

Article

A Comparative Analysis of Advanced Machine Learning Techniques for River Streamflow Time-Series Forecasting

Antoifi Abdoulhalik * and Ashraf A. Ahmed

Department of Civil and Environmental Engineering, Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK; ashraf.ahmedmohamed@brunel.ac.uk

* Correspondence: antoifi.abdoulhalik@brunel.ac.uk

Abstract: This study examines the contribution of rainfall data (RF) in improving the streamflow-forecasting accuracy of advanced machine learning (ML) models in the Syr Darya River Basin. Different sets of scenarios included rainfall data from different weather stations located in various geographical locations with respect to the flow monitoring station. Long short-term memory (LSTM)-based models were used to examine the contribution of rainfall data on streamflow-forecasting performance by investigating five scenarios whereby RF data from different weather stations were incorporated depending on their geographical positions. Specifically, the All-RF scenario included all rainfall data collected at 11 stations; Upstream-RF (Up-RF) and Downstream-RF (Down-RF) included only the rainfall data measured upstream and downstream of the streamflow-measuring station; Pearson-RF (P-RF) only included the rainfall data exhibiting the highest level of correlation with the streamflow data, and the Flow-only (FO) scenario included streamflow data. The evaluation metrics used to quantitatively assess the performance of the models included the RMSE, MAE, and the coefficient of determination, R^2 . Both ML models performed best in the FO scenario, which shows that the diversity of input features (hydrological and meteorological data) did not improve the predictive accuracy regardless of the positions of the weather stations. The results show that the P-RF scenarios yielded better prediction accuracy compared to all the other scenarios including rainfall data, which suggests that only rainfall data upstream of the flow monitoring station tend to make a positive contribution to the model's forecasting performance. The findings evidence the suitability of simple monolayer LSTM-based networks with only streamflow data as input features for high-performance and budget-wise river flow forecast applications while minimizing data processing time.

Keywords: river flow forecasting; water resource management; artificial intelligence; machine learning; LSTM networks; Central Asia



Citation: Abdoulhalik, A.; Ahmed, A.A. A Comparative Analysis of Advanced Machine Learning Techniques for River Streamflow Time-Series Forecasting. *Sustainability* **2024**, *16*, 4005. <https://doi.org/10.3390/su16104005>

Academic Editors: Bahram Gharabaghi, Hossein Bonakdari and Silvio José Gumiere

Received: 12 March 2024

Revised: 29 April 2024

Accepted: 30 April 2024

Published: 10 May 2024



Copyright: © 2024 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/>).

1. Introduction

The expected population growth in the Syr Darya region will increase water demand in the forthcoming years [1], while climate change will further exacerbate water stress in the forthcoming years [2]. Climate change-induced effects mostly include temperature rises, which will increase the water's need for irrigation due to higher evapotranspiration rates [3,4] and potentially reduce the streamflow in the Syr Darya River due to glacier loss in the Tien Shan Mountains, which greatly sustains the streamflow through glaciers and snow melting [5–7]. Notwithstanding, anthropogenic activities such as the construction of dams and irrigation canals have already significantly reduced river streamflow, especially after the 1970s, and thus subsequently impacted the water supply and water-demand patterns in the Syr Darya region. These pluralities of multifaceted factors warrant the need for effective river streamflow forecasting tools to improve planning and water management efficiency in the Syr Darya region for the benefit of human well-being and environmental sustainability in light of the known and expected climate change impacts.

River streamflow prediction has recently gained significant research relevance in many studies over the past few years. The application of streamflow modeling has been applied to a wide range of applications, including, but not limited to, agriculture, water supply, recreation, flood defense, and energy production. Streamflow forecasting can be classified as short-term and long-term forecasting. Short-term forecasts correspond to real-time predictions (i.e., daily or hourly outputs) and are more suitable for flood-management applications. Long-term forecasts may include weekly, monthly, or annual predictions, and are more relevant for reservoir operation, irrigation-system management, and hydropower generation [8].

Additionally, three models have been adopted for the forecasting of river streamflow, namely the conceptual, physical, or data-driven models [9]. Conceptual models are based on correlations between various observable hydrological and meteorological variables. The dependency on the observed variables of this model diminishes the practicality of this approach in catchments where data availability is limited. Physical-based models have been widely used for river-flow predictions, mostly due to better process understanding, increased computing power, improved observation systems, which include flow data for model calibration, and remote sensing [10]. Despite this substantial improvement, physical models require a large amount of data, including, but not limited to, that of land-use, slopes, soils, and climate [11], and the limited process understanding of subsurface data and inadequate grid resolution represents a considerable constraint [10].

The increasingly wide availability of hydrological and meteorological data, in addition to their rapid temporal evolution, often leads to a preference for data-driven models over physically based or conceptual forecasting models. Data-driven models overcome these limitations and produce reliable predictions of physical systems, even without prior knowledge of the underlying physical relationships and the catchment information. Data-driven models, often incorporating machine learning algorithms, have widely been applied to streamflow forecasting in recent decades. Thanks to the considerable development in computation capabilities and data availability [12], machine learning models, particularly those based on artificial neural networks (ANNs), have been widely used to unravel the non-linear input–output relationship. Specifically, a particular type of ANN, referred to as a long short-term memory network (LSTM), has gained increasing attention in the field of hydrology, given its strong learning ability to process sequential time-series data.

In the realm of river streamflow forecast application, LSTM-based networks have successfully used and yielded promising outcomes e.g., [13–15]. Cheng et al. compared ANNs and LSTM in long lead-time forecasting capability in the Nan River Basin and Ping River Basin and demonstrated that the LSTM model outperformed the ANN model in long lead-time daily forecasting [16]. Xu et al. assessed the performance of LSTM networks in the task of the 10-day average flow predictions and the daily flow predictions for the Hun River and Upper Yangtze River basins, respectively [17]. Their study evidenced the superiority of LSTM over traditional hydrological and data-driven models tested, which included the Soil and Water Assessment Tool (SWAT) SWAT, the Xinjiang model (XAJ), the multiple linear regression model (MLR), and back-propagation neural networks (BP). Mehedi et al. used the LSTM neural network to forecast river discharge and demonstrated its higher performance compared to other neural network regression models, including for longer lead periods [18].

The performance of LSTM networks has also proved satisfactory when compared to other advanced machine learning algorithms [19]. Dehghani et al. compared LSTM, CNN, and Convolutional long short-term memory (ConvLSTM) in hourly streamflow prediction in two rivers in Malaysia, namely the Kelantan and Muda River basins [20]. They showed that all three deep learning methods performed with high accuracy in predicting streamflow, but LSTM outperformed CNN and ConvLSTM in small basins with well-spatially distributed rainfall stations, while it underperformed in moderate to high streamflow and large river basins. Le et al. demonstrated the superiority of LSTM-based models over a feed-forward neural network (FFNN), a CNN model adapted from

a performance and stability standpoint [21]. While their study showed that LSTM-based models could achieve remarkable forecasts even in the presence of upstream dams and reservoirs, they also evidenced that the complexity of the RNN family models—Stacked LSTM and BiLSTM models—is not accompanied by performance improvement.

