02.115.

23. Humphrey, G.B.; Gibbs, M.S.; Dandy, G.C.; Maier, H.R. A Hybrid Approach to Monthly Streamflow Forecasting: Integrating Hydrological Model Outputs into a Bayesian Artificial Neural Network. *J. Hydrol.* **2016**, *540*, 623–640, doi:10.1016/j.jhydrol.2016.06.026.
24. Mosavi, A.; Ozturk, P.; Chau, K. Flood Prediction Using Machine Learning Models: Literature Review. *Water* **2018**, *10*, 1536, doi:10.3390/w10111536.
25. Costabile, P.; Macchione, F. Enhancing River Model Set-up for 2-D Dynamic Flood Modelling. *Environ. Model. Softw.* **2015**, *67*, 89–107, doi:10.1016/j.envsoft.2015.01.009.
26. Cheng, M.; Fang, F.; Kinouchi, T.; Navon, I.M.; Pain, C.C. Long Lead-Time Daily and Monthly Streamflow Forecasting Using Machine Learning Methods. *J. Hydrol.* **2020**, *590*, 125376, doi:10.1016/j.jhydrol.2020.125376.
27. Alvisi, S.; Franchini, M. Fuzzy Neural Networks for Water Level and Discharge Forecasting with Uncertainty. *Environ. Model. Softw.* **2011**, *26*, 523–537, doi:10.1016/j.envsoft.2010.10.016.
28. Prasad, R.; Deo, R.C.; Yan, L.; Maraseni, T. Input Selection and Performance Optimization of ANN-Based Streamflow Forecasts in the Drought-Prone Murray Darling Basin Region Using IIS and MODWT Algorithm. *Atmospheric Res.* **2017**, *197*, 42–63.
29. Rathinasamy, M.; Adamowski, J.; Khosa, R. Multiscale Streamflow Forecasting Using a New Bayesian Model Average Based Ensemble Multi-Wavelet Volterra Nonlinear Method. *J. Hydrol.* **2013**, *507*, 186–200, doi:10.1016/j.jhydrol.2013.09.025.
30. Yaseen, Z.M.; El-shafie, A.; Jaafar, O.; Afan, H.A.; Sayl, K.N. Artificial Intelligence Based Models for Stream-Flow Forecasting: 2000–2015. *J. Hydrol.* **2015**, *530*, 829–844, doi:10.1016/j.jhydrol.2015.10.038.
31. Mehedi, M.A.A.; Hosseiny, H.; Smith, V.; Jiao, X. Unraveling Urban Fluvial Flooding Complexities through AI. **2021**, *2021*, H22G-15.
32. Poole, G.C.; Fogg, S.K.; O'Daniel, S.J.; Amerson, B.E.; Reinhold, A.M.; Carlson, S.P.; Mohr, E.J.; Oakland, H.C. Hyporheic Hydraulic Geometry: Conceptualizing Relationships among Hyporheic Exchange, Storage, and Water Age. *PLOS ONE* **2022**, *17*, e0262080, doi:10.1371/journal.pone.0262080.
33. Wohl, E. Time and the Rivers Flowing: Fluvial Geomorphology since 1960. *Geomorphology* **2014**, *216*, 263–282, doi:10.1016/j.geomorph.2014.04.012.
34. Myronidis, D.; Ioannou, K.; Fotakis, D.; Dörflinger, G. Streamflow and Hydrological Drought Trend Analysis and Forecasting in Cyprus. *Water Resour. Manag. Int. J. Publ. Eur. Water Resour. Assoc. EWRA* **2018**, *32*, 1759–1776.
35. Wang, W.; Chau, K.; Xu, D.; Chen, X.-Y. Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition. *Water Resour. Manag.* **2015**, *29*, 2655–2675, doi:10.1007/s11269-015-0962-6.
36. Long, J.; Sun, Z.; Pardalos, P.M.; Hong, Y.; Zhang, S.; Li, C. A Hybrid Multi-Objective Genetic Local Search Algorithm for the Prize-Collecting Vehicle Routing Problem. *Inf. Sci.* **2019**, *478*, 40–61, doi:10.1016/j.ins.2018.11.006.
