

Proceeding Paper

Artificial Neural Networks and Regression Modeling for Water Resources Management in Indus Basin [†]

Muhammad Imran ¹, Muhammad Danish Majeed ^{1,2,*}, Muhammad Zaman ^{1,*}, Muhammad Adnan Shahid ^{1,2}, Danrong Zhang ³, Syeda Mishal Zahra ^{1,2}, Rehan Mehmood Sabir ^{1,2}, Muhammad Safdar ^{1,2} and Zahid Maqbool ²

¹ Department of Irrigation & Drainage, Faculty of Agricultural Engineering & Technology, University of Agriculture, Faisalabad, Pakistan; email1@email.com (M.I.); email2@email.com (M.A.S.); email3@email.com (S.M.Z.); email4@email.com (R.M.S.); email5@email.com (M.S.)

² Agricultural Remote Sensing Lab-(ARSL)-NCGSA, University of Agriculture, Faisalabad, Pakistan; email6@email.com

³ College of Hydrology and Water Resources, Hohai University; email7@email.com

* Correspondence: m.imrankbr@gmail.com; Tel.: +92-346-3805181

† Presented at the 7th International Electronic Conference on Water Sciences, 15–30 March 2023; Available online: <https://ecws-7.sciforum.net>.

Abstract: A flood is a natural disaster. Heavy rainfall and overflow frequently cause enclosed land areas to fill with water, resulting in considerable loss of human life and property, including damage to buildings, bridges, electric supply networks transportation, and economic concern. This work was carried out in the Upper Indus Basin (UIB). We were developing an artificial intelligence model and forecasting the flood events in this study. Long-Short Term Memory (LSTM) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) were used in this study to forecast flood events. This study dataset was used from 1971–2009 and divided into training, testing, and forecasting from 1971–2004, 2005–2009, and 2010–2014 respectively. The best statistical analysis result was observed from the LSTM model which documented the value of root mean squared error (RMSE) at 22.79 and 35.05 for training and testing respectively. Hence, the results of the study highlight that the LSTM model was the most suitable among the artificial neural networks for flood event forecasts. This current study will help in the forecasting of high storms for effective water resources management.

Keywords: upper indus basin; flood forecasting; LSTM; SARIMA; water resources management

Citation: Imran, M.; Majeed, M.D.; Zaman, M.; Shahid, M.A.; Zhang, D.; Zahra, S.M.; Sabir, R.M.; Safdar, M.; Maqbool, Z. Artificial Neural Networks and Regression Modeling for Water Resources Management in Indus Basin. *Environ. Sci. Proc.* **2023**, *5*, x. <https://doi.org/10.3390/xxxxx>

Academic Editor(s):

Published: 15 March 2023



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Most Asian countries depend on agriculture, and Pakistan is most of the dependable country on agriculture. A huge part of the Gross Domestic Production (GDP) of Pakistan is based on agriculture and the agriculture sector of this country is mostly reliant on irrigation water generated by the upper Indus basin. The water scarcity report of the international monetary fund (IMF) ranked Pakistan in the third position among water scarcity in the year 2018. The international panel for climate change (IPCC) estimated that the total temperature increased by 0.72 °C in the period from 1951 to 2012. The expected temperature will likely increase from 1 °C to 3 °C till 2050 and from 2 °C to 5 °C till 2100, depending on the different gas emission circumstances documented by IPCC-2013. Water reserves are the core of most crises in countries like Pakistan, where the economy, culture, and textiles are intimately connected to irrigation water. The effects of warming drift on the outer circle are particularly unique and complex [1].

Throughout history, floods had recorded as one of the most devastating natural disasters capable of destroying severe personal damage as well as property. In recent years, the expansion of flood events has been a problem resulting in a large number of deaths

every year. Another main reason is that as the human population is increasing, human communities are becoming closer to water resources. Flood-affected people’s infrastructure and lives have been severely damaged and disturbed [2].