In a more recent study, Le et al. also evidenced that only a single hidden layer would be sufficient for LSTM/GRU models to deliver highly accurate predictions while minimizing the data processing time [22]. In addition, their study also demonstrated that the use of multiple-input data types—rainfall and streamflow data—does not necessarily yield an increase in prediction accuracy when compared to using only one data type (only streamflow). In other words, they concluded that the inclusion of rainfall data does not improve streamflow forecasting performance. However, it is very important to examine the contribution of rainfall data on model performance while considering the geographical locations of the weather stations relative to streamflow-measuring stations, as some may have more impact than others depending on their positions—upstream or downstream.

This study aims to examine the contribution of rainfall data in improving the model's forecasting accuracy by considering different sets of scenarios, which include rainfall data from different weather stations based on their geographical locations with respect to the flow monitoring station. Specifically, the streamflow data were used as an input feature in all the scenarios, whereas rainfall data (RF) were only used in four scenarios, such that all All-RF included rainfall data collected in all 11 stations; Up-RF and Down-RF included only the rainfall data measured upstream and downstream of the streamflow-measuring station; and P-RF only included the rainfall data exhibiting the highest level of correlation with the streamflow data (Pearson coefficient > 0.3), which was only the case for Naryn and Tian-Shan stations. Also, two LSTM-based models were tested, namely LSTM and BiLSTM, and their forecasting performance was compared quantitatively and qualitatively. The outcomes of this study will provide a first insight into the potential applicability of LSTM-based models in forecasting streamflow data in the Syr Darya River and better understand the influence of hydrological data properties on the model's predictive forecasting performance.

2. Study Area and Data Acquisition

The Syr Darya River is the second major river in Central Asia and is a transboundary river shared by four countries, namely Kyrgyzstan, Kazakhstan, Tajikistan, and Uzbekistan. The catchment area of the Syr Darya extends over around 219,000.0 km². Along with the Amu Darya, which originates in Tajikistan, the Syr Darya is the second major tributary of the Aral Sea and is one of the most anthropologically altered water systems in the world.

The Syr Darya possesses two tributaries, namely the Naryn and the Kara Darya. While the former originates in the high mountainous areas of Central Tianshan, the latter originates in the Fergana/Alai Range. In the Naryn catchment basin, the total annual precipitation ranges between 280 and 450 mm, where the largest amount of annual precipitation occurs during spring and summer. In the Kara Darya, the total annual precipitation ranges between 350 and 1050 mm, where the largest portion falls in winter and spring while very little falls during the summer period [23]. These two tributaries eventually merge into the Syr Darya in the Fergana Valley of Uzbekistan.

The coupled impact of climate change and water governance contribute towards exacerbating tensions over the already limited water resources within and across the borders, particularly given the conflicting nature of the water demands. While water is mainly needed for energy-production (hydropower) purposes in winter in Kirgizstan, water is mostly abstracted for agriculture (water-intensive crops) during the vegetation season (spring/summer) in Uzbekistan. This conflicting interest in water use combined with river streamflow reduction will further compromise the relationship between the two riparian states in the future.

Figure 1 provides a map showing the demonstration catchments investigated herein, as well as the location of the hydrological and meteorological stations. The mean flow data

from the Syr Darya River were provided by the World Meteorological Organization (WMO) and were collected as part of the framework of the First GARP Global Experiment (FGGE).

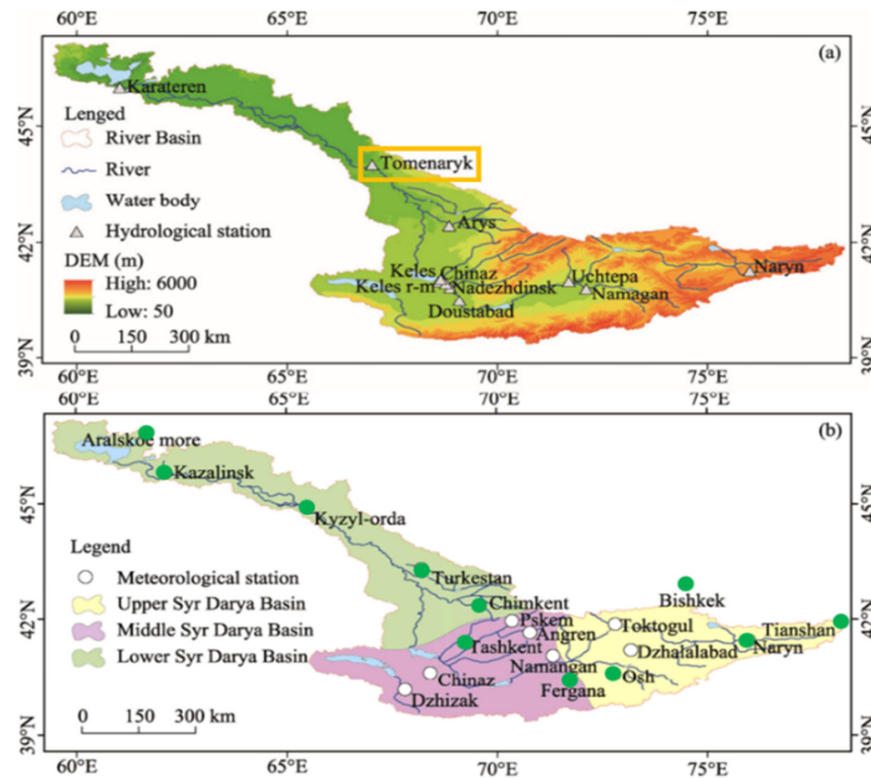


Figure 1. Overview of the Syr Darya River Basin as well as the location of the hydrological stations (a) and the meteorological stations in green (b). Figure adapted from Figure 1 in [24].

The flow data were collected at the Tomenaryk hydrological station, located in Kazakhstan (Figure 1a), and were recorded daily from January 1936 to December 1986. Data collected before 1957 were discarded from the dataset due to some recurring portions of missing values. The resulting dataset comprised 30 years of streamflow data spanning from January 1957 to December 1986. The entire length of the sequence of the flow dataset did not present any missing values; thus, no deletion or imputation processes were required. The mean flow data were then converted into monthly data for the challenge of forecasting. The graph of the flow data of the Syr Darya River is shown in Figure 2.