37. Abdollahzadeh, M.; Khosravi, M.; Hajipour Khire Masjidi, B.; Samimi Behbahan, A.; Bagherzadeh, A.; Shahkar, A.; Tat Shahdost, F. Estimating the Density of Deep Eutectic Solvents Applying Supervised Machine Learning Techniques. *Sci. Rep.* **2022**, *12*, 4954, doi:10.1038/s41598-022-08842-5.
38. Chang Fi [Chang, F.J.; LiChiu, C.; ChienWei, H.; IFeng, K. Prediction of Monthly Regional Groundwater Levels through Hybrid Soft-Computing Techniques. *J. Hydrol. Amst.* **2016**, *541*, 965–976.
39. Daliakopoulos, I.N.; Coulibaly, P.; Tsanis, I.K. Groundwater Level Forecasting Using Artificial Neural Networks. *J. Hydrol.* **2005**, *309*, 229–240, doi:10.1016/j.jhydrol.2004.12.001.
40. Parchami-Araghi, F.; Mirlatifi, S.M.; Dashtaki, S.G.; Mahdian, M.H. Point Estimation of Soil Water Infiltration Process Using Artificial Neural Networks for Some Calcareous Soils. *J. Hydrol. Amst.* **2013**, *481*, 35–47.
41. Zhu, X.; Khosravi, M.; Vaferi, B.; Nait Amar, M.; Ghriga, M.A.; Mohammed, A.H. Application of Machine Learning Methods for Estimating and Comparing the Sulfur Dioxide Absorption Capacity of a Variety of Deep Eutectic Solvents. *J. Clean. Prod.* **2022**, *363*, 132465, doi:10.1016/j.jclepro.2022.132465.
42. Rozos, E.; Dimitriadis, P.; Mazi, K.; Koussis, A.D. A Multilayer Perceptron Model for Stochastic Synthesis. *Hydrology* **2021**, *8*, 67, doi:10.3390/hydrology8020067.
43. Elbeltagi, A.; Di Nunno, F.; Kushwaha, N.L.; de Marinis, G.; Granata, F. River Flow Rate Prediction in the Des Moines Watershed (Iowa, USA): A Machine Learning Approach. *Stoch. Environ. Res. Risk Assess.* **2022**, doi:10.1007/s00477-022-02228-9.
44. Belayneh, A.; Adamowski, J.; Khalil, B.; Ozga-Zielinski, B. Long-Term SPI Drought Forecasting in the Awash River Basin in Ethiopia Using Wavelet Neural Network and Wavelet Support Vector Regression Models. *J. Hydrol.* **2014**, *508*, 418–429, doi:10.1016/j.jhydrol.2013.10.052.
45. Mirzavand, M.; Ghazavi, R. A Stochastic Modelling Technique for Groundwater Level Forecasting in an Arid Environment Using Time Series Methods. *Water Resour. Manag.* **2015**, *29*, 1315–1328, doi:10.1007/s11269-014-0875-9.
46. 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, doi:10.1016/j.jhydrol.2010.11.002.
47. Khosravi, M.; Tabasi, S.; Hossam Eldien, H.; Motahari, M.R.; Alizadeh, S.M. Evaluation and Prediction of the Rock Static and Dynamic Parameters. *J. Appl. Geophys.* **2022**, *199*, 104581, doi:10.1016/j.jappgeo.2022.104581.
48. Karimi, M.; Khosravi, M.; Fathollahi, R.; Khandakar, A.; Vaferi, B. Determination of the Heat Capacity of Cellulosic Biosamples Employing Diverse Machine Learning Approaches. *Energy Sci. Eng.* **2022**, *10*, 1925–1939, doi:10.1002/ese3.1155.
