


Article

Water Level Forecasting Using Spatiotemporal Attention-Based Long Short-Term Memory Network

Fahima Noor ¹, Sanaulla Haq ¹, Mohammed Rakib ¹, Tarik Ahmed ¹, Zeeshan Jamal ¹, Zakaria Shams Siam ^{1,2}, Rubyat Tasnuva Hasan ¹, Mohammed Sarfaraz Gani Adnan ^{3,4,*}, Ashraf Dewan ⁵ and Rashedur M. Rahman ¹

¹ Department of Electrical and Computer Engineering, North South University, Plot-15, Block-B, Bashundhara Residential Area, Dhaka 1229, Bangladesh; fahima.noor@northsouth.edu (F.N.); sanaulla.haq@northsouth.edu (S.H.); mohammed.rakib@northsouth.edu (M.R.); tarik.ahmed@northsouth.edu (T.A.); zeeshan.jamal@northsouth.edu (Z.J.); zakaria.siam@northsouth.edu (Z.S.S.); rubyat.tasnuva@northsouth.edu (R.T.H.); rashedur.rahman@northsouth.edu (R.M.R.)

² Department of Electrical and Computer Engineering, Presidency University, Dhaka 1212, Bangladesh

³ Department of Urban and Regional Planning, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh

⁴ Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, UK

⁵ School of Earth and Planetary Sciences, Curtin University, Perth, WA 6102, Australia; a.dewan@curtin.edu.au

* Correspondence: sarfarazadnan@cuet.ac.bd



Citation: Noor, F.; Haq, S.; Rakib, M.; Ahmed, T.; Jamal, Z.; Siam, Z.S.; Hasan, R.T.; Adnan, M.S.G.; Dewan, A.; Rahman, R.M. Water Level Forecasting Using Spatiotemporal Attention-Based Long Short-Term Memory Network. *Water* **2022**, *14*, 612. <https://doi.org/10.3390/w14040612>

Academic Editors: Celestine Iwendu and Thippa Reddy Gadekallu

Received: 10 January 2022

Accepted: 15 February 2022

Published: 17 February 2022

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



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

Abstract: Bangladesh is in the floodplains of the Ganges, Brahmaputra, and Meghna River delta, crisscrossed by an intricate web of rivers. Although the country is highly prone to flooding, the use of state-of-the-art deep learning models in predicting river water levels to aid flood forecasting is underexplored. Deep learning and attention-based models have shown high potential for accurately forecasting floods over space and time. The present study aims to develop a long short-term memory (LSTM) network and its attention-based architectures to predict flood water levels in the rivers of Bangladesh. The models developed in this study incorporated gauge-based water level data over 7 days for flood prediction at Dhaka and Sylhet stations. This study developed five models: artificial neural network (ANN), LSTM, spatial attention LSTM (SALSTM), temporal attention LSTM (TALSTM), and spatiotemporal attention LSTM (STALSTM). The multiple imputation by chained equations (MICE) method was applied to address missing data in the time series analysis. The results showed that the use of both spatial and temporal attention together increases the predictive performance of the LSTM model, which outperforms other attention-based LSTM models. The STALSTM-based flood forecasting system, developed in this study, could inform flood management plans to accurately predict floods in Bangladesh and elsewhere.

Keywords: deep learning; time series; flood forecasting; attention mechanism LSTM; water-level prediction

1. Introduction

Bangladesh is one of the most vulnerable countries to climate-change-induced stresses because of its geographical position [1–3]. Among all natural hazards, it is highly prone to flooding [4,5]. The country is in the floodplains of the Ganges, the Brahmaputra, and the Meghna (GBM) River systems, making it highly susceptible to flooding of various types and magnitudes [6–11]. Around two-thirds of the country has elevations <5 m above mean sea level [12]. Every year, it receives copious rainfall during the monsoon season and high river discharge along with intense rainfall results in floods of varying magnitudes [6], which are recurrent and lead to significant damages to lives and properties [8,9,12–14]. The Bangladesh Water Development Board (BWDB) is primarily responsible for providing

flood warnings [15]. The Flood Forecasting and Warning Centre (FFWC), developed in 1972 under BWDB, creates flood forecasting and warning systems as part of nonstructural intervention [16]. Bangladesh experienced extreme flood events during 1987, 1988, 1998, 2004, and 2007 that submerged more than half of the country [1,4–6]. Flood vulnerability in Bangladesh is increasing due to changes in the natural environment and anthropogenic forcing, as well as climate change [5,17].

Flooding causes significant damage to the national economy, as it results in a loss of approximately 1.5% of gross domestic product (GDP) each year. The flood events of 1988 and 1998 were extremely devastating, affecting more than 60% of Bangladesh and resulting in losses of more than 8% of GDP. The recurring floods killed 15,033 individuals between 1972 and 2013, averaging 358 deaths per year [18]. Consequently, there is a great need to design an accurate flood forecasting system that could enhance flood preparedness for saving lives and properties [19–21]. Since the late 1980s, the government has taken structural and nonstructural mitigation and adaptation measures to limit damage and protect the natural and anthropogenic environments [19,22]. Structural mitigation measures alter the physical environment of the location to prevent disaster. The most common approaches include the use of dams and embankment and drainage channel improvement [23]. However, structural flood protection has caused more frequent rainfall-induced flooding and has promoted potential damage during extreme events [24]. Contrarily, nonstructural measures have proven to be more effective in adapting to floods [7,23]. Such measures do not rely on altering physical space, and rather aim to influence human behavior to abate disaster losses. Nonstructural mitigation options include flood forecasting, community awareness, and environmental control, such as increasing vegetation to prevent the land from erosion [7]. These also include tools such as early warning systems, which can also aid in disaster preparedness [22,25]. Early warning systems provide more lead time to flood-affected people so that they can move to safer places [4,25]. An accurate flood forecasting system is thus considered to be the most prioritized and effective nonstructural mitigation measure [16].

Several studies have explored various ways of minimizing flooding and its associated risks [4,18,26–29]. One of them is the design principle of forecasting, which can be divided into two groups: physical models and data-driven models. Hydrological models are more concerned with the physical aspects of flooding. These models use a deterministic approach by exploiting mathematical equations to define the relationship between the input and output variables [26–29]. However, such deterministic models are subject to considerable uncertainties [30].

On the other hand, the data-driven model utilizes a probabilistic approach using historical data to make predictions [31–37]. Unlike physical models, the latter model analyzes the relationship between the input and output variables from observed data. However, these models are data intensive and require a large number of in situ observations to make predictions reliable [38,39]. The outcomes of both probabilistic and deterministic models are, however, influenced by the choice of flood influencing factors, such as topography.

Recent advancement of deep learning models has enabled researchers to perform quick and accurate prediction of flooding [16,31–40]. These models, including long short-term memory (LSTM), are introduced in sequence learning, which has proved to be powerful for time series analysis [38,39,41]. The LSTM is a recurrent neural network that can retain useful information from sequential data to make future predictions.

As flood in Bangladesh is governed by major rivers and their tributaries, a robust model is required to predict river water levels [9,11,16,31–34,42]. The complexity of river networks in Bangladesh makes flood prediction a challenging task. The dynamic nature of rivers frequently changes their courses, causing bank erosions [43]. Hence, an effective neural network model can discern the complexity of river networks [7,16,38,41].

Neural network models are effective in situations where the relationships between the independent and dependent variables are difficult to establish [9,16]. Such data-driven models enable researchers to predict future events, such as floods [16,38]. These models can provide improved real-time flood forecasting. Artificial neural network (ANN) and

other traditional machine learning models are, however, not robust enough to capture the entire relationship between the input and output features [9,16]. In search of an accurate forecasting model, an attempt is made in the present study for time series flood forecasting in Bangladesh using the attention-based LSTM model. We further compared the predictive performance of different LSTM architectures and the ANN model.

The contribution of this study is as follows:

1. To the best of our knowledge, this is the first attempt to compare the performance of traditional backpropagation and deep neural network techniques on real-time BWDB data for river water level forecasting.
2. This is also the first attempt where traditional LSTM and different attention-based models are being proposed and implemented to perform the water-level forecasting of a complex river system.

This article is organized as follows: In Section 2, we provide an overview of water-level prediction using deep learning works that are being conducted in Bangladesh as well as in other countries. In Section 3, study area, data used, models developed, and model evaluation criteria are described. Section 4 presents the results of this study and their critical discussion. Finally, in Section 5, we provide concluding remarks with general findings and an understanding of the work.

2. Literature Review

Flood forecasting in Bangladesh plays a crucial role in mitigating flood damage. The Flood Forecasting and Warning Centre (FFWC) uses traditional hydrological models, such as MIKE 11, to issue warning against flooding. The Mike 11 model simulates deterministic water levels and discharges in rivers for up to about 48 to 72 h [37]. The experimental model produces 1 to 10 days' probabilistic discharge forecasting. It requires a specialized skill to execute the model. However, river water levels in Bangladesh are influenced by complex and nonlinear interactions of hydro-climatological and hydrogeomorphic factors [16,19,22,23,25–29,44,45]. Hydrological models are both data and computing intensive, which can cause calibration difficulty in case of data unavailability. Thus, the probabilistic nature causes limitations in predicting and interpreting early warning messages [16,31–35]. The complexities in predicting floods using physical models have led many researchers to use machine learning.

Interventional studies involving animals or humans, and other studies that require ethical approval, must list the authority that provided approval and the corresponding ethical approval code and deep learning models, enabling learning of complex and non-linear relationships without making hypotheses about the pattern of the relationships. ANN is one of the most widely used flood forecasting models in the world [4,8,16]. ANNs are perceived to be conclusively valuable for modeling time series hydrologic problems as their architecture allows us to learn key information from a collection of inputs [9,16]. Neural network models are becoming more favorable for flood forecasting as, unlike linear regression, moving average (MA), and autoregressive integrated moving average (ARIMA), they can handle nonlinearity and nonstationary features. As such, more researchers are improving data-driven models for improving flood prediction accuracy.