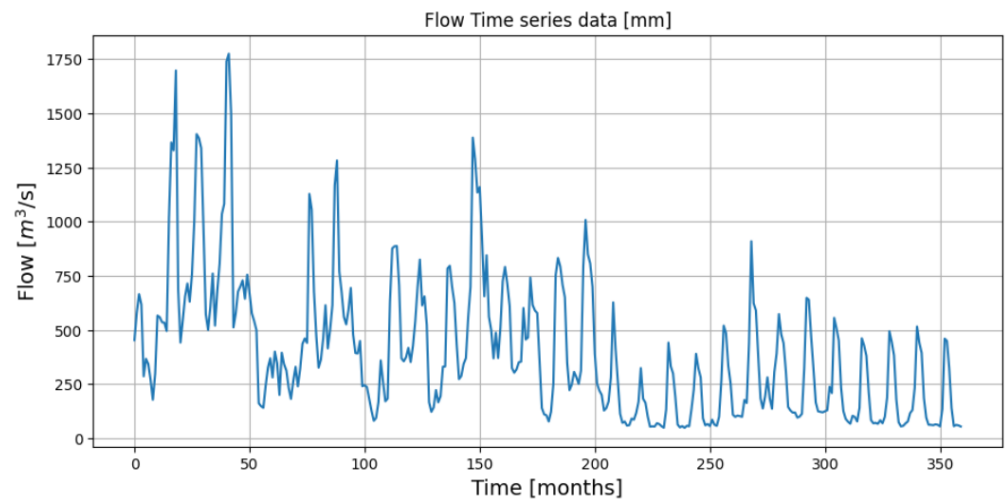


Figure 2. Monthly streamflow data for the Syr Darya River at the Tomenaryk hydrological station.

The rainfall data were sourced from the Global Historical Climate Network (GHCN) database. The rainfall data were collected at 11 stations located throughout the catchment area on the upstream and downstream sides of the Tomenaryk station. The locations of the meteorological stations are highlighted by the green dots in Figure 1b. Missing values were recorded as zeros (i.e., no precipitation). The graph of the rainfall data for the Naryn and Tien-Shan stations is shown in Figure 3 (other rainfall graphs are not shown for brevity).

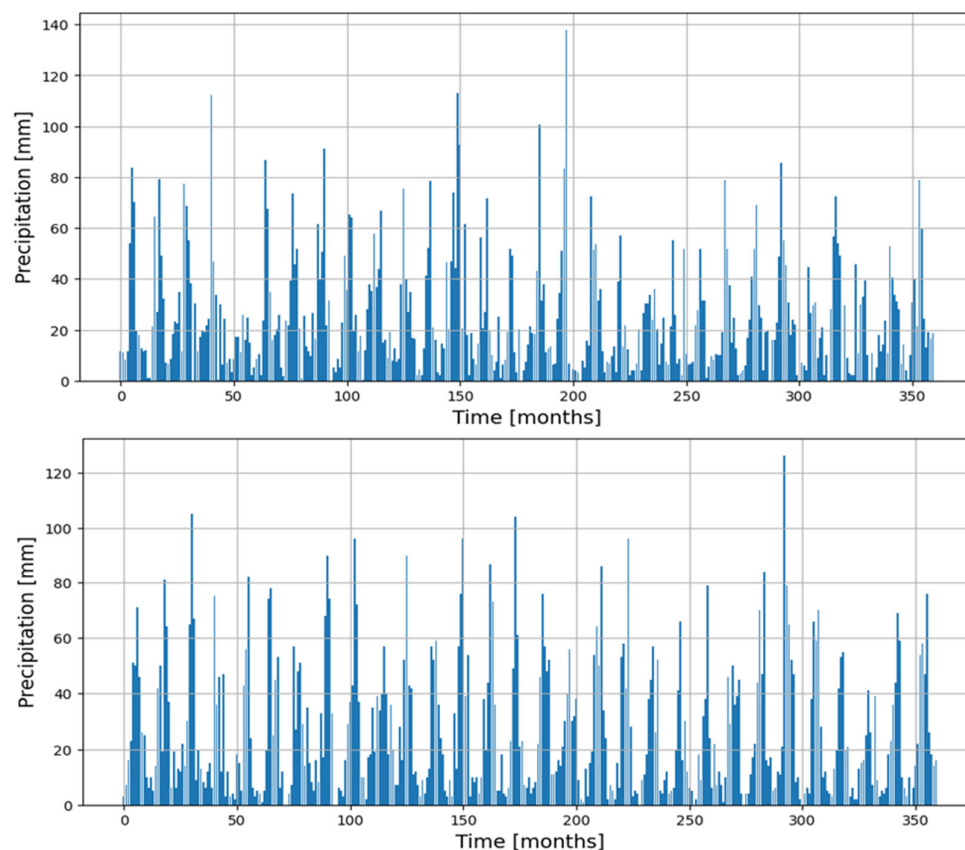


Figure 3. Monthly rainfall data in the Naryn (top) and Tian Shan (bottom) stations.

3. Materials and Method

3.1. Data Preparation

The complete dataset comprised additional information such as records of the country, station name, latitude and longitude coordinates, catchment-area size, altitude, and the date of the observation. These additional data were not required and were therefore removed to prepare the data for the model-training and -validation phases.

The dataset from 1 January 1957 to 31 December 1986 was subdivided into two portions, referred to as the training and validation sets. In this study, the data were subsequently split into training and validation sets using a ratio of 90:10. Specifically, the data recorded from 1 January 1957 to 31 December 1983 were used for the training phase, and the data recorded from 1 January 1984 to 31 December 1986 were used for the validation phase.

Sliding windows were used to sample the time-series dataset. The sliding window technique consists of breaking down the observations of the entire time series into pairs of inputs and outputs, thereby transforming the forecasting problem into a supervised learning problem. For a given window size n (i.e., the number of time steps), a subsample of n data measures is read as an input to generate a predicted output that is to be evaluated using real measurements. In other words, the first predictions from past iterations are used to predict a new output. The second prediction from the n past iterations is always discarded. The impact of the window size on the forecasting accuracy was examined by

testing different window-size values in the models. Note that the data were not shuffled in the partitioning phases of the datasets nor during the training process of the streamflow-forecasting prediction model.

3.2. Flow Forecasting Approach

In this study, LSTM and BiLSTM networks were compared in the task of forecasting monthly river flow using time-series data with up to 1 month of lead time. Long short-term memory (LSTM) is a recurrent artificial neural network (RNN) architecture that is widely used in real deep learning forecasting problems. The cells that compose LSTM networks can capture long-term dependencies in sequences while minimizing the gradient vanishing/exploding problem. LSTM networks involve the processing of data in a single direction (forward), i.e., important information is either stored or predicted from previous data. A BiLSTM, on the other hand, extends the capabilities of LSTM by incorporating information from both past and future time steps. A BiLSTM involves the simultaneous processing of data in both forward and backward directions through two distinct LSTM layers: one processes the sequence in the forward direction, and the other processes it in the backward direction. The outputs of the two LSTM layers are typically concatenated at each time step to form the final output. Figure 4 shows the architecture of a single LSTM cell and a bidirectional LSTM network.