49. Jothiprakash, V.; Kote, A.S. Effect of Pruning and Smoothing While Using M5 Model Tree Technique for Reservoir Inflow Prediction. *J. Hydrol. Eng.* **2011**, *16*, 563–574, doi:10.1061/(ASCE)HE.1943-5584.0000342.
50. Khosravi, M.; Arif, S.B.; Ghaseminejad, A.; Tohidi, H.; Shabanian, H. Performance Evaluation of Machine Learning Regressors for Estimating Real Estate House Prices 2022.

51. Allawi, M.F.; Jaafar, O.; Mohamad Hamzah, F.; Mohd, N.S.; Deo, R.C.; El-Shafie, A. Reservoir Inflow Forecasting with a Modified Coactive Neuro-Fuzzy Inference System: A Case Study for a Semi-Arid Region. *Theor. Appl. Climatol.* **2018**, *134*, 545–563, doi:10.1007/s00704-017-2292-5.
52. Xu, X.; Zhang, X.; Fang, H.; Lai, R.; Zhang, Y.; Huang, L.; Liu, X. A Real-Time Probabilistic Channel Flood-Forecasting Model Based on the Bayesian Particle Filter Approach. *Environ. Model. Softw.* **2017**, *88*, 151–167, doi:10.1016/j.envsoft.2016.11.010.
53. 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, doi:10.1016/j.jhydrol.2018.04.065.
54. Bai, Y.; Xie, J.; Wang, X.; Li, C. Model Fusion Approach for Monthly Reservoir Inflow Forecasting. *J. Hydroinformatics* **2016**, *18*, 634–650, doi:10.2166/hydro.2016.141.
55. Sahoo, S.; Jha, M.K. Groundwater-Level Prediction Using Multiple Linear Regression and Artificial Neural Network Techniques: A Comparative Assessment. *Hydrogeol. J.* **2013**, *21*, 1865–1887, doi:10.1007/s10040-013-1029-5.
56. Mehedi, M.A.A.; Reichert, N.; Molkenthin, F. *SENSITIVITY ANALYSIS OF HYPORHEIC EXCHANGE TO SMALL SCALE CHANGES IN GRAVEL-SAND FLUMEBED USING A COUPLED GROUNDWATER-SURFACE WATER MODEL*; 2020;
57. Karandish, F.; Šimůnek, J. A Comparison of Numerical and Machine-Learning Modeling of Soil Water Content with Limited Input Data. *J. Hydrol.* **2016**, 19.
58. Mohanty, S.; Jha, M.K.; Kumar, A.; Panda, D.K. Comparative Evaluation of Numerical Model and Artificial Neural Network for Simulating Groundwater Flow in Kathajodi–Surua Inter-Basin of Odisha, India. *J. Hydrol.* **2013**, *495*, 38–51, doi:10.1016/j.jhydrol.2013.04.041.
59. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780, doi:10.1162/neco.1997.9.8.1735.
60. Hu, R.; Fang, F.; Pain, C.C.; Navon, I.M. Rapid Spatio-Temporal Flood Prediction and Uncertainty Quantification Using a Deep Learning Method. *J. Hydrol.* **2019**, *575*, 911–920, doi:10.1016/j.jhydrol.2019.05.087.
61. Shin, M.-J.; Moon, S.-H.; Kang, K.G.; Moon, D.-C.; Koh, H.-J. Analysis of Groundwater Level Variations Caused by the Changes in Groundwater Withdrawals Using Long Short-Term Memory Network. *Hydrology* **2020**, *7*, 64, doi:10.3390/hydrology7030064.
62. Granata, F.; Di Nunno, F.; de Marinis, G. Stacked Machine Learning Algorithms and Bidirectional Long Short-Term Memory Networks for Multi-Step Ahead Streamflow Forecasting: A Comparative Study. *J. Hydrol.* **2022**, *613*, 128431, doi:10.1016/j.jhydrol.2022.128431.
63. Kao, I.-F.; Liou, J.-Y.; Lee, M.-H.; Chang, F.-J. Fusing Stacked Autoencoder and Long Short-Term Memory for Regional Multi-step-Ahead Flood Inundation Forecasts. *J. Hydrol.* **2021**, *598*, 126371, doi:10.1016/j.jhydrol.2021.126371.