Islam [31] used an ANN model for water level forecasting in Dhaka City. The ANN model was trained using data from 1998 to 2004, and validated with data from 2005 to 2007. Similar studies have also been conducted in the Sylhet District by applying ANN to predict the peak flow of the Surma River, where an ANN was able to identify nonlinear relationships between two different hydrological data series [7,8]. Liong, Lim, and Paudyal [9] utilized ANN for water level prediction in Dhaka for a lead time of 7 days with accuracy. Biswas and Jayawardena [32] predicted water level in the Surma River, Bangladesh, using ANN. Siddiquee and Hossain [16] designed ANN to predict the river water levels in different parts of Bangladesh. Several other studies used random forest (RF) and support vector machine (SVM) along with fuzzy logic to predict floods [9,35,37].

SVM is also used to analyze the water level in the Dhaka station [35]. SVM, ANN, hybrid wavelet-ANN (W-ANN), and RF have been employed on the daily water flow of the Punarbhaba River for making predictions [37]. However, other than ANN and convolutional neural network (CNN) models, not many diverse and robust deep learning models have been implemented for water level prediction in countries such as Bangladesh, which are highly prone to flooding [16].

Globally, in different regions, the LSTM network has been used to design flood prediction models. Recent studies have shown LSTM outperforming the ANN model for flood forecasting [36–39]. Thus, the high accuracy of LSTM has attracted researchers to use this model in a variety of time series predictions, such as predicting electricity prices and sales [46,47].

Similar studies have been conducted in different regions worldwide. A local spatial sequential long short-term memory (LSS-LSTM) network was used for flood susceptibility mapping in Shangyou County, China [39]. Wu et al. used a reduced order model (ROM) with the LSTM to create an LSTM-ROM model for flood forecasting by representing the spatiotemporal distribution of floods. LSTM models have also been used in transportation studies, where accurate prediction of traffic volume is a challenging task. Stacked bidirectional and unidirectional LSTM models are proposed by using a spatiotemporal dataset for traffic prediction [48]. These studies delineate the beneficial use of the LSTM in the field of flood forecasting. Despite the higher prediction accuracy of LSTM models, the use of such models in river-water-level forecasting in flood-prone areas is still underexplored.

After the discovery of the attention-based mechanism [49], many studies show that the predictive performance of the LSTM model greatly improves after the implementation of the attention-based mechanism [38,39,49–54]. Attention mechanism has been deployed in various tasks such as language translation [50] and image processing tasks such as image captioning [51,52]. Such a mechanism has been used to overcome the inability to accurately predict human action [53]. The attention-based mechanism has also been used with CNN to improve image labeling [51] and traffic prediction as it can overcome and interpret the nonlinearity and complexity of the spatiotemporal pattern of the problems [54]. Therefore, attention can be used to improve the accuracy of multivariate flood forecasting.

3. Materials and Methods

Figure 1 presents an overview of the methodological approach utilized in this study. First, we selected appropriate gauges for developing a multivariate water level prediction model. The acquired water level dataset, however, contains gaps. Thus, an imputation technique is applied to replace the missing data with some substitute value. The processed dataset is split into training and testing sets. The first 80% of the dataset is taken as the training set, and the remaining 20% is used to evaluate the performance of the model. Using the normalized training data, we established five neural network models: ANN, LSTM, spatial attention LSTM (SALSTM), temporal attention LSTM (TALSTM), and spatiotemporal attention LSTM (STALSTM). The performance of these models was assessed using various evaluation indices.

All neural network models had the same structures and were trained under the same hyperparameters, which helped to compare the models. Using the grid search technique, the optimum dimension of the hidden layer for the models was chosen to be 300. The Adam algorithm was chosen as the optimizer, and the mean square error (MSE) was used as the loss function for all the predictive models. The batch size was set at 600, with the epoch set at 100 and the learning rate at 0.1.

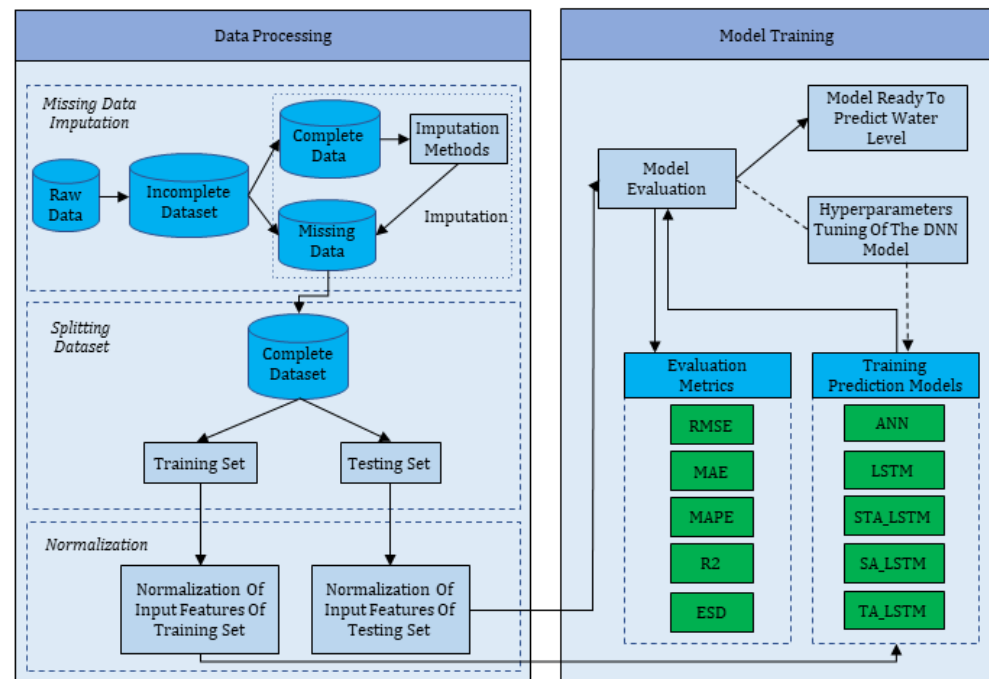


Figure 1. Water level forecasting stations in the study area.

3.1. Study Area

In this study, water level forecasting was carried out in two major cities of Bangladesh: Dhaka and Sylhet [7–9,14,16]. These two case study regions were selected as they are highly vulnerable to riverine flooding [55,56]. Besides, forecasting model structure and time series water level data are available for all the stations relative to other regions of the country.

Dhaka, the capital, is located in the central part of Bangladesh and at the confluence of three major rivers: the Brahmaputra, the Meghna, and the Ganges. The geographical location of the city makes it extremely susceptible to flooding [9,14].

Sylhet, located in the northeastern part of Bangladesh, is characterized by the country's longest river network (i.e., the Surma–Meghna River System) [7,8]. The geographical structure and the dense river network have made it extremely susceptible to flooding during the monsoon months when the rivers receive a heavy discharge from the upstream hilly regions. The locations of these stations are shown in Figure 2.

3.2. Dataset

This study used daily water level data from May 1985 to October 2008 at the selected stations collected from the Bangladesh Water Development Board (BWDB) [15]. For each of the two regions, several station data were considered independent variables to establish multivariate models. Since Bangladesh is a country filled with intricate networks of rivers, it is difficult to determine which river networks have a strong influence on Dhaka's water level without having the necessary domain knowledge. However, a study by Liong, Lim, and Paudyal [9] used stations near the borders of the country to forecast the water level of the Dhaka station as 90% of the annual water in all the major stations flows from outside the country. For this study, we selected the same stations to predict the water level of the Dhaka station.

For Sylhet, no such multivariate studies have been conducted from which we could select appropriate stations to utilize as input. Since the Sylhet station is one of the few stations that monitor the water level of Bangladesh's largest river system, the Surma–Meghna Rivers, we took all the existing stations as independent features. Moreover, a dense network of hydrological features is recommended to be used in order to take the spatial and temporal variation of the stations into account [39].

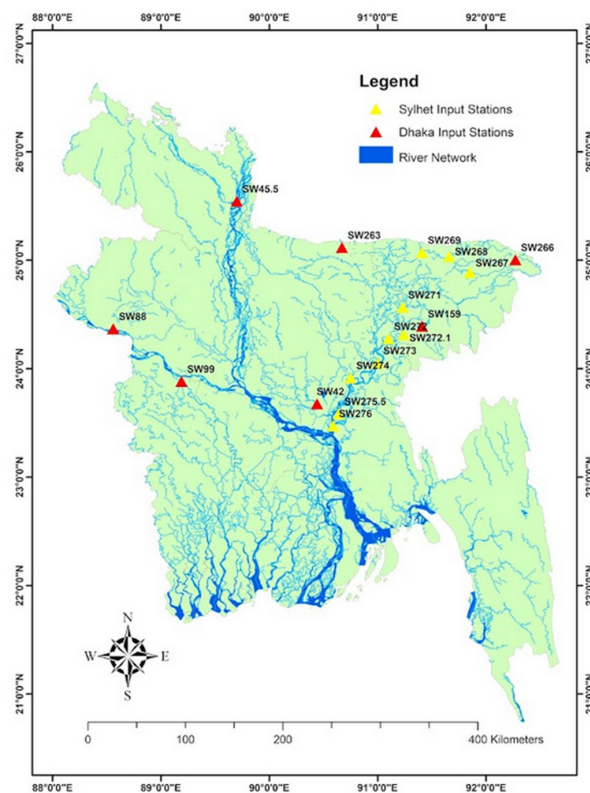


Figure 2. Location of the water-level stations.

In this study, we analyzed the water levels of the Sylhet and Dhaka basins and predicted flood water levels at the main stations. Sylhet (SW267) and Dhaka (SW42) are the main stations of the basins, respectively. Floods at these two stations are forecasted using the water level data of all other stations. For both of these stations, the input feature for a period of 7 days was taken to make predictions up to 7 days into the future. For this study, we developed five neural network models. The input and output of the predictive models to forecast the stations of Dhaka and Sylhet are shown in Table 1.