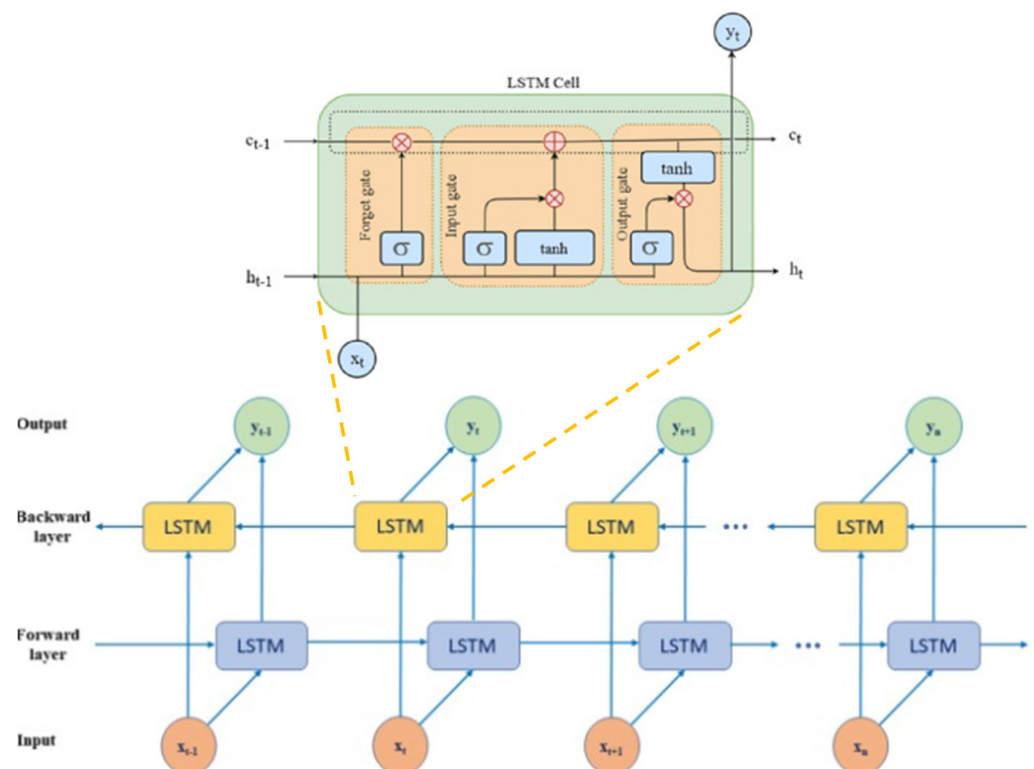


Figure 4. Architecture of a single LSTM cell (**top**) and a bidirectional LSTM network (**bottom**).

Five scenarios were established using hydrological data recorded over 30 years in the Syr Darya River Basin to examine the effect of input variables and hyperparameter selection on model flow-forecasting performance. While the streamflow data were used as an input feature in all the scenarios, the rainfall data were only used in only four scenarios: the RF included rainfall data collected in all 11 stations; Up-RF and Down-RF included only the rainfall data measured upstream and downstream of the streamflow-measuring station; and P-RF only included the rainfall data exhibiting the highest level of correlation with the streamflow data (Pearson coefficient > 0.3), which was only the case for the Naryn and Tian-Shan stations.

The impact of the main hyperparameters was examined so that the design of the network architecture that yielded the best performance was adopted for the flow-forecasting challenge. These parameters included the number of units within the LSTM cell and the batch size. The number of units varied from 64 to 512 (in terms of powers of two), while the batch size varied from 2 to 8 (also in terms of powers of two). Preliminary runs showed that batch size values exceeding 8 yielded a considerable loss in prediction accuracy.

The correct set-up of the hyperparameters (training parameters) is also an important step affecting the performance of the model. Amongst the most existing common optimization techniques, which include trial-and-error, grid, random, and probabilistic techniques, the trial-and-error approach was adopted herein to find the optimal values of the hyperparameters.

The most prevalent hyperparameters include the optimization algorithm, loss function and learning rate, batch size, window size, and the number of epochs. In this study, the number of units within the LSTM cell and the batch size were varied to determine the most suitable network architecture design to yield the best forecasting performance. The LSTM model comprised a single hidden layer. In all the experiments, the activation function within the LSTM was maintained as its standard value, i.e., the tanh function. Common practice suggests that the number of neurons in the hidden layer should be a power function of 2, so the number of neurons varied from 64 to 512. Likewise, the batch size was varied from 2 to 8 (also in terms of powers of two). Preliminary runs showed that batch-size values exceeding 8 yielded a loss in prediction accuracy.

The other model hyperparameters were kept constant in all cases. Specifically, the Adam optimization algorithm [25] was selected along with its standard learning-rate value of 0.001. The Adam optimization algorithm is one of the most prevalent algorithms used in deep learning studies and has demonstrated good performance [26,27]. The Huber loss was adopted as a loss function in all the models. The number of epochs was set to 250. Despite their known aptitude to prevent overfitting issues in some cases, no early stopping and/or dropout procedures [28] were implemented in our models. All models were developed using the programming language Python along with several open-source software libraries [29], including TensorFlow v2.16.1 [30], Numpy [31], Keras [32], and Pandas [33]. All figures were plotted using the matplotlib library [34].

The contribution of rainfall data in improving the model's forecasting accuracy was examined by creating different sets of scenarios which included rainfall data from different weather stations based on their geographical locations. Specifically, the streamflow data were used as an input feature in all the scenarios, whereas rainfall data (RF) were only used only four scenarios, such that all All-RF included rainfall data collected in all 11 stations; Up-RF and Down-RF included only the rainfall data measured upstream and downstream of the streamflow-measuring station; and P-RF only included the rainfall data exhibiting the highest level of correlation with the streamflow data (Pearson coefficient > 0.3), which was only the case for Naryn and Tian-Shan stations (Figure 5). Table 1 describes the various rainfall data incorporated in each scenario.

Table 1. Summary of the different scenarios investigated in this study.

Rainfall Data	All-RF	Up-RF	Down-RF	P-RF	FO
Chimkent	Yes	Yes	No	No	No
Turkestan	Yes	Yes	No	No	No
Bishkek	Yes	Yes	No	No	No
Naryn	Yes	Yes	No	Yes	No
Osh	Yes	Yes	No	No	No
Tien-Shan	Yes	Yes	No	Yes	No
Fergana	Yes	Yes	No	No	No
Tashkent	Yes	Yes	No	No	No
Aralskoe-more	Yes	No	Yes	No	No

Table 1. Cont.

