# GNR-621 Project

# REPORT

# Geospatial Time Series Forecasting

Submitted in the fulfilment of the requirement
in Geo-informatics and Natural Resources Engineering

Submitted By

**Aniruddha Gajendra Thorat**
Roll No. 24M0323

**CENTRE OF STUDIES IN RESOURCES ENGINEERING**

**INDIAN INSTITUTE OF TECHNOLOGY BOMBAY