Table 1. Input and output stations for water level forecasting in Dhaka and Sylhet.

Locations	Input	Output
Dhaka	SW42, SW88, SW45.5, SW263, SW266, SW159, SW99	SW42
Sylhet	SW267, SW268, SW271, SW269, SW272.1, SW272, SW273, SW274, SW275.5, SW276	SW267

Correlation analysis was carried out to verify how correlated the input variables were with the output, estimating the Pearson correlation coefficients [57]. Pearson correlation coefficients are used to measure the relationship between two sets of data. Table 2 presents the interpretation of correlation values. The correlation coefficients of stations in Dhaka and Sylhet are given in Tables 3 and 4, respectively.

Table 2. Interpretation of correlation values.

Range of Correlation Coefficient	Interpretation
$0.8 < r < 1.0$	Very High Correlation
$0.6 < r < 0.79$	High Correlation
$0.4 < r < 0.59$	Moderate Correlation
$0.2 < r < 0.39$	Low Correlation
$0 < r < 0.19$	Very Low Correlation

Table 3. Correlation analysis for Dhaka.

Stations	SW42	SW88	SW45.5	SW263	SW266	SW159	SW99
SW42	1	0.8676	0.9208	0.7044	0.8614	0.5394	0.9032

Table 4. Correlation analysis for Sylhet.

Stations	SW267	SW268	SW269	SW271	SW272.1	SW272	SW273	SW274	SW275.5	SW276
SW267	1	0.9817	0.9557	0.9040	0.8764	0.8994	0.8915	0.8770	0.8543	0.8248

The correlation coefficients for all the inputs to their corresponding output are greater than 0.5. It can be noted that the majority of the input stations chosen for the study are highly correlated to the output stations. To perform the multivariate time series forecasting, we chose 7 stations as our input features to forecast the water level of the Dhaka station. To forecast the water level in Sylhet, we chose 10 stations as independent variables.

3.3. Dataset Preprocessing

3.3.1. Imputation of Missing Data through Mice

Figures 3 and 4 provide a summary of the missing values of input features. The bar plot (left side) represents the percentage of missing data for each of the stations in the dataset. The level plot (right side) represents the combination of data—both missing and present. In the level plot, the available data are shown in blue color, and missing data in yellow color. The level plots are developed for a specific time frame, so the number of yellow/blue bars in the level plot will only count the data for a specific time. The highest proportion of missing data is found in stations SW268 and SW271, where more than 30% of data are missing.

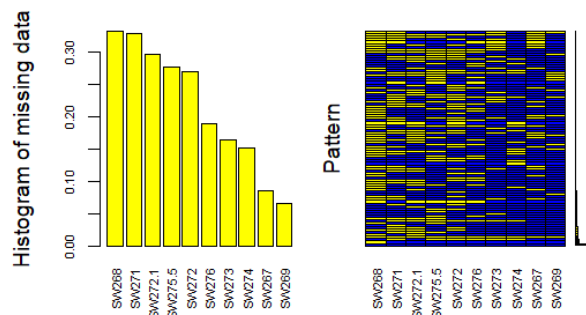


Figure 3. Illustration of the missing data for input stations of Sylhet.

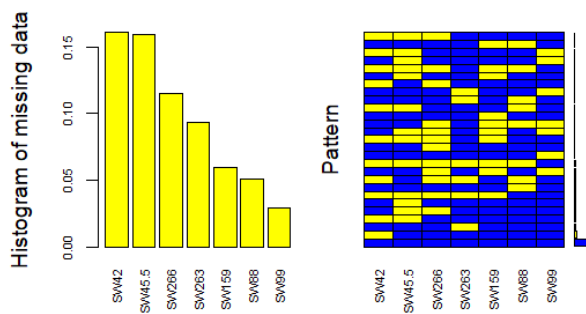


Figure 4. Illustration of the missing data for input stations of Dhaka.

The number of missing values is larger in stations around Sylhet than those located adjacent to Dhaka. Discarding missing data would significantly reduce the number of observations at each station, which may lead to bias and loss of critical information. Hence,

we performed data imputations to address the data gap. Since water level data at different stations are missing randomly, we used the multiple imputation by chained equations (MICE) technique to impute the data gap [58]. MICE is robust and one of the widely used approaches to overcome a large amount of missing data [58]. It imputes missing data by assuming that the data missing are random in the series. It creates different sets of imputed datasets, which allows for using the best-imputed set. The data imputation process involves five steps:

- Step 1: Perform simple imputation, such as mean imputation, on the dataset.
- Step 2: Select the variable with the lowest missing data. Remove the imputed value of this variable.
- Step 3: Use a regression model to predict the missing value of the selected variable by using other existing variables. The selected variable is considered to be the dependent variable, and the other variables are to be independent variables.
- Step 4: Impute the missing value of the dependent variable using the regression model.
- Step 5: Select the next variable with a missing value.
- Step 6: Repeat steps 2 to 5 for all variables.

3.3.2. Min–Max Normalization

Since different stations had various magnitudes of water level, we performed min–max normalization on the input features using Equation (13). The dataset was first split into the train and test set, and then normalization was applied to prevent losing information of the training set.

$$x_{\text{normalization}} = \frac{x - X_{\min}}{x_{\max} - X_{\min}} \tag{1}$$

3.4. Models Employed

3.4.1. Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are mathematical function systems developed based on the inspiration of biological neurons. Instead of synapses in the human brain, an ANN model has artificial neurons, and instead of axons, it has weights. The model is typically constructed with various connections among nodes arranged in layers. It can have different numbers of layers. Two nodes from two layers are connected by an edge, and every edge has its weight. Each node in ANNs receives a signal, and then processes it and signals the node connected next to it. The “signal” or “response” from a neuron is a real number, and this response or signal is calculated by some linear or nonlinear function with the basic weighted sum approach. Neurons may have a threshold such that a signal is sent only if the aggregated signal crosses that threshold, which is controlled by functions called activation functions. Figures 5 and 6 represent an ANN structure that will be used in this study to forecast the water level.

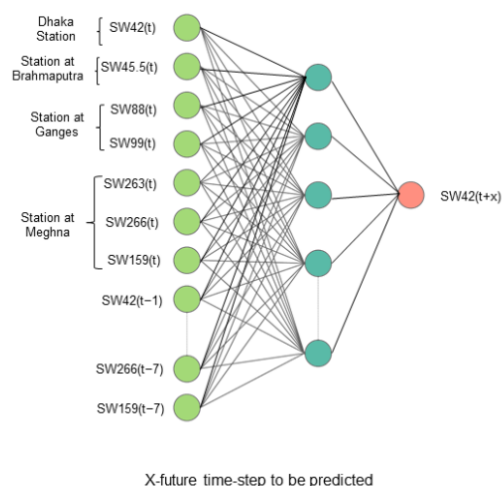


Figure 5. ANN model for forecasting the water level of the Dhaka station.

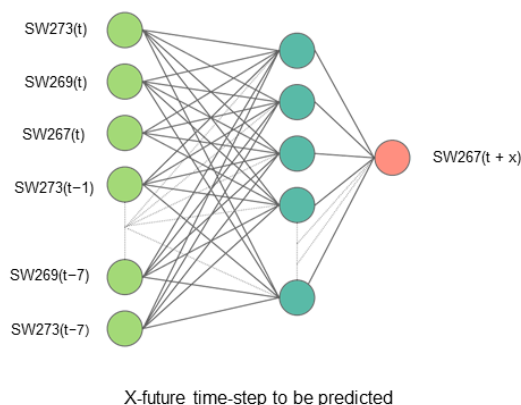


Figure 6. ANN model for forecasting the water level of the Sylhet station.

3.4.2. Long Short-Term Memory (LSTM)

The long short-term memory (LSTM) network [38] is an artificial recurrent neural network (RNN) model that can remember the order in sequential data. The two major problems of the traditional RNN architecture are the vanishing gradient problem and its inability to retain information of long sequences. LSTM overcomes these issues with the help of its cell state and gates. The architecture of the LSTM model allows it to perform well in time series forecasting.

LSTM consists of three gates that control the flow of information. These three gates are input gate (it), output gate (ot), and most importantly, forget gate (ft). The hidden state of the LSTM network is the short-term memory, while the cell state is the long-term memory. Operations within LSTM cells help the model retain the information of sequential data. LSTM uses cells as the memory box for the model. The forget gate, shown in Equation (2), determines how much information from the previous timestamp needs to be forwarded to the current cell state:

$$f_t = \sigma(x_t * U_f + H_{t-1} * W_f) \tag{2}$$

where U_f represents the weight matrix that connects the input layer of its cell to the hidden layer. The input gate controls what extent of information is necessary to be added to the new cell state (C_t) In Equation (3), it is mathematically described that input gate takes in the previous hidden state (H_{t-1}) value and the input of the current time step (x_t) value.

$$i_t = \sigma(x_t * U_i + H_{t-1} * W_i) \tag{3}$$

In Equation (4), the output gate values are generated by passing the information of the current input and previous hidden state H_{t-1} through the sigmoid activation function. In Equation (5), \tilde{C}_t is produced based on the current input, x_t , and the previous hidden state. Then, taking the previous cell state, C_{t-1} , \tilde{C}_t is used in Equation (6) to produce the current cell state, C_t . Lastly in Equation (7), the output gate as well as the new cell state output helps to determine the new hidden state, H_t .

$$o_t = \sigma(x_o * U_o + H_{t-1} * W_o) \tag{4}$$

$$\tilde{C}_t = \tanh(U_c x_t + W_c H_{t-1}) \tag{5}$$

$$C_t = \sigma(f_t * C_{t-1} + I_t * \tilde{C}_t) \tag{6}$$

$$H_t = o_t * \tanh(C_t) \tag{7}$$

Due to the usage of the sigmoid function, the value of the gates remains between 0 and 1. The closer the value of the output of these gates gets to 0, the more it indicates that the information is not worth remembering. Similarly, a value closer to 1 increases the

importance of the information. Moreover, the LSTM uses the hyperbolic tangent function to determine the current hidden state (H_{t-1}), which helps increase and decrease the cell state.