Rainfall Data	All-RF	Up-RF	Down-RF	P-RF	FO
Kazalinsk	Yes	No	Yes	No	No
Kyzyl-orda	Yes	No	Yes	No	No

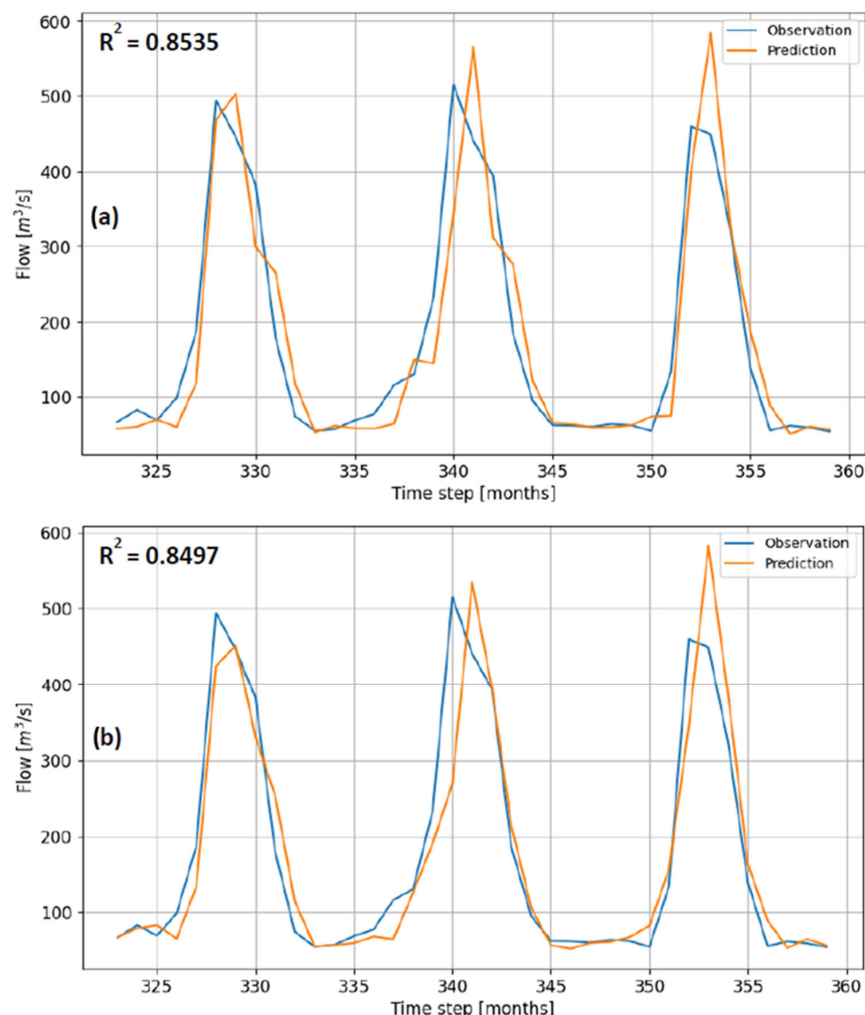


Figure 5. Comparison between forecasted and observed streamflow in the Syr Darya River during the validation period in the FO scenario, (a) LSTM, and (b) BILSTM.

3.3. Performance Criterion

The performances of the models were objectively and quantitatively compared during the training, validation, and testing phases. Three performance metrics were selected to evaluate the forecasting performance of the proposed DL models. These metrics included the mean absolute error (MAE), the root mean square error (RMSE), and the coefficient of determination (R^2). These performance indicators are commonly used to objectively evaluate the correlation between two data series both in statistical and hydrological models. The MAE and RMSE are expressed as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}$$

where x_i and y_i correspond to the observed and forecasted streamflow at time t , respectively; \bar{x} refers to the mean of observed discharges; and n is the total number of observation points. The coefficient of determination, R^2 , which characterizes the goodness of the fit between the predicted and observed values, was also considered. The value of R^2 ranges between 0 and 1, whereby values closer to 1 denote a good fit.

4. Results and Discussion

The analysis of the effects of rainfall on the model's forecasting performance required the comparison of the outputs achieved in the various scenarios including rainfall data to the results of the scenario without rainfall (FO), which was thus considered as the base case. The forecasting accuracy of the LSTM and BILSTM models in the FO scenario was examined by comparing the performance metrics at the validation stage. Figure 5 shows the comparison between predictive and observation curves in the FO scenario, in both LSTM and BILSTM cases. Overall, the streamflow was relatively well forecasted by the two models during the validation phase. The agreement between observed and forecasted streamflow was relatively similar, with comparable R^2 values. In the two cases, the low flow values were better predicted than the peak flows, which were slightly overestimated, apart from the first peak flow in the BILSTM case, which was slightly underestimated.

Figure 6 shows the comparison between observation and prediction in the All-RF, Up-RF, Down-RF, and P-RF scenarios in the LSTM case. The prediction curves appear less smothered and present intermittent fluctuations compared to the FO scenarios. The P-RF scenarios visibly yielded a slightly better fit compared to the other scenarios, which are also reflected in a slightly greater R^2 value than that recorded in the other scenarios. Neither scenario could reproduce the peak values accurately, all being slightly shifted forward over time and slightly misestimated in magnitude. The first peak was overestimated in the All-RF scenario but underestimated in the Up-RF scenario. The closest fit to the first peak was observed in the P-RF scenario. The second peak was overestimated in the Up-RF and P-RF, and was visibly better in the All-RF scenario. The third peak was also overestimated in the Up-RF, and the All-RF, but slightly less in the latter scenario. The worst prediction–observation fit was exhibited in the Down-RF scenario, where all peaks were noticeably overestimated. On the other hand, all scenarios presented some odd fluctuations during the lows, especially in the All-RF, Up-RF, and Down-RF scenarios. The P-RF scenarios also exhibited similar behavior, but only in the second low period ($t = 335$ – 350 months), while a relatively smoother match was observed in the first low period.

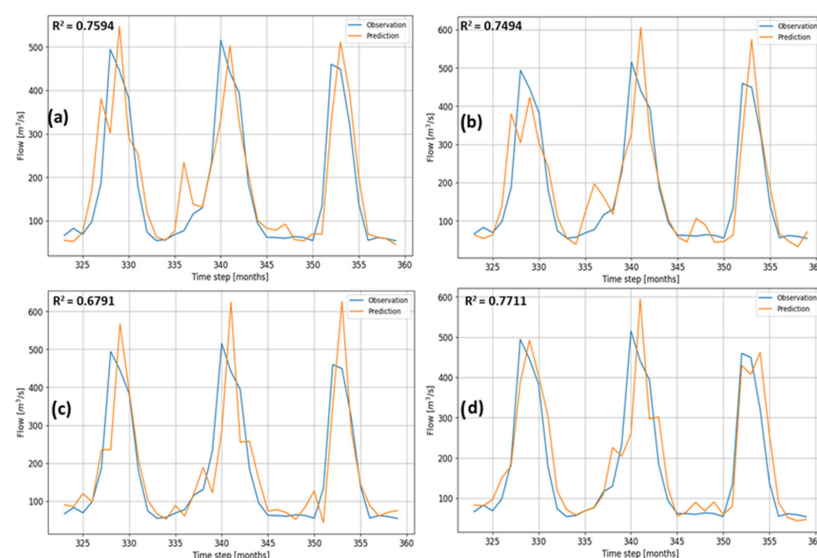


Figure 6. Comparison between forecasted and observed streamflow in the LSTM case during the validation period in (a) the All-RF scenario, (b) the Up-RF scenario, (c) the Down-RF scenario, and (d) the P-RF scenario.

