Deep Learning-Based LSTM Model for Forecasting Water Quality in the Godavari River

Sai Bargav Reddy M[1], Ayush Kushwah[1] , Vinnarasi R [1]

[1,] Department of Civil Engineering, [1]Indian Institute of Technology Roorkee, Uttarakhand, India

Corresponding Author: muskula_sbr@ce.iitr.ac.in

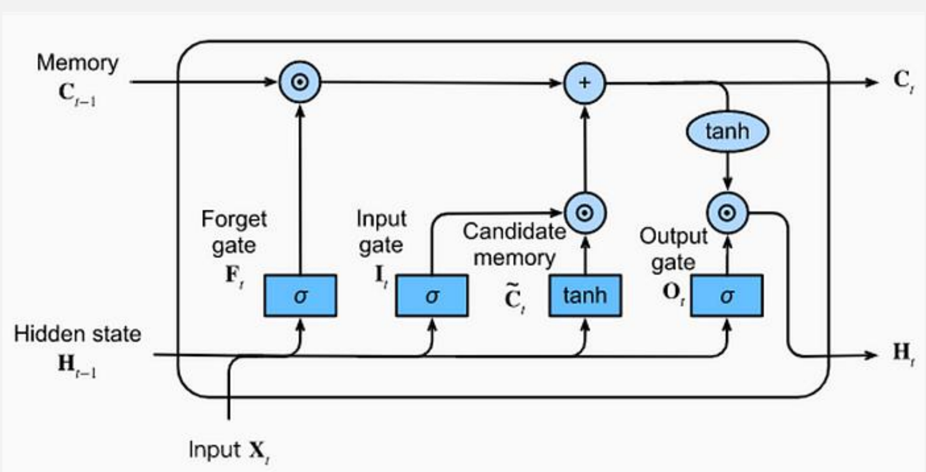
AGU23
San Francisco, CA & Online Everywhere
11-15 December 2023

1. Introduction

- Accurate forecasting of water quality is of utmost importance in preserving water resources, safeguarding public health, and protecting the environment.
- While traditional physical-based models have been utilized for water quality assessment over the years, recent advancements in data-driven approaches have shown superior performance in forecasting water quality parameters.
- Among these data-driven models, deep learning methods such as Long Short-Term Memory (LSTM) hold significant advantages in handling sequential data, capturing long-term dependencies, and thus, proving to be a powerful choice for tasks involving time series data compared to traditional Artificial Neural Network (ANN) models.
- In this context, we present a study focusing on the application of an advanced data-driven model for forecasting Dissolved Oxygen (DO), a vital river water quality parameter, that involves the development of an LSTM-based model, considering river water temperature as the forcing variable for the Godavari river.
- The developed model was trained, validated, and tested for the 1990-2006, 2006-2014, and 2014-2017 periods respectively.
- Three evaluation criteria, namely root mean square error (RMSE), coefficient of determination (R), and Nash-Sutcliffe efficiency (NSE), were used to assess the model's accuracy in forecasting DO.

2. Methodology

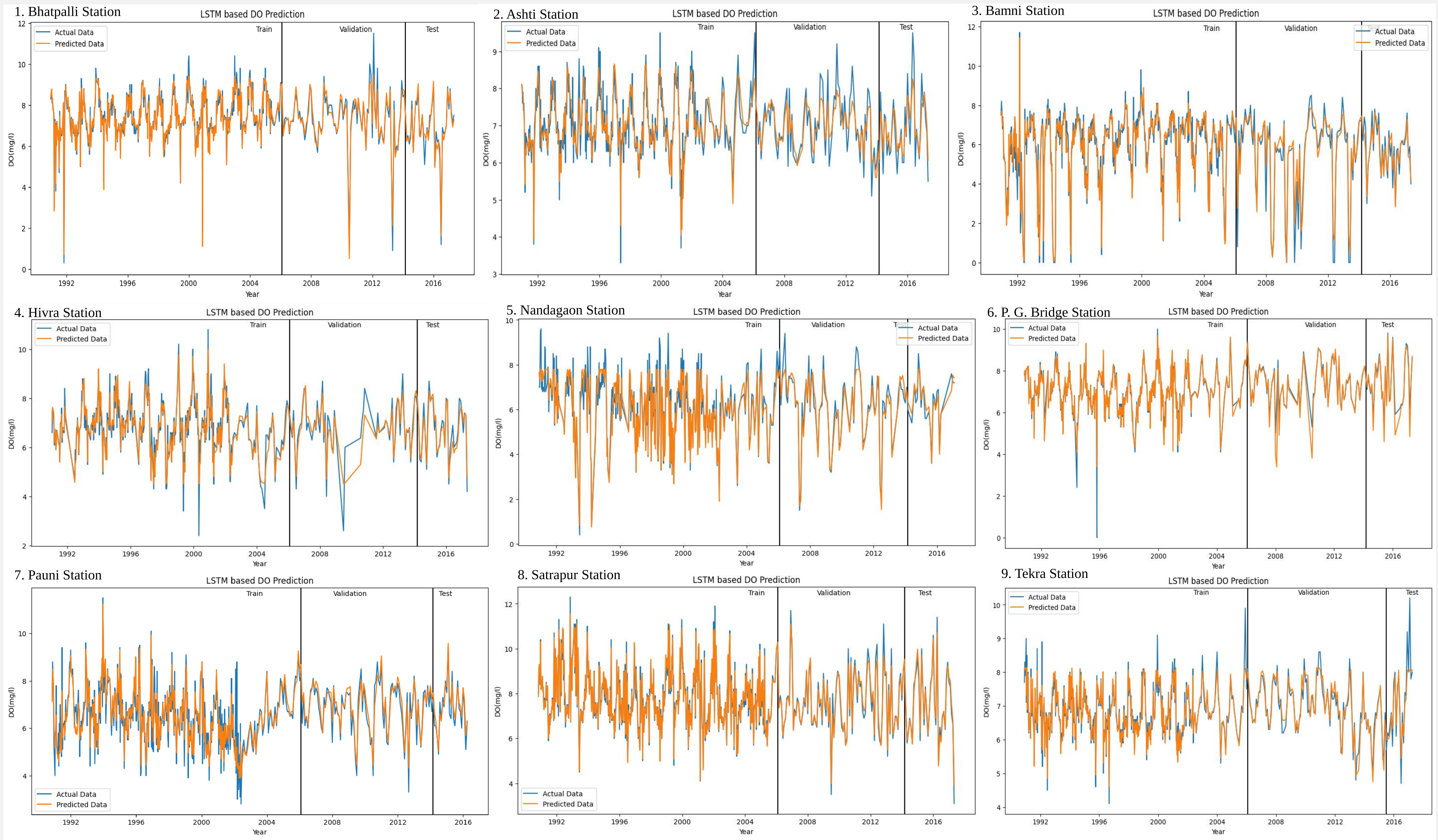
- Long Short-Term Memory Model (LSTM) is an artificial RNN proposed by Sepp Hochreiter et al (1997)
- LSTM contains feedback connections unlike the typical feed forward networks
- An LSTM consists of a cell, an input gate, forget gate and an output gate.
- The cell remembers time series and the gates regulate the flow of data in and out.