3.4.3. Attention

All the input stations used for prediction may or may not influence the water level forecasting. However, it is known that the river networks that are closer to the output station or more deeply connected to the output station will have a greater influence [9]. However, the nature of the water flow of the rivers in Bangladesh is complex to determine the actual river network, which might be having a greater influence. This influence may keep changing with time. As the river network system is affected by many geographical factors, it is difficult to identify which hydrological factors influence the most without having expert domain knowledge. Moreover, using irrelevant data for the prediction can lead to model inaccuracy. Hence, it is a challenge to make an accurate prediction as basic predictive models are unable to understand the importance of different features and how the importance of one feature can be more than the other.

Therefore, a deep learning model that uses an attention mechanism can overcome such a challenge. Attention has been widely used in sequence models to improve the accuracy of the prediction [38,39,51]. Attention allows models to narrow down their focus on important features. Attention is a module that helps deep learning models to gain the ability to focus on crucial information from all the data available to it. The attention module has also proven to increase the accuracy of deep learning models.

Attention is of two types: hard and soft attention. Hard attention is useful when the model needs weight with a value of either 0 or 1. On the other hand, soft attention distributes the weights within the range of 0 to 1. Hence, compared with hard attention, soft attention provides more flexibility in assigning weights. For our study, we will be using soft attention.

From all the information presented to a model, the attention mechanism can help the model to focus on critical information. With spatial and temporal features of a time series dataset, we can use attention to find correlations among the different features changing over time and thus put more focus on the important features.

3.4.4. Spatial and Temporal Attentions

The spatial attention LSTM uses the spatial attention module to find out the correlation between independent feature and output. In this study, we are forecasting the water level of a certain station using the water level from its nearby stations. Water levels from different stations and locations are taken as input features to predict the water level of a particular station. However, not all the input features need to have a positive effect on the output feature. Giving excess weight to a feature that does not provide any contribution to the output variable will lead to inaccuracy. In the case of spatial data like this, it becomes difficult for the deep learning models to find out the correlation of different spatial data to the output. Therefore, implementing spatial attention to the base neural network will allow the model to understand the spatial correlation of the stations to the station that will be predicted.

To produce spatial attention, we used probabilistic activation sigmoid function as the function generates a value based on the input feature, and then the softmax is used to normalize the values. The production of attention weights is shown in Equation (9). Then, by element-wise product, shown in Equation (10), as \odot , the spatial attention is being implemented to the feature matrices in Equation (8):

$$x_t = [f_1^t, f_2^t, \dots, f_n^t]_{n \times 1} \tag{8}$$

$$\text{Spatial Attention, } \alpha_t = [\alpha_1^t, \alpha_2^t, \dots, \alpha_n^t]_{1 \times n}^T \tag{9}$$

$$x'_t = \alpha_t \odot x_t = [\alpha_1^t f_1^t, \alpha_2^t f_2^t, \dots, \alpha_n^t f_n^t]_{n \times 1} \tag{10}$$

where n represents the total input features that will be processed in a single time step, while k represents the different time steps. The element-wise product is also known as the Hadamard product, where identical matrices are being multiplied to produce another matrix of a similar dimension.

Temporal attention can be used to model the dynamic correlation between different time intervals in the target time series. Incorporating the spatial attention module and temporal attention module in an LSTM network can create a spatial-LSTM and temporal-LSTM, respectively [38]. The temporal attention allows the model to focus on moments, which is most influential for the forecasting of the output for that time step. For example, in our study, we are taking the water levels of multiple water-level stations for 7 days to predict water levels in the next time step (i.e., $n + 1$ to $n + 7$ time step). The temporal attention enables deep learning models to distribute weights of the independent variables based on time. In Equation (12), the temporal attention is generated after the weights are passed through ReLU and then through the softmax. In Equation (13), the temporal attention takes in the hidden layer states from Equation (11) to produce modified hidden weights with temporal attention incorporated within it. \otimes denotes the matrix product.

$$H = [h_1, h_2, \dots, h_k]_{k \times s} \quad (11)$$

$$\text{Temporal Attention, } \beta = [\beta_1, \beta_2, \dots, \beta_k]_{1 \times k} \quad (12)$$

$$H_{\text{attention}} = \beta \otimes H \quad (13)$$

3.4.5. Spatiotemporal Attention LSTM (STALSTM)

By implementing both the spatial and temporal attention to a neural network model, it will allow the model to narrow its focus on the spatiotemporal aspect at once.

One of the major benefits of the spatiotemporal attention LSTM model is that it can automatically select informative factors based on the inherent characteristics of the collected historical data. The STALSTM can identify key variables and moments important for forecasting and more suitable for multivariate time series data. With the implementation of both spatial and temporal attention, the neural network model is also able to identify the dynamic changing relationship of the independent variable to the dependent variable. Thus, a spatiotemporal attention-based model can overcome problems, faced by basic deep learning architecture.

The attention mechanism can find out key features that are responsible for the changes in the water level of the station to be forecasted. Thus, a model like the spatiotemporal attention neural network model can prove to be useful in such cases.

Implementing a spatial attention module and a temporal attention module in the LSTM can help the predictive deep learning model to figure out which of the input features is important for forecasting. The spatial module of the STALSTM will identify the complex relationship among the features and allow the model and would put more focus on those important features. For spatial attention, the feature vectors are passed through the sigmoid function to generate weights with respect to each feature. The higher is the importance of a feature, the more attention weight is generated for it.

The structure of the spatiotemporal attention LSTM used in this study is shown in Figure 7. This modified LSTM structure has been proven to increase the accuracy of LSTM models [38,39]. In Figure 7, the input features represent the input stations that will be used by the model to forecast the water level of a station. After the element-wise operation of the input features and the spatial attention module, the data are then passed to the LSTM cells. The temporal attention would then be incorporated into the hidden layer states, as shown in Equation (11), to produce modified hidden weights with temporal attention incorporated within it. Thus, these attention modules allow the STALSTM to take spatial and temporal variation for making predictions.

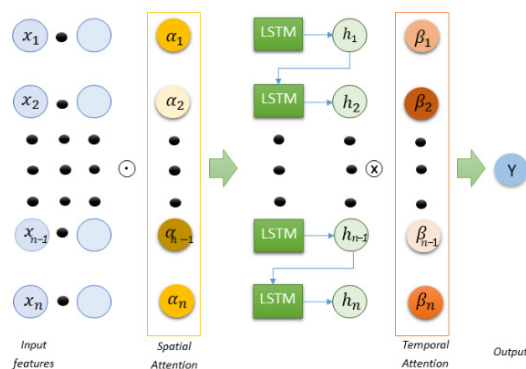


Figure 7. Structure of spatiotemporal attention LSTM.

3.4.6. Evaluation Criteria

To select the best-performing model, we selected six most commonly used evaluation metrics used in time-series analysis. The model, showing the best performance under most of the evaluation metrics, is selected as most efficient and accurate to forecast water levels.

For these evaluation metrics, x represents actual value, y represents predicted value, and n represents the total number of test samples. The evaluation metrics used in the study are as follows:

$RMSE$ is the root mean square error, which is one of the widely used metrics to analyze the accuracy of forecasting. The formula is shown in Equation (14).

$$RMSE = \sqrt{\frac{\sum (x - y)^2}{n}} \tag{14}$$

MAE is the mean absolute error. MAE helps us to understand how well the models are able to accurately predict and by what amount these values are deviating from the actual one. The formula is shown in Equation (15).

$$MAE = \frac{\sum |(x - y)|}{n} \tag{15}$$

$MAPE$ is the mean absolute percentage error. The formula is shown in Equation (16).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x - y}{x} \right| * 100\% \tag{16}$$

R^2 is the coefficient of determination. It helps to analyze the performance of the model. A higher value of R^2 represents a better fit for the model. The formula is shown in Equation (17).

$$R^2 = \frac{[\sum (x - \bar{x})(y - \bar{y})]^2}{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2} \tag{17}$$

ESD is the error standard deviation. Here, e represents error, and \bar{e} represents the dispersion between the predicted and observed data. The formula is shown in Equation (18).

$$ESD = \sqrt{\frac{\sum (e - \bar{e})^2}{n}} \tag{18}$$

NSE is the Nash–Sutcliffe model efficiency coefficient. This statistical criterion is used to determine model performance. An NSE value greater than 0.75 represents good performance of the model [59].

$$NSE = 1 - \frac{\sum_{i=1}^n (x - y)^2}{\sum_{i=1}^n (x - \bar{x})^2} \quad (19)$$

4. Results and Discussions

Table 5 shows the average performance of the models for forecasting water levels in main stations, located in Dhaka and Sylhet. From the evaluation metrics, it can be seen that the STALSTM model has the best forecasting results. Lower values of *RMSE* and a higher value of R^2 indicate a higher level of prediction accuracy. The overall performance of the neural network is better for stations in Dhaka than Sylhet. This may reflect a greater number of missing data in the Sylhet stations. Any changes in the imputation method can influence the accuracy of predictions [55]. Data cleaning methods may also impact the performance of the model. Overall, the STALSTM model outperformed other neural network models tested in the study. This is because the model inherits both spatial and temporal attention and utilizes a mechanism to make reliable predictions [38]. The STALSTM model outperformed the basic LSTM model. This indicates that the attention mechanism has a positive effect on its performance [49]. The error trend of all attention models is similar to the LSTM as it is used as the base. Among all the neural network models employed in the study, the performance of the ANN is least for both the Dhaka and Sylhet stations. Unlike the LSTM, the ANN lacks a memory cell, which is needed to retain long sequences of information to make predictions [38,39,41].