64. Younger, A.S.; Hochreiter, S.; Conwell, P.R. Meta-Learning with Backpropagation. In Proceedings of the IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222); July 2001; Vol. 3, pp. 2001–2006 vol.3.
65. Mouatadid, S.; Adamowski, J.F.; Tiwari, M.K.; Quilty, J.M. Coupling the Maximum Overlap Discrete Wavelet Transform and Long Short-Term Memory Networks for Irrigation Flow Forecasting. *Agric. Water Manag.* **2019**, *219*, 72–85.
66. Kratzert, F.; Klotz, D.; Brenner, C.; Schulz, K.; Herrnegger, M. Rainfall–Runoff Modelling Using Long Short-Term Memory (LSTM) Networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022, doi:10.5194/hess-22-6005-2018.
67. Ni, L.; Wang, D.; Singh, V.P.; Wu, J.; Wang, Y.; Tao, Y.; Zhang, J. Streamflow and Rainfall Forecasting by Two Long Short-Term Memory-Based Models. *J. Hydrol.* **2020**, *583*, 124296, doi:10.1016/j.jhydrol.2019.124296.
68. 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, doi:10.3390/w10111543.
69. Mehedi, M.A.A.; Amur, A.; McGauley, M.; Metcalf, J.; Wadzuk, B.; Smith, V. Quantifying the Benefits of AI vs. Numerical Modeling for Urban Green Stormwater Infrastructure. **2021**, *2021*, H45N-1330.
70. Kilinc, H.C.; Haznedar, B. A Hybrid Model for Streamflow Forecasting in the Basin of Euphrates. *Water* **2022**, *14*, 80, doi:10.3390/w14010080.
71. Xayasouk, T.; Lee, H.; Lee, G. Air Pollution Prediction Using Long Short-Term Memory (LSTM) and Deep Autoencoder (DAE) Models. *Sustainability* **2020**, *12*, 2570, doi:10.3390/su12062570.
72. Rozos, E.; Dimitriadis, P.; Bellos, V. Machine Learning in Assessing the Performance of Hydrological Models. *Hydrology* **2022**, *9*, 5, doi:10.3390/hydrology9010005.
73. Staudemeyer, R.C.; Morris, E.R. *Understanding LSTM -- a Tutorial into Long Short-Term Memory Recurrent Neural Networks*; arXiv, 2019;
74. Tsang, G.; Deng, J.; Xie, X. Recurrent Neural Networks for Financial Time-Series Modelling. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR); August 2018; pp. 892–897.
75. Maulik, R.; Egele, R.; Lusch, B.; Balaprakash, P. Recurrent Neural Network Architecture Search for Geophysical Emulation. In Proceedings of the Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis; IEEE Press: Atlanta, Georgia, November 9 2020; pp. 1–14.
76. Gupta, H.V.; Kling, H. On Typical Range, Sensitivity, and Normalization of Mean Squared Error and Nash-Sutcliffe Efficiency Type Metrics. *Water Resour. Res.* **2011**, *47*, doi:10.1029/2011WR010962.
77. Willmott, C.J.; Robeson, S.M.; Matsuura, K. A Refined Index of Model Performance. *Int. J. Climatol.* **2012**, *32*, 2088–2094, doi:10.1002/joc.2419.
78. Hossain, M.D.; Ochiai, H.; Fall, D.; Kadobayashi, Y. LSTM-Based Network Attack Detection: Performance Comparison by Hyper-Parameter Values Tuning. In Proceedings of the 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom); August 2020; pp. 62–69.

 doi:10.20944/preprints202209.0398.v1

79. Gorgolis, N.; Hatzilygeroudis, I.; Istenes, Z.; Gyenne, L.– G. Hyperparameter Optimization of LSTM Network Models through Genetic Algorithm. In Proceedings of the 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA); July 2019; pp. 1–4.