In this study, the data-driven approaches based on the statistical relationship between input and output data were used to forecast the flood events in the UIB. A possible alternative to the current approaches for the hydrological forecast of streamflow may be the data-driven approach like as ANNs [3].

2. Materials and Methods

2.1. Study Area

The area of research is come across within the following geographical points as 33–75° N and 72–78° E gambol by the swat hillock in the northwards the Mohamad, Mangla Complex, and the District of Charsada in the southwesterly, and the Kotli, Mangla, Kalam Complex in the westerly are shown in Figure 1. Kalam station is at a high mean sea level of 5821 km and Munda Dam is at low mean sea level of 376 km. The research region is characterized by climatic lateral tropical and wet temperate zones, with thunderstorms and snowfall. Summers are hot (41.9 °C) and winters are frigid (0.8 °C) in the chosen location. The effluvium ornament is a cardinal aspect in determining water potential. The flow of a locality is defined by its effluvium compactness, which is the relative amount of rainfall that enters. As a result, the lower the runoff, the higher the chances of recharging.

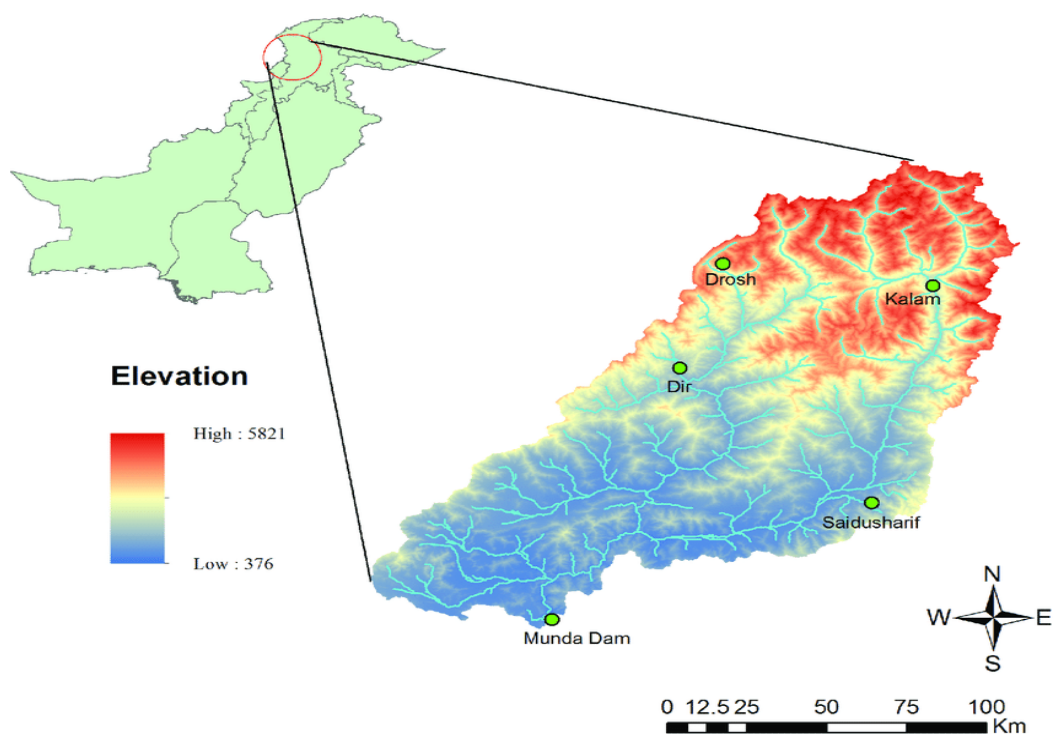


Figure 1. Digital elevation model of Indus Basin with Gauge Stations.

2.2. Data Collection and Model Description

Pakistan is one of the most climatically varied countries due to its wide temperature range, and more occurrence of rainfall in the UIB including extreme flood events. Pakistan has faced more financial and wastage of precious lives due to extreme flood events such as in the years 2011, 2020, and 2022. Data on extreme events of stream flow was collected for the duration from 1971–2009. The data was collected from the Water and Power Development Authority (WAPDA) and Pakistan Meteorological Department (PMD) [4].