Table 5. Average performance of the predictive models.

	Dhaka Station						Sylhet Station					
	<i>RMSE</i>	<i>MAE</i>	<i>MAPE</i>	R^2	<i>ESD</i>	<i>NSE</i>	<i>RMSE</i>	<i>MAE</i>	<i>MAPE</i>	R^2	<i>ESD</i>	<i>NSE</i>
ANN	0.39	0.34	16.1%	0.93	0.25	0.91	1.08	0.73	16.16%	0.86	0.78	0.84
LSTM	0.40	0.29	12.8%	0.94	0.23	0.94	1.06	0.71	15.53%	0.87	0.75	0.86
SALSTM	0.34	0.24	9.79%	0.95	0.21	0.95	1.00	0.69	13.99%	0.90	0.72	0.88
TALSTM	0.35	0.25	10.81%	0.95	0.19	0.94	1.02	0.70	14.48%	0.89	0.73	0.89
STALSTM	0.31	0.21	9.36%	0.97	0.17	0.96	0.94	0.65	12.41%	0.93	0.70	0.91

Figures 8 and 9 represent the performance of the predictive models on forecasting the Dhaka station and the Sylhet station for all different time steps for each of the evaluation metrics. For all the future forecasting, the STALSTM model shows the least *RMSE*. Figures 8 and 9 also show how the performance of the models changes with the lead day. As the number of days increases, the error for each model increases. There is a difference in the performance of the ANN and LSTM due to the difference in the model architecture.

Compared with other models, the incorporation of attention methods drastically improved the forecasting accuracy. The spatial attention module puts more weight on the appropriate input features for a certain time step prediction [38,39]. By providing adequate relevant features to the input, the STALSTM was able to outperform all other models. The lowest error achieved in this model suggests that it is successful in capturing the importance of variables in the prediction of the water level at a certain time step.

To assess the performance of the STALSTM, the attention weights are extracted and visualized. Figure 10 illustrates weights of the STALSTM model used to predict the water level of the Dhaka stations. Light color represents low attention weights, and dark color denotes high attention weights. The bar presented on the right of the heat maps represents the attention weight values. Here, the attention weights are not only different for varying moments in time, but also different on input features. If the model is given an n -previous time step to forecast one step ahead, the model is most likely to take values from the moment it fits. For forecasting of the Dhaka station at a one-time step ahead, the SW159 station with $t - 2$ has been assigned with more weights compared with other stations. This

is because, at a certain time step, the attention mechanism increases the weights of input information, which are useful for that time step [38].

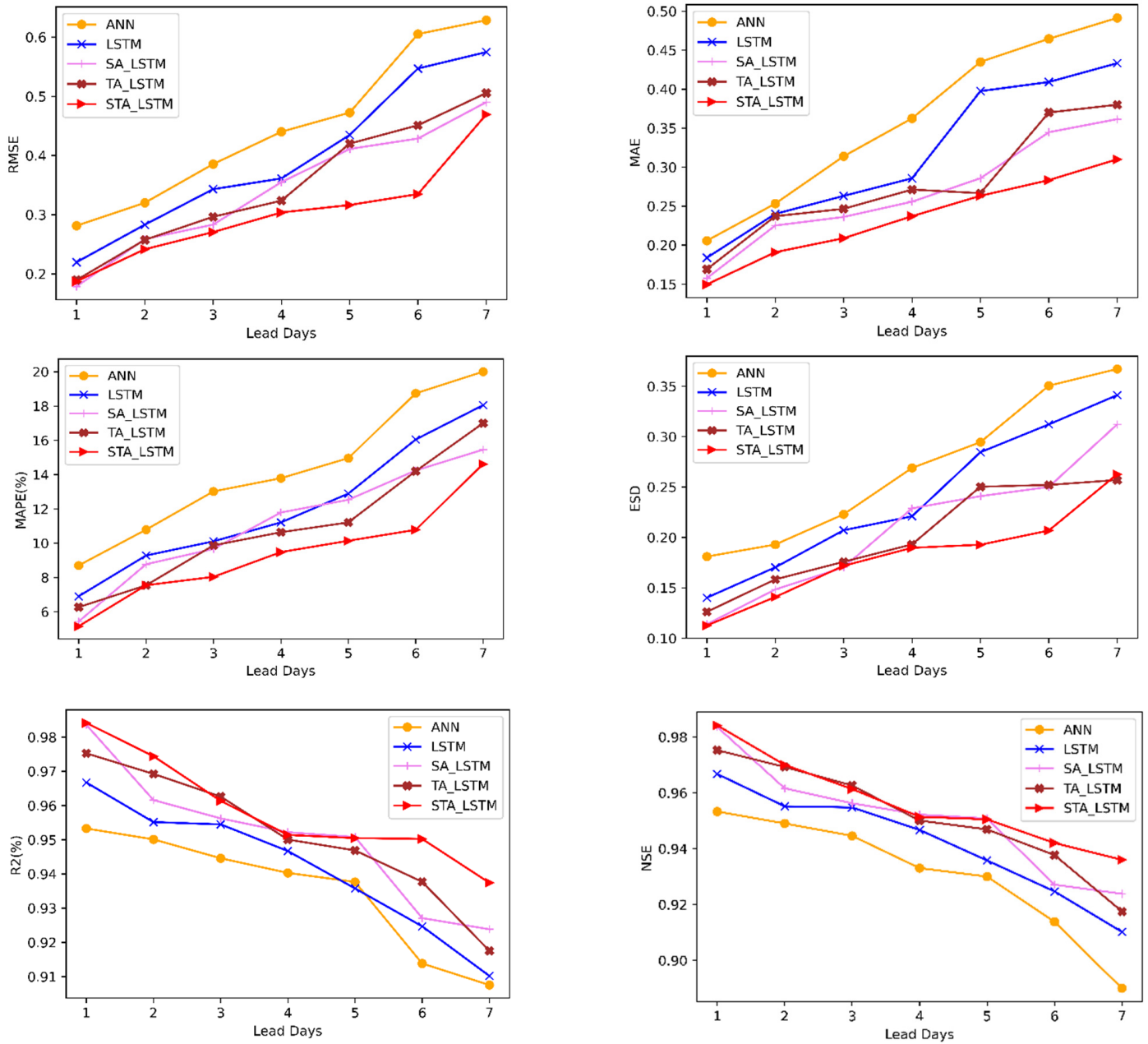


Figure 8. Comparison of the performance of the predictive model of the Dhaka station under different evaluation metrics.

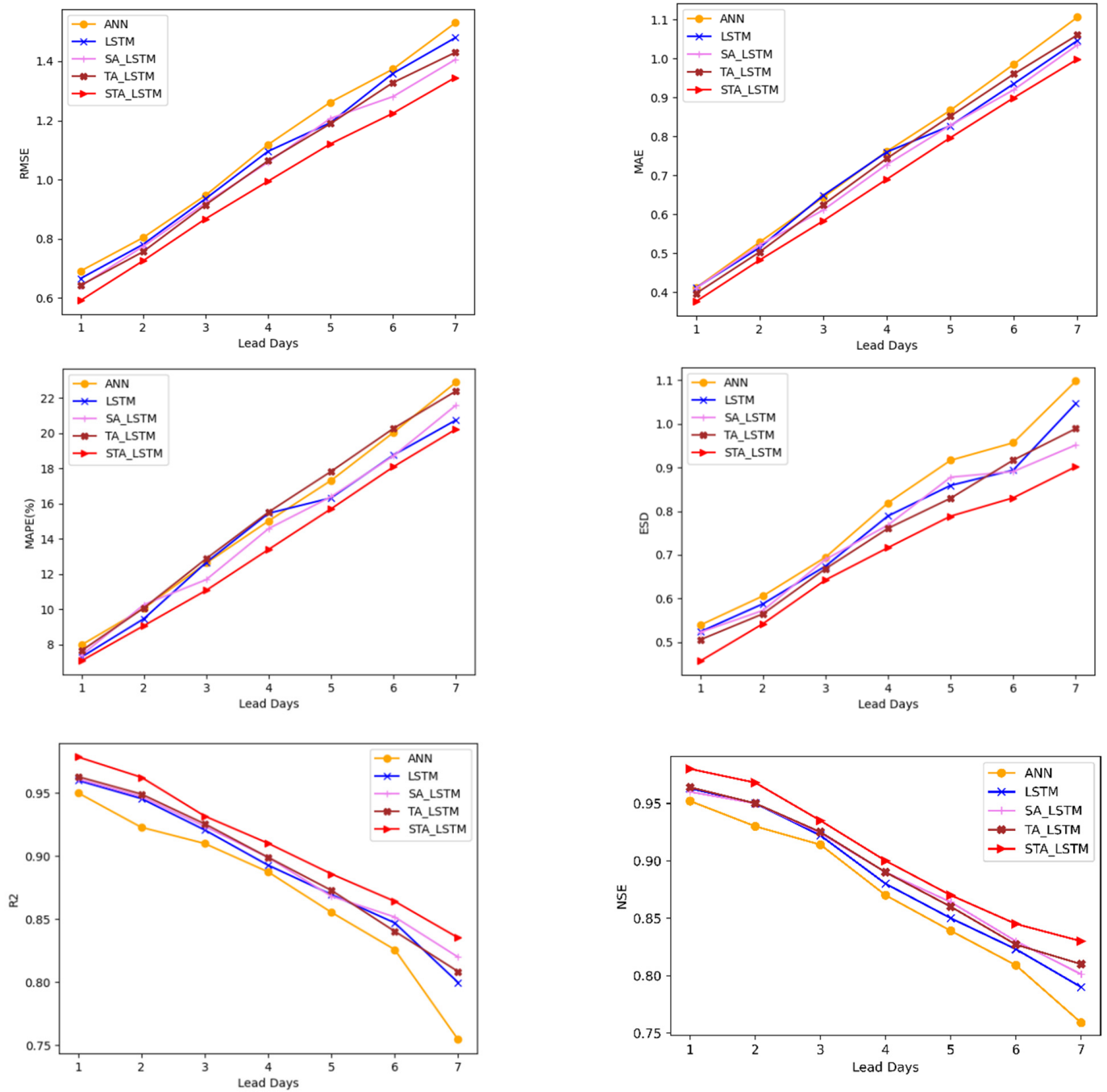


Figure 9. Comparison of the performance of the predictive model of the Sylhet station under different evaluation metrics.