Figure 7 shows the comparison between the observation and prediction in the All-RF, Up-RF, Down-RF, and P-RF scenarios in the BILSTM case. Similarly, the predicted curves exhibited some odd fluctuation patterns compared to the FO scenario. However, the forecasted curve of the P-RF scenario did not present such discrepancies, but rather presented a smoother behavior, as observed in the FO scenario. Nonetheless, the All-RF scenario produced the best overall fit with an R^2 value of 0.8598, while the worst overall prediction–observation fit was observed in the Down-RF scenario, with an R^2 of 0.6425.

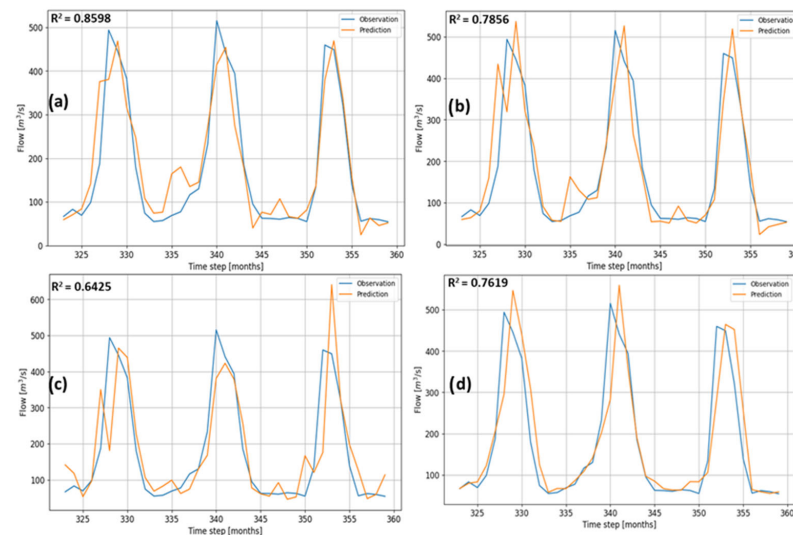


Figure 7. Comparison between forecasted and observed streamflow in the BILSTM case during the validation period in (a) the All-RF scenario, (b) the Up-RF scenario, (c) the Down-RF scenario, and (d) the P-RF scenario.

Again, all multivariate scenarios presented slight discrepancies in predicting the peak times, all being slightly shifted forward time-wise, and slightly misestimated in magnitude. The first peak was overestimated in the Up-RF and P-RF scenarios but underestimated in the All-RF and Down-RF scenarios. The second peak was overestimated in the Up-RF and P-RF scenarios, albeit slightly less in the former scenario. The Down-RF and All-RF scenarios both presented underestimated peak values. The third peak was overestimated in the Up-RF and Down-RF scenarios, albeit more in the latter than the former. The P-RF reproduced a somewhat similar magnitude to the observed values, but shifted forward. The All-RF scenario exhibited a relatively good match throughout the occurrence of the peak from a timing and magnitude standpoint, but the tip presented a slightly sharper shape.

Figure 8 shows the performance metrics recorded in all the scenarios in the LSTM and BILSTM cases during the validation phase. In the LSTM case, the FO scenario exhibited the best forecasting performance among all the scenarios, while the Down-RF scenario yielded the least accurate predictions. Specifically, the RMSE values were $75.59 \text{ m}^3/\text{s}$, $77.14 \text{ m}^3/\text{s}$, $87.29 \text{ m}^3/\text{s}$, $73.73 \text{ m}^3/\text{s}$, and $58.97 \text{ m}^3/\text{s}$ in the All-RF, Up-RF, Down-RF, P-RF, and FO scenarios, respectively. In other words, the RMSE was increased by about 28%, 31%, 48%, and 25% compared to the FO scenario in the All-RF, Up-RF, Down-RF, and P-RF scenarios, respectively. The MAE values were $50.72 \text{ m}^3/\text{s}$, $52.46 \text{ m}^3/\text{s}$, $57.09 \text{ m}^3/\text{s}$, $48.21 \text{ m}^3/\text{s}$, and $40.88 \text{ m}^3/\text{s}$ in the All-RF, Up-RF, Down-RF, P-RF, and FO scenarios, respectively. This means that the MAE was increased by 24%, 28%, 40%, and 18% in the All-RF, Up-RF, Down-RF, P-RF, and FO scenarios, respectively. Regarding the overall fit, the highest and lowest R^2 values were in the FO and Down-RF scenarios, respectively, with 0.8535 and 0.6791 recorded in the former and the latter. The scenarios All-RF, Up-RF, and P-RF yielded comparable fits.

In the BILSTM case, the All-RF and P-RF were able to achieve performance metrics more comparable to the FO scenario, albeit the MAE was still smaller in the latter scenario. The worst performance was recorded in the Down-RF scenario, as observed in the LSTM

case. The RMSE was $57.70 \text{ m}^3/\text{s}$, $71.35 \text{ m}^3/\text{s}$, $92.13 \text{ m}^3/\text{s}$, $57.53 \text{ m}^3/\text{s}$, and 59.74 in the All-RF, Up-RF, Down-RF, P-RF and FO scenarios, respectively. The results show that multivariate scenarios containing multiple input features, as opposed to the univariate scenario containing a single data type (flow), did not improve the predictive accuracy of either model, regardless of the positions of the weather stations. In other words, feeding an ML model with a large disparity in input variables (hydrological and meteorological data) does not necessarily lead to a gain in the model's forecasting performance. Thus, these findings show that rainfall may not necessarily be among the controlling input variables for streamflow forecasting in the Syr Darya River.