2.2.1. LSTM Model

Deep learning is a type of neural network that uses a larger number of layers and layer types to model complex systems and interactions. Because traditional neural networks cannot retain temporal information, recurrent neural networks were developed using previous time step information. LSTMs are a deep learning version of recurrent neural networks that can remember information for longer. To change the data, LSTM cells use gates, vector addition, and multiplication to remove or add information.

2.2.2. SARIMA Model

SARIMA is a regression model, and all regression models assume that the values in a data set are independent of one another. When using regression to predict time series, it is critical to ensure that the data is stationary, which means that statistical properties such as variance do not change over time. In ARIMA, “AR” denotes the “autoregressive” component, which is the lag of the stationary series, “MA” denotes the lags of the forecast errors, and “I” denotes the order of differentiation to make the series stationary. SARIMA (1, 0, 1) × (0, 1, 1) 12 model was used in this study for forecasting of flow in UIB. To ensure that the time-series data were stationary, the Dickey-Fuller test was used. The resulting *p*-value was less than 0.03 and the test static was −3.739768, allowing us to reject the null hypothesis and conclude that the data is stationary, seasonality period(s) was 12 month moving average and minimum AIC score was 138.065 at 29-time steps.

2.2.3. Model Evaluation Criteria

Root Mean Square Error (RMSE) is a statistical method commonly used to compare predicted values with observed values in hydrology fields to evaluate the performance of forecasting models. Based on the relative range of the data, the RMSE is frequently used to evaluate how closely the predicted values match the observed values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y - Y_i)^2} \quad (1)$$

In Equation (1), *Y* and *Y_i* are actual and predicted discharges at time *t*, respectively; *Y* is the mean of actual discharges; and *n* is the total number of observations.

2.2.4. Model Structure

Our study is related to open-source software libraries. According to the research, Python [5] is the programming language of choice. NumPy [6], Pandas [7], and Matplotlib [8] libraries are also imported for data processing, management, and visualization. We created the LSTM model and TensorFlow [9], a Google open-source software library. TensorFlow was originally designed to conduct machine learning, deep learning, and numerical computation research using data flow graphs. However, this framework is comprehensive enough to be applicable to a wide range of domains.

3. Results and Discussion

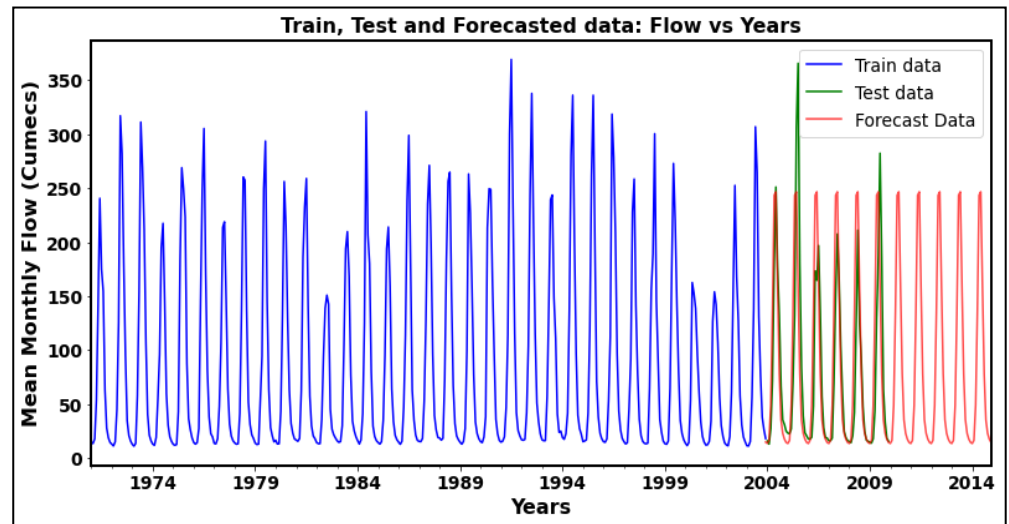
The forecasting models developed were validated and tested using independent data. When comparing actual data and forecast values made during the validation and testing process, the RMSE values were used to evaluate the qualitative and quantitative performance of the scenarios. The size of the training data sets varies depending on the scenario. A thirty-four-year dataset (1971–2004) was used for training while a five-year dataset (2005–2009) was used for testing and a five-year dataset (2010–2014) was used for forecasting.