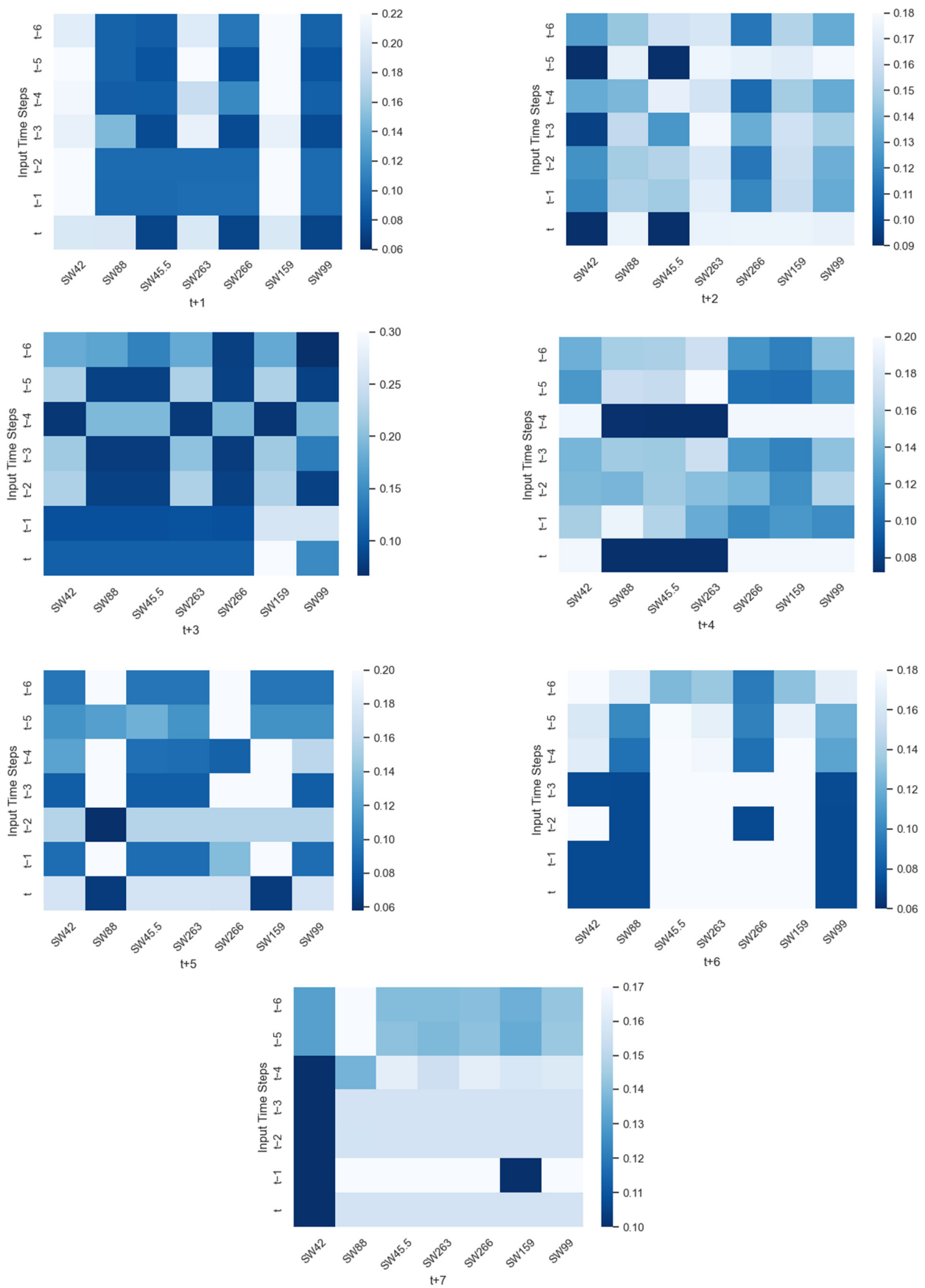


Figure 10. Distribution of the STALSTM attention weights for the Dhaka station at different prediction moments.

It is worth noting that the further into the future the model is used to make a prediction, the more the model focuses its attention on the values that are nearby to the present time, t . The attention moves its focus towards more recent water level values of the input. When predicting water levels 7 days into the future, we can see how the attention module focuses by putting more weight on the values in the current day. Hence, attention weights for $t + 7$ are more for the more recent historical moment of the Dhaka station. Therefore, it can be observed how the distribution of spatial and temporal attention modules can enable the LSTM architecture to capture the complex relationship to perform accurate flood forecasting [54,60].

Figure 11 represents a subsection of the predictive models forecasting on the testing set for the Dhaka station, where prediction can be compared with ground truth information. Here, increasing the lead day makes the prediction performed by the model more uncertain. For $t + 1$, the prediction model is close to the actual water level. However, with an increasing lead time, the predictions become more far off than the actual value. Among all the models used in this study, the STA_LSTM shows the most stable performance. The prediction of the ANN model fluctuated more heavily compared with other models. The fluctuations are observed near the peaks due to sudden changes in water level.

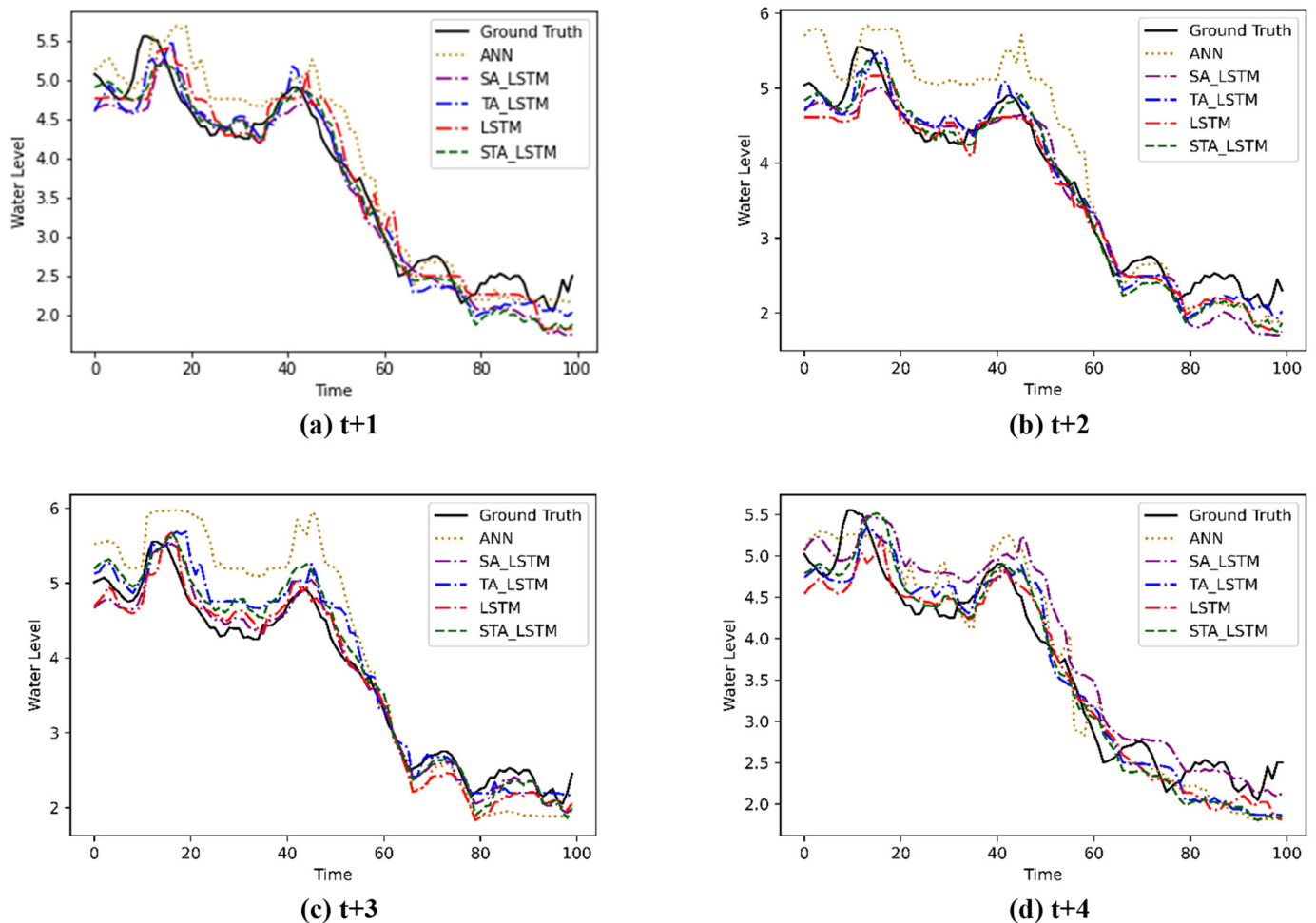


Figure 11. Cont.