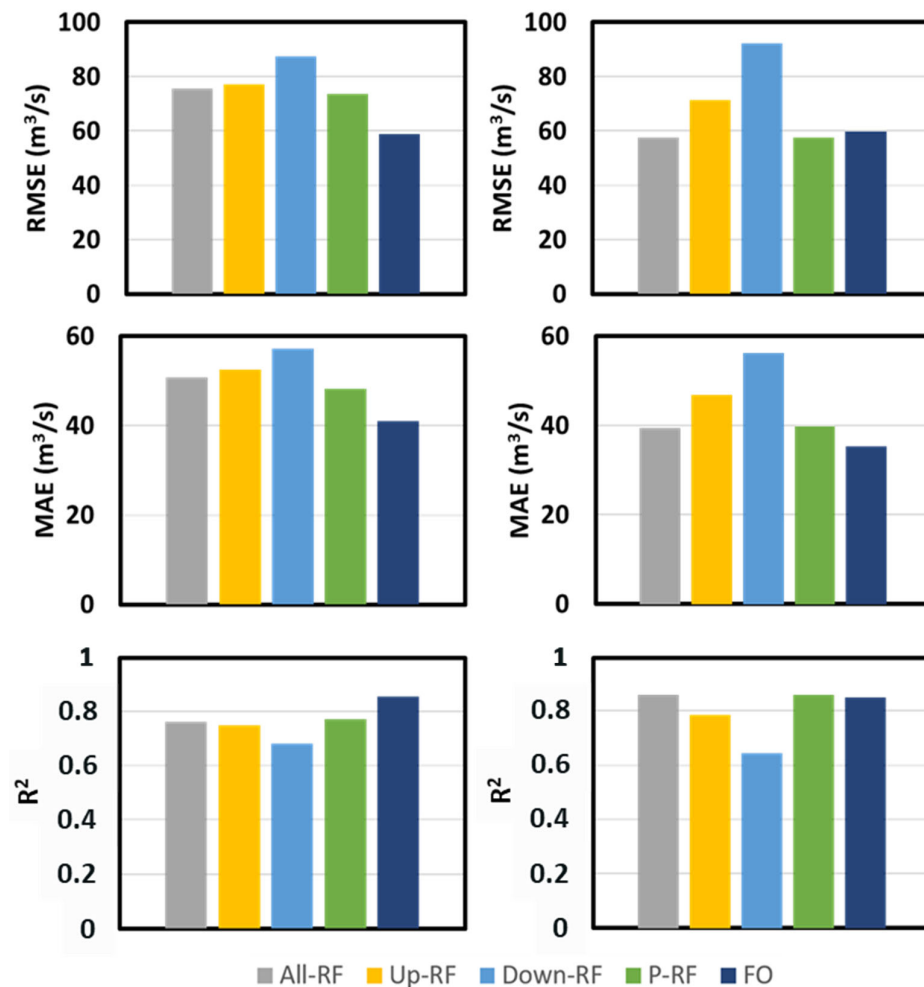


Figure 8. Evolution of the RMSE, MAE, and R^2 in all the scenarios in the LSTM case (left) and BILSTM case (right).

Hence, the All-RF and P-RF scenarios induced very comparable model performances in the FO scenario. However, the Up-RF scenario induced an increase of 19% compared to the FO scenario, and the Down-RF scenario achieved an RMSE 54% greater. Regarding the MAE values, the BILSTM achieved similar values in the All-RF and P-RF scenarios, with $39.40 \text{ m}^3/\text{s}$ and $39.80 \text{ m}^3/\text{s}$. In other words, the MAE achieved in these two scenarios was greater than in the FO scenario, by 12% and 13% in the former and the latter, respectively. The Up-RF and Down-RF scenarios yielded higher MAE values compared to the FO scenario, with a respective 33% and 59.40% increase. The overall fit was very comparable in All-RF, P-RF, and FO scenarios, with R^2 values of 0.8598, 0.8606, and 0.8497, respectively. The scenarios of Up-RF and Down-RF induced a lower fit, with an R^2 of 0.7857 and 0.6426, respectively.

These findings suggest that selecting only the rainfall data upstream of the flow monitoring station tends to make a positive contribution to the model's forecasting performance

while, also, including data from downstream rain gauges complicates training and thus may reduce the model's predictive accuracy, which is particularly important in the context of data scarcity, as is the case in most developing countries.

Also, in comparison to the LSTM case, the BILSTM model achieved RMSE values of 24%, 7.5%, and 22% in All-RF, Up-RF, and P-RF scenarios. However, the Down-RF was increased by 5.5% compared to the LSTM model, while the FO scenario yielded relatively similar outputs in both models. Also, the BILSTM achieved 22%, 11%, 17%, and 14% smaller MAE values than the LSTM model in All-RF, Up-RF, P-RF, and FO scenarios, respectively. However, the Down-RF scenario achieved similar MAE values in the two models. The coefficient of determination, R^2 , recorded in the BILSTM case was higher in All-RF and P-RF scenarios. In the Up-RF and FO scenarios, the R^2 values were comparable, albeit slightly higher in the Up-RF scenario. However, the Down-RF scenario yielded a slightly worse fit in the BILSTM.

Amongst all the multivariate scenarios, the results show that the models performed the least well in the Down-RF scenario. In other words, the forecasting performance was substantially reduced when the models were fed with rainfall data from stations located downstream of the streamflow monitoring station. In the LSTM case, the results also show that the models exhibited comparable performances in the All-RF, the Up-RF, and the P-RF scenarios, although the model performed slightly better in the latter scenario. This means that the LSTM model trained with rainfall data from all the stations yielded comparable forecasts as when trained with rainfall data only from upstream stations, which is particularly important in the context of data scarcity, as is the case in most developing countries.

In the BILSTM case, the performance achieved by the model was equivalent in the All-RF and P-RF scenarios but was worsened in the Up-RF scenario. Furthermore, the results suggest that the two models trained with only rainfall data with some level of correlation with streamflow data (P-RF) also helped improve the forecasting performance compared to when feeding with all available rainfall data. The results also show that the forecast was the most satisfactory in the FO scenario, regardless of the multitude of rainfall datasets considered. In other words, none of the multivariate LSTM-based models could outperform the univariate LSTM-based models, i.e., considering only streamflow data. While Le et al. proved that incorporating rainfall data into their LSTM forecasting model did not significantly improve forecasting performance [22], our results show that the multivariate scenarios containing multiple input features as opposed to the univariate scenario containing a single data type (flow) did not noticeably improve the predictive accuracy of either forecasting model, regardless of the positions of the weather stations with respect to the position of the flow monitoring station. Hence, these results call into question the relevance of including rainfall data as a predictor in streamflow-forecasting applications in the Syr Darya River basin.

5. Summary and Conclusions

In this study, advanced machine learning models such as LSTM and bidirectional LSTM Networks were compared in the task of forecasting monthly river flow using time-series data. Five scenarios were established using hydrological data recorded over 30 years in the Syr Darya River Basin to examine the effect of input variables and hyperparameter selection on models' forecasting performances. While streamflow data were used as an input feature in all the scenarios, rainfall data were only used in four scenarios: RF included rainfall data collected in all 11 stations; Up-RF and Down-RF included only the rainfall data measured upstream and downstream of the streamflow-measuring station; and P-RF only included the rainfall data exhibiting the highest level of correlation with the streamflow data (Pearson coefficient > 0.3). The evaluation metrics used to quantitatively assess the performance of the models included the RMSE, MAE, and coefficient of determination R^2 . The findings of the study can be summarized as follows:

- The results show that multivariate scenarios containing multiple input features, as opposed to the univariate scenario containing a single data type (flow), did not improve the predictive accuracy of either model, regardless of the positions of the weather stations. In other words, feeding ML models with large disparities in input variables (hydrological and meteorological data) does not necessarily lead to a gain in model forecasting performance;
- The P-RF scenarios yielded better prediction accuracy compared to all the other scenarios including rainfall data. These findings suggest that selecting only the rainfall data upstream of the flow monitoring station tends to make a positive contribution to the model's forecasting performance. The inclusion of data from downstream rain gauges tends to complexify training and thus may reduce the model's predictive accuracy, which is particularly important in the context of data scarcity;
- The results show that rainfall may not necessarily be among the controlling input variables for streamflow forecasting in the Syr Darya River;
- The findings evidence the suitability of simple monolayer LSTM-based networks, especially BILSTM, with only streamflow data as input features for high-performance and budget-wise river-flow forecast applications while minimizing data processing time.