The RMSE in the two-prediction model for the test and train is shown in Table 1. From the observation of below RMSE results the LSTM model performs better for training as well as testing as compared to the SARIMA model.

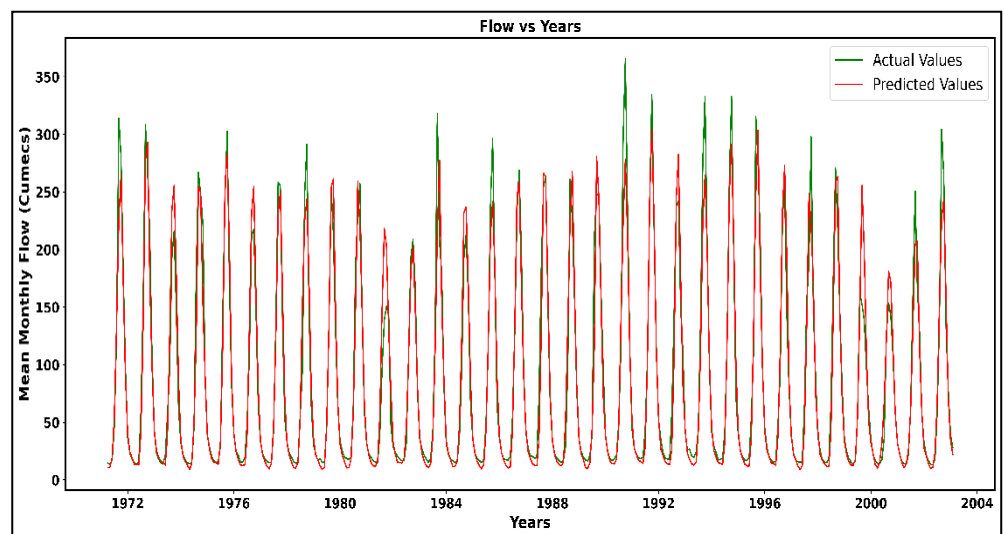
RMSE	LSTM	SARIMA
Training	22.79	27.82
Testing	35.05	54.42

Train, Test, and Forecast Results

Figure 2a–c show the visual comparison results of actual and forecasted flow data from Kalam Station. Figure 2a shows that the comparison result of the SARIMA model in Kalam Station. The blue line shows the training data and the green line shows the testing data and the red line shows the forecasted data, which is the output data of our model. Figure 2b,c show the comparison result of the LSTM model in Kalam Station. In Figure 2b the green line shows the training data (Actual Values) and the red line shows the testing data (Predicted Values) while in Figure 2c the green line are the observed values from the model and the red line is the forecasted values, which is output data of our model.



(a)



(b)