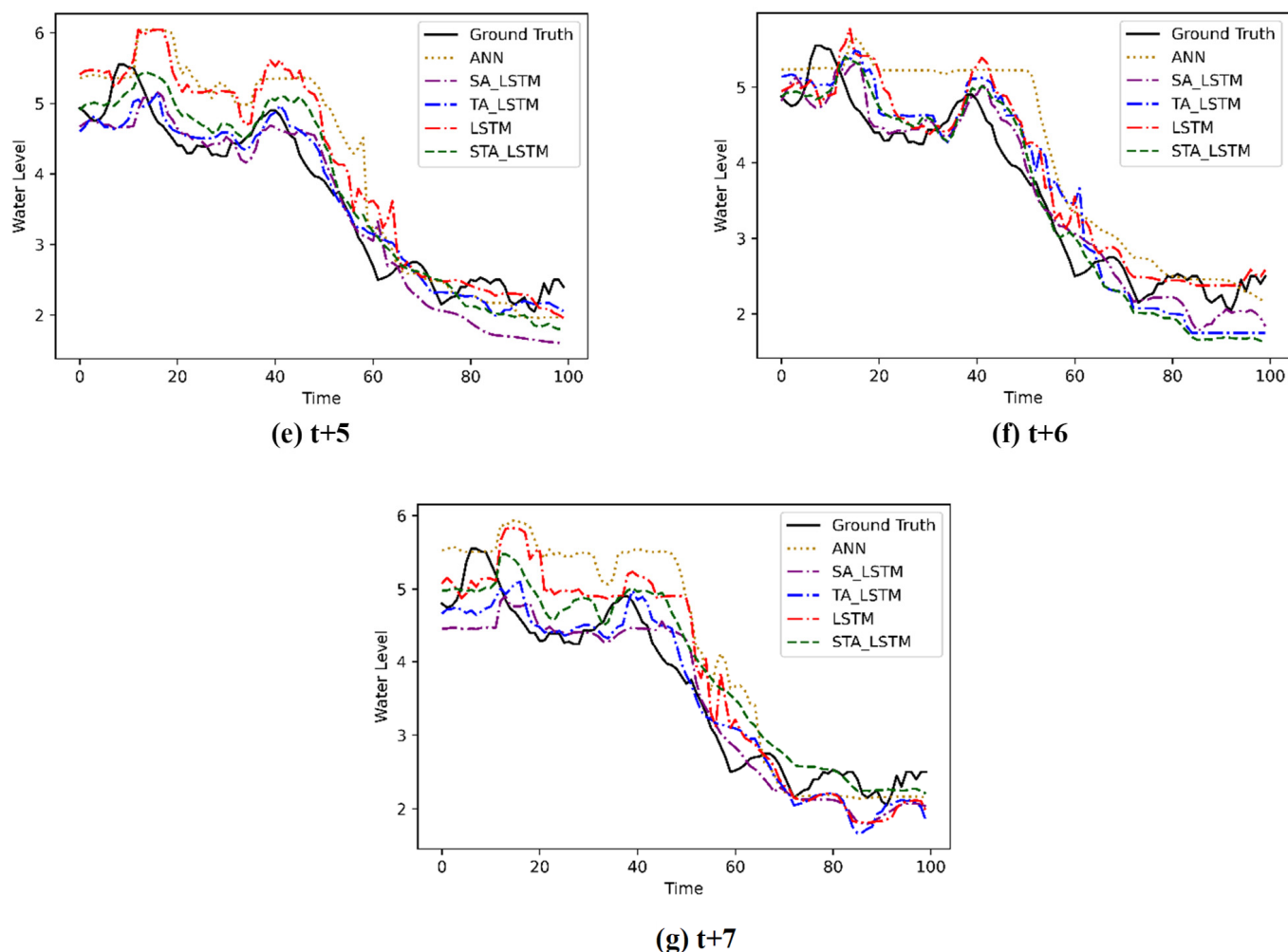


Figure 11. The predicted water levels in Dhaka station, with ground observations, for different time-steps: (a) one-day ahead, (b) two days ahead, (c) three days ahead, (d) four days ahead, (e) five days ahead, (f) six days ahead, and (g) seven days ahead.

The results, obtained from the study, illustrate that when using multivariate time series to make multistep ahead prediction, the LSTM or its variants can perform better than the ANN model due to the advantage of having the ability to retain a long sequence of information. The performance of the models can be ranked as follows:

$$\text{STALSTM} > \text{SALSTM} > \text{TALSTM} > \text{LSTM} > \text{ANN} \quad (20)$$

5. Conclusions

Bangladesh is highly prone to flooding, particularly during the monsoon season. To improve the flood forecasting system, an accurate lead time of the water level of the rivers is essential. The results of this study demonstrated that the use of spatial and temporal modules in the LSTM models can predict riverine flooding accurately.

This study proposes a sequential model to enhance the water level forecast. It explored how deep learning models can be used for multivariate water level prediction for multistep forecasting. The performance of the prediction models has been analyzed using widely known evaluation metrics followed in the field of time-series analysis. The results show that the STALSTM outperforms the basic LSTM and its other attention models and the ANN in accurately predicting the water level. The attention modules are shown to improve the accuracy of the deep learning model for forecasting.

It is worth noting that the attention models are data intensive, requiring a large number of in situ data. The performance of the STALSTM model tends to be poor when smaller datasets are used [38]. Despite being sensitive to data volume, the LSTM model can address nonlinearity issues in flood prediction in complex river network systems. Future works can focus on exploring the effects of different input lengths of the sequential data and missing data imputation method on the accuracy of flood forecasting.

Author Contributions: Conceptualization, R.M.R., A.D., F.N. and M.S.G.A.; methodology, F.N., S.H., M.R., T.A., Z.J., Z.S.S. and R.T.H.; validation, R.M.R., A.D., F.N. and M.S.G.A.; formal analysis, F.N., S.H., M.R., T.A., Z.J., Z.S.S. and R.T.H.; investigation, R.M.R., A.D. and M.S.G.A.; resources, R.M.R., A.D. and M.S.G.A.; data curation, R.M.R., A.D., F.N., Z.S.S., R.T.H. and M.S.G.A.; writing—original draft preparation, F.N., S.H., M.R., T.A., Z.J., Z.S.S. and R.T.H.; writing—review and editing, F.N., R.M.R., A.D. and M.S.G.A.; visualization, F.N., Z.S.S. and R.T.H.; supervision, R.M.R., A.D. and M.S.G.A.; project administration, R.M.R. and M.S.G.A.; funding acquisition, R.M.R. and M.S.G.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Post, Telecommunication and Information Technology, Bangladesh through ICT Innovation Fund (2020-21) round 3: Grant Number 12.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the first author upon reasonable request.

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

References

1. Karim, N. Disasters in Bangladesh. *Nat. Hazards* **1995**, *11*, 247–258. [[CrossRef](#)]
2. Cutter, S.L. Vulnerability to environmental hazards. *Prog. Hum. Geogr.* **1996**, *20*, 529–539. [[CrossRef](#)]
3. Kisinger, C.; Matsui, K. Responding to climate-induced displacement in Bangladesh: A governance perspective. *Sustainability* **2021**, *13*, 7788. [[CrossRef](#)]
4. Kabir, H.; Hossen, N. Impacts of flood and its possible solution in Bangladesh. *Disaster Adv.* **2019**, *12*, 48–57.
5. Dasgupta, S.; Huq, M.; Khan, Z.H.; Sohel Masud, M.; Ahmed, M.M.Z.; Mukherjee, N.; Pandey, K. Climate proofing infrastructure in Bangladesh: The incremental cost of limiting future flood damage. *J. Environ. Dev.* **2011**, *20*, 167–190. [[CrossRef](#)]
6. Mirza, M.M.Q. Three recent extreme floods in Bangladesh: A hydro-meteorological analysis. In *Flood Problem and Management in South Asia*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 35–64.
7. Faisal, I.; Kabir, M.; Nishat, A. Non-structural flood mitigation measures for Dhaka City. *Urban Water* **1999**, *1*, 145–153. [[CrossRef](#)]
8. Biswas, R.; Jayawardena, A.; Takeuchi, K. Prediction of water levels in the Surma River of Bangladesh by artificial neural network. In Proceedings of the 22nd Annual Conference (2009), Japan Society of Hydrology and Water Resources, Hachioji, Tokyo, 25 December 2009.
9. Liong, S.-Y.; Lim, W.-H.; Paudyal, G.N. River stage forecasting in Bangladesh: Neural network approach. *J. Comput. Civ. Eng.* **2000**, *14*, 1–8. [[CrossRef](#)]
10. Adnan, M.S.G.; Dewan, A.; Zannat, K.E.; Abdullah, A.Y.M. The use of watershed geomorphic data in flash flood susceptibility zoning: A case study of the Karnaphuli and Sangu river basins of Bangladesh. *Nat. Hazards* **2019**, *99*, 425–448. [[CrossRef](#)]
11. Adnan, M.S.G.; Talchabhadel, R.; Nakagawa, H.; Hall, J.W. The potential of tidal river management for flood alleviation in south western Bangladesh. *Sci. Total Environ.* **2020**, *731*, 138747. [[CrossRef](#)]
12. Dastagir, M.R. Modeling recent climate change induced extreme events in Bangladesh: A review. *Weather Clim. Extrem.* **2015**, *7*, 49–60. [[CrossRef](#)]
13. Dewan, T.H. Societal impacts and vulnerability to floods in Bangladesh and Nepal. *Weather Clim. Extrem.* **2015**, *7*, 36–42. [[CrossRef](#)]
14. Huq, M.E.; Shoeb, A.; Javed, A.; Shao, Z.; Hossain, M.A.; Sarven, M.S. Measuring vulnerability for city dwellers exposed to flood hazard: A case study of Dhaka City, Bangladesh. In *Urban Intelligence and Applications*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 207–215.
15. BWDB. *Processing and Flood Forecasting Circle*; Bangladesh Water Development Board (BWDB): Dhaka, Bangladesh, 2021.
16. Siddiquee, M.S.A.; Hossain, M.M.A. Development of a sequential Artificial Neural Network for predicting river water levels based on Brahmaputra and Ganges water levels. *Neural Comput. Appl.* **2015**, *26*, 1979–1990. [[CrossRef](#)]
17. Shaw, R.; Mallick, F.; Islam, A. *Climate Change Adaptation Actions in Bangladesh*; Springer: Berlin/Heidelberg, Germany, 2013.
18. Ozaki, M. *Disaster Risk Financing in Bangladesh*; Asian Development Bank (ADB): Manila, Philippines, 2016.