Author Contributions: Conceptualization, A.A.; Methodology, A.A.; Validation, A.A.; Formal analysis, A.A.; Data curation, A.A.; Writing—original draft, A.A.; Writing—review & editing, A.A.A.; Supervision, A.A.A.; Project administration, A.A.A.; Funding acquisition, A.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 101083481]. This research is part of the Horizon Europe WE-ACT project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

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

References

1. Siegfried, T.; Bernauer, T.; Guiennet, R.; Sellars, S.; Robertson, A.W.; Mankin, J.; Bauer-Gottwein, P.; Yakovlev, A. Will climate change exacerbate water stress in Central Asia? *Clim. Chang.* **2012**, *112*, 881–899. [[CrossRef](#)]
2. Liang, W.; Chen, Y.; Fang, G.; Kaldybayev, A. Machine learning method is an alternative for the hydrological model in an alpine catchment in the Tianshan region, Central Asia. *J. Hydrol. Reg. Stud.* **2023**, *49*, 101492. [[CrossRef](#)]
3. Aizen, V.B.; Kuzmichenok, V.A.; Surazakov, A.B.; Aizen, E.M. Glacier changes in the central and northern Tien Shan during the last 140 years based on surface and remote-sensing data. *Ann. Glaciol.* **2006**, *43*, 202–213. [[CrossRef](#)]
4. Ibatullin, S.; Yasinsky, V.; Mironenkov, A. *Impacts of Climate Change on Water Resources in Central Asia*; Sector Report; Eurasian Development Bank: Almaty, Kazakhstan, 2009; p. 44.
5. Chen, Y.; Li, W.; Fang, G.; Li, Z. Review article: Hydrological modeling in glacierized catchments of central Asia—Status and challenges. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 669–684. [[CrossRef](#)]
6. Golubtsov, V.; Lineitseva, A.; Merz, B.; Dukhovny, V.; Unger-Shayesteh, K. Receipt of water in the rivers of Northern slope of Jetisu Alatau because of glacier degradation. In Proceedings of the International Scientific Symposium, "Water in Central Asia", Tashkent, Uzbekistan, 24–26 November 2010; p. 87.
7. Lioubimtseva, E.; Henebry, G.M. Climate and environmental change in arid Central Asia: Impacts, vulnerability, and adaptations. *J. Arid Environ.* **2009**, *73*, 963–977. [[CrossRef](#)]
8. Yaseen, Z.M.; El-Shafie, A.; Jaafar, O.; Afan, H.A.; Sayl, K.N. Artificial intelligence-based models for streamflow forecasting: 2000–2015. *J. Hydrol.* **2015**, *530*, 829–844. [[CrossRef](#)]
9. Yaseen, Z.M.; Mohtar, W.H.M.W.; Ameen, A.M.S.; Ebtehaj, I.; Razali, S.F.M.; Bonakdari, H.; Salih, S.Q.; Al-Ansari, N.; Shahid, S. Implementation of univariate paradigm for streamflow simulation using hybrid data-driven model: Case study in tropical region. *IEEE Access* **2019**, *7*, 74471–74481. [[CrossRef](#)]
10. Hunt, K.M.R.; Matthews, G.R.; Pappenberger, F.; Prudhomme, C. Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 5449–5472. [[CrossRef](#)]
11. Jaiswal, R.K.; Ali, S.; Bharti, B. Comparative evaluation of conceptual and physical rainfall-runoff models. *Appl. Water Sci.* **2020**, *10*, 48. [[CrossRef](#)]

12. Ghaith, M.; Siam, A.; Li, Z.; El-Dakhkhni, W. Hybrid hydrological data-driven approach for daily streamflow forecasting. *J. Hydrol. Eng.* **2020**, *25*, 04019063. [[CrossRef](#)]
13. 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. [[CrossRef](#)]
14. Li, X.; Xu, W.; Ren, M.; Jiang, Y.; Fu, G. Hybrid CNN-LSTM models for river flow prediction. *Water Supply* **2022**, *22*, 4902–4919. [[CrossRef](#)]
15. Wegayehu, E.B.; Muluneh, F.B. Short-Term Daily Univariate Streamflow Forecasting Using Deep Learning Models. *Adv. Meteorol.* **2022**, *2022*, 1860460. [[CrossRef](#)]
16. 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. [[CrossRef](#)]
17. Xu, W.; Jiang, Y.; Zhang, X.; Li, Y.; Zhang, R.; Fu, G. Using long short-term memory networks for river flow prediction. *Hydrol. Res.* **2020**, *51*, 1358–1376. [[CrossRef](#)]
18. Mehedi, M.A.A.; Khosravi, M.; Yazdan, M.M.S.; Shabaniyan, H. Exploring Temporal Dynamics of River Discharge Using Univariate Long Short-Term Memory (LSTM) Recurrent Neural Network at East Branch of Delaware River. *Hydrology* **2022**, *9*, 202. [[CrossRef](#)]
19. Ahmed, A.A.; Sayed, S.; Abdoulhalik, A.; Moutari, S.; Oyedele, L. Applications of machine learning to water resources management: A review of present status and future opportunities. *J. Clean. Prod.* **2024**, *441*, 140715. [[CrossRef](#)]
20. Dehghani, A.; Moazam, H.M.Z.H.; Mortazavizadeh, F.; Ranjbar, V.; Mirzaei, M.; Mortezaei, S.; Ng, J.L.; Dehghani, A. Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches. *Ecol. Inform.* **2023**, *75*, 102119. [[CrossRef](#)]
21. Le, X.-H.; Nguyen, D.-H.; Jung, S.; Yeon, M.; Lee, G. Comparison of Deep Learning Techniques for River Streamflow Forecasting. *IEEE Access* **2021**, *9*, 71805–71820. [[CrossRef](#)]
22. Le, X.-H.; Van, L.N.; Nguyen, G.V.; Nguyen, D.H.; Jung, S.; Lee, G. Towards an efficient streamflow forecasting method for event-scales in Ca River basin, Vietnam. *J. Hydrol. Reg. Stud.* **2023**, *46*, 101328. [[CrossRef](#)]
23. Duethmann, D.; Peters, J.; Blume, T.; Vorogushyn, S.; Güntner, A. The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. *Water Resour. Res.* **2014**, *50*, 2002–2021. [[CrossRef](#)]
24. Bissenbayeva, S.; Abuduwaili, J.; Saparova, A.; Ahmed, T. Long-term variations in runoff of the Syr Darya River Basin under climate change and human activities. *J. Arid Land* **2021**, *13*, 56–70. [[CrossRef](#)]
25. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
26. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent neural networks for time series forecasting: Current status and future directions. *arXiv* **2019**, arXiv:1909.00590. [[CrossRef](#)]
27. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans. Smart Grid* **2019**, *10*, 841–851. [[CrossRef](#)]
28. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
29. Rossum, G. *Python Tutorial*; CWI (Centre for Mathematics and Computer Science): Amsterdam, The Netherlands, 1995.
30. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* **2016**, arXiv:1603.04467.
31. Van Der Walt, S.; Colbert, S.C.; Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30. [[CrossRef](#)]
32. Chollet, F. Keras. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 1 April 2018).
33. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 51–56.
34. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.