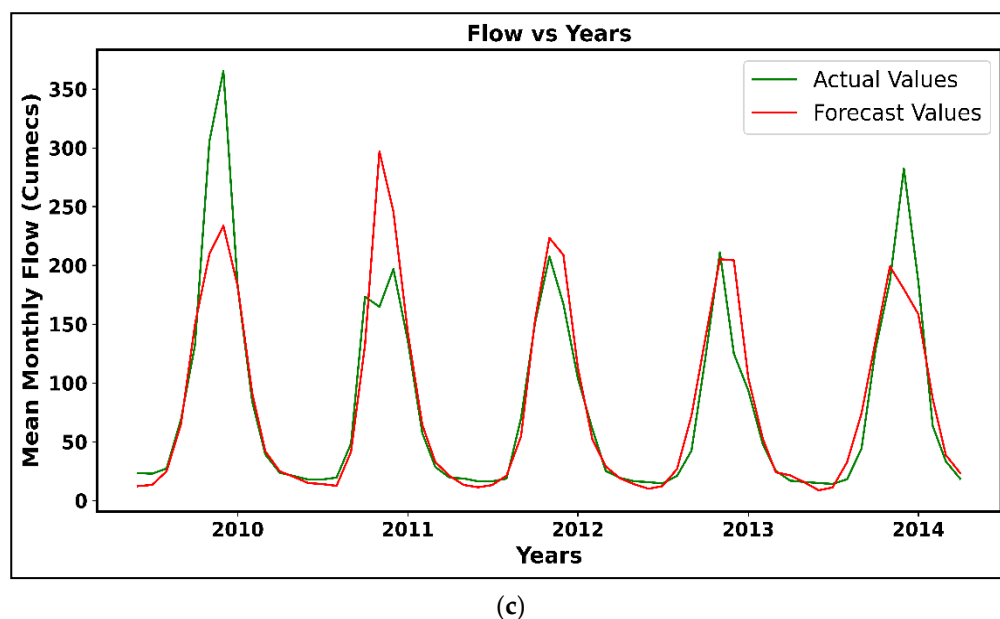


Figure 2. Visual Comparison of Actual and Forecasted Data (a) SARIMA (b) LSTM training and testing (c) LSTM Forecasting.

4. Conclusions

Accurately time series forecasting for upcoming flood events is important but it is a challenging task like in Pakistan. The flood forecasting at the Swat River flow gauging station, particularly for downstream stations that lack discharge information, such as Saidu Sharif in this study, plays an important role in the early flood-warning system. In this research, we examined the classical statistical model used such as the Deep learning model, LSTM, and the Regression model, SARIMA. In our comparison of the models for Kalam station, the LSTM model achieved better results and more accurate forecasting performance than the SARIMA model. SARIMA at modeling linear data, whereas the other statistical deep learning models at modeling nonlinear data. The RMSE values of LSTM model fit to the series found for training and testing, for one-year-ahead forecasting are 22.79, and 35.05, respectively. These results indicates that the Deep Learning algorithm is a dependable ideal solution for flood prediction due to its high precision.

References

- Shahid, F. Climate Change: Impacts on Pakistan and Proposed Solutions. *Pak. Soc. Sci. Rev.* **2021**, *5*, 223–235. [https://doi.org/10.35484/pssr.2021\(5-ii\)18](https://doi.org/10.35484/pssr.2021(5-ii)18).
- Gude, V.; Corns, S.; Long, S. Flood Prediction and Uncertainty Estimation Using Deep Learning. *Water* **2020**, *12*, 884. <https://doi.org/10.3390/w12030884>.
- Atashi, V.; Gorji, H.T.; Shahabi, S.M.; Kardan, R.; Lim, Y.H. Water Level Forecasting Using Deep Learning Time-Series Analysis: A Case Study of Red River of the North. *Water* **2022**, *14*, 1971. <https://doi.org/10.3390/w14121971>.
- Jawad, M.; Nadeem, M.S.A.; Shim, S.-O.; Khan, I.R.; Shaheen, A.; Habib, N.; Hussain, L.; Aziz, W. Machine Learning Based Cost Effective Electricity Load Forecasting Model Using Correlated Meteorological Parameters. *IEEE Access* **2020**, *8*, 146847–146864. <https://doi.org/10.1109/access.2020.3014086>.
- 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. <https://doi.org/10.3390/w11071387>.
- 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. <https://doi.org/10.1109/mcse.2011.37>.
- McKinney, W. Data structures for statistical computing in python. *Proc. 9th Python Sci. Conf.* **2010**, *1*, 56–61. <https://doi.org/10.25080/majora-92bf1922-00a>.
- Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- 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.

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.