19. Hasan, A.; Saha, R.; Biswas, B. A Study on the Flood Damage and Mitigation Measures of Floods Occurring in Bangladesh at the Last Decade. In Proceedings of the 1st National Conference on Water Resources Engineering (NCWRE), Chittagong University of Engineering and Technology, Chittagong, Bangladesh, 21–22 March 2018; pp. 21–22.
20. Dewan, A.M.; Kumamoto, T.; Nishigaki, M. Flood hazard delineation in greater Dhaka, Bangladesh using an integrated GIS and remote sensing approach. *Geocarto Int.* **2006**, *21*, 33–38. [[CrossRef](#)]
21. Hayat, H.; Akbar, T.A.; Tahir, A.A.; Hassan, Q.K.; Dewan, A.; Irshad, M. Simulating current and future river-flows in the Karakoram and Himalayan regions of Pakistan using snow-melt-runoff model and RCP scenarios. *Water* **2019**, *11*, 761. [[CrossRef](#)]
22. GoB. *Action for Disaster Risk Management towards Resilient Nation*; Government of Bangladesh (GoB): Dhaka, Bangladesh, 2020.
23. Rahman, M.H.; Rahman, M.S.; Rahman, M.M. Disasters in Bangladesh: Mitigation and Management. *Barisal Univ. J. Part* **2017**, *1*, 1.
24. Adnan, M.S.G.; Haque, A.; Hall, J.W. Have coastal embankments reduced flooding in Bangladesh? *Sci. Total Environ.* **2019**, *682*, 405–416. [[CrossRef](#)]
25. Fakhruddin, S.; Kawasaki, A.; Babel, M.S. Community responses to flood early warning system: Case study in Kaijuri Union, Bangladesh. *Int. J. Disaster Risk Reduct.* **2015**, *14*, 323–331. [[CrossRef](#)]
26. Ali, M.H.; Bhattacharya, B.; Islam, A.; Islam, G.; Hossain, M.S.; Khan, A. Challenges for flood risk management in flood-prone Sirajganj region of Bangladesh. *J. Flood Risk Manag.* **2019**, *12*, e12450. [[CrossRef](#)]
27. Roy, B.; Khan, M.S.M.; Islam, A.S.; Mohammed, K.; Khan, M.J.U. Climate-induced flood inundation for the Arial Khan River of Bangladesh using open-source SWAT and HEC-RAS model for RCP8.5-SSP5 Scenario. *SN Appl. Sci.* **2021**, *3*, 648. [[CrossRef](#)]
28. Chowdhury, A.; Reshad, S.; Kumruzzaman, M. Hydrodynamic Flood Modelling for the Jamuna River using HEC-RAS & MIKE 11. In Proceedings of the 5th International Conference on Advances in Civil Engineering (ICACE-2020), Chattogram, Bangladesh, 21–23 December 2020.
29. Rahman, M.M.; Goel, N.; Arya, D. Development of the Jamuneswari flood forecasting system: Case study in Bangladesh. *J. Hydrol. Eng.* **2012**, *17*, 1123–1140. [[CrossRef](#)]
30. Di Baldassarre, G.; Schumann, G.; Bates, P.D.; Freer, J.E.; Beven, K.J. Flood-plain mapping: A critical discussion of deterministic and probabilistic approaches. *Hydrol. Sci. J. J. Sci. Hydrol.* **2010**, *55*, 364–376. [[CrossRef](#)]
31. Islam, A. Improving flood forecasting in Bangladesh using an artificial neural network. *J. Hydroinform.* **2010**, *12*, 351–364. [[CrossRef](#)]
32. Biswas, R.; Jayawardena, A. Water level prediction by artificial neural network in a flashy transboundary river of Bangladesh. *Glob. Nest J.* **2014**, *16*, 432–444.
33. Ullah, N.; Choudhury, P. Flood flow modeling in a river system using adaptive neuro-fuzzy inference system. *Env. Manag. Sustain. Dev.* **2013**, *2*, 54–68. [[CrossRef](#)]
34. Liong, S.Y.; Lim, W.H.; Kojiri, T.; Hori, T. Advance flood forecasting for flood stricken Bangladesh with a fuzzy reasoning method. *Hydrol. Processes* **2000**, *14*, 431–448. [[CrossRef](#)]
35. Liong, S.Y.; Sivapragasam, C. Flood stage forecasting with support vector machines 1. *J. Am. Water Resour. Assoc.* **2002**, *38*, 173–186. [[CrossRef](#)]
36. Widiasari, I.R.; Nugoho, L.E.; Efendi, R. Context-based hydrology time series data for a flood prediction model using LSTM. In Proceedings of the 2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, 27–28 September 2018; pp. 385–390.
37. Fang, Z.; Wang, Y.; Peng, L.; Hong, H. Predicting flood susceptibility using LSTM neural networks. *J. Hydrol.* **2021**, *594*, 125734. [[CrossRef](#)]
38. Ding, Y.; Zhu, Y.; Feng, J.; Zhang, P.; Cheng, Z. Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing* **2020**, *403*, 348–359. [[CrossRef](#)]
39. Wu, Y.; Ding, Y.; Zhu, Y.; Feng, J.; Wang, S. Complexity to forecast flood: Problem definition and spatiotemporal attention LSTM solution. *Complexity* **2020**, *2020*, 7670382. [[CrossRef](#)]
40. Siam, Z.S.; Hasan, R.T.; Anik, S.S.; Noor, F.; Adnan, M.S.G.; Rahman, R.M. Study of Hybridized Support Vector Regression Based Flood Susceptibility Mapping for Bangladesh. In Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Kuala Lumpur, Malaysia, 26–29 July 2021; pp. 59–71.
41. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
42. Sai, F.; Cumiskey, L.; Weerts, A.; Bhattacharya, B.; Haque Khan, R. Towards impact-based flood forecasting and warning in Bangladesh: A case study at the local level in Sirajganj district. *Nat. Hazards Earth Syst. Sci. Discuss.* **2018**, 1–20. [[CrossRef](#)]
43. Dewan, A.; Corner, R.; Saleem, A.; Rahman, M.M.; Haider, M.R.; Rahman, M.M.; Sarker, M.H. Assessing channel changes of the Ganges-Padma River system in Bangladesh using Landsat and hydrological data. *Geomorphology* **2017**, *276*, 257–279. [[CrossRef](#)]
44. Iwendi, C.; Maddikunta, P.K.R.; Gadekallu, T.R.; Lakshmana, K.; Bashir, A.K.; Piran, M.J. A metaheuristic optimization approach for energy efficiency in the IoT networks. *Softw. Pract. Exp.* **2021**, *51*, 2558–2571. [[CrossRef](#)]
45. Dhanamjayulu, C.; Nizhal, N.U.; Maddikunta, P.K.R.; Gadekallu, T.R.; Iwendi, C.; Wei, C.; Xin, Q. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Processing* **2021**, *16*, 647–658. [[CrossRef](#)]
46. Fatema, I.; Kong, X.; Fang, G. Electricity demand and price forecasting model for sustainable smart grid using comprehensive long short term memory. *Int. J. Sustain. Eng.* **2021**, *14*, 1714–1732. [[CrossRef](#)]

47. Sagheer, A.; Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **2019**, *323*, 203–213. [[CrossRef](#)]
48. Cui, Z.; Ke, R.; Pu, Z.; Wang, Y. Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transp. Res. Part C Emerg. Technol.* **2020**, *118*, 102674. [[CrossRef](#)]
49. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
50. Luong, M.-T.; Pham, H.; Manning, C.D. Effective approaches to attention-based neural machine translation. *arXiv* **2015**, arXiv:1508.04025.
51. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 2048–2057.
52. Mnih, V.; Heess, N.; Graves, A. Recurrent models of visual attention. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, USA, 8–13 December 2014; pp. 2204–2212.
53. Song, S.; Lan, C.; Xing, J.; Zeng, W.; Liu, J. An end-to-end spatio-temporal attention model for human action recognition from skeleton data. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
54. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 922–929.
55. Dewan, A.M.; Kankam-Yeboah, K. Using synthetic aperture radar (SAR) data for mapping river water flooding in an urban landscape: A case study of Greater Dhaka, Bangladesh. *J. Jpn. Soc. Hydrol. Water Resour.* **2006**, *19*, 44–54. [[CrossRef](#)]
56. Rahman, M.; Ningsheng, C.; Islam, M.M.; Mahmud, G.I.; Pourghasemi, H.R.; Alam, M.; Rahim, M.A.; Baig, M.A.; Bhattacharjee, A.; Dewan, A. Development of flood hazard map and emergency relief operation system using hydrodynamic modeling and machine learning algorithm. *J. Clean. Prod.* **2021**, 127594. [[CrossRef](#)]
57. Schober, P.; Boer, C.; Schwarte, L. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 5. [[CrossRef](#)]
58. Van Buuren, S.; Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67. [[CrossRef](#)]
59. Moriasi, D.N.; Gitau, M.W.; Pai, N.; Daggupati, P. Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Am. Soc. Agric. Biol. Eng.* **2015**, *58*, 1763–1785.
60. Lasheras, F.S.; Rodríguez, J.G.; Nieto, P.J.G.; García-Gonzalo, E.; Valverde, G.F. A Multivariate Approach to Time Series Forecasting of Copper Prices with the Help of Multiple Imputation by Chained Equations and Multivariate Adaptive Regression Splines. In Proceedings of the International Workshop on Soft Computing Models in Industrial and Environmental Applications, Burgos, Spain, 16–18 September 2020; pp. 691–701.