3. Datasets

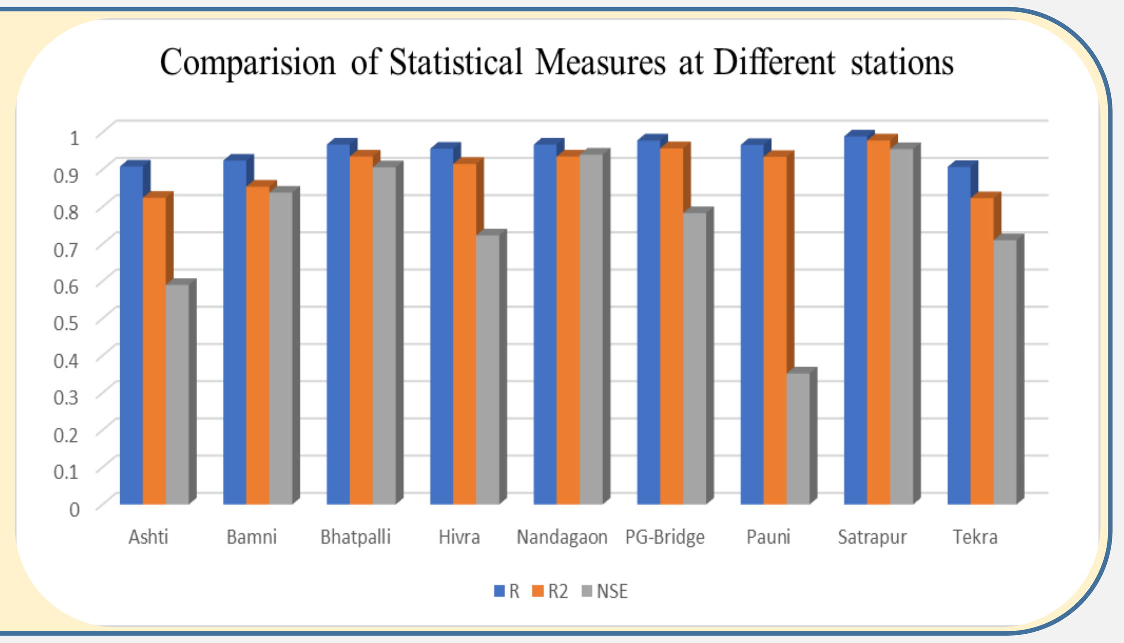
- In this study, we use the water quality data of Dissolved oxygen from 1990-2017.
- The data is obtained from Global Freshwater Quality Database (GEM Stat) data for all the stations of river Godavari.
- The DO dataset of 9 stations is selected since these had a long-term and consistent set of observed data.
- The dataset is divided into 3 sets i.e. for training, validation, and testing, and spanned over a time period of 1990 – 2005, 2006-2013, and 2014-2017.

4. Results



5. Highlights

- To evaluate the performance of the developed model statistical measures like the coefficient of correlation (R), coefficient of determination (R^2), and Nash Sutcliffe Efficiency (NSE) were calculated at every station.



- We used LSTM to develop a new method to predict water quality parameter DO at 9 stations of river Godavari.
- From the study, it is found that by the optimum setting of parameters, the LSTM model exhibited a better model performance.
- It is observed that most of the stations have a reasonable correlation (R^2 >0.7) and NSE (NSE > 0.5) values with observed data, except for the stations Ashti and Pauni since these had inconsistent datasets.
- The results are reasonable at most of the stations. The model performance can be improved by further tuning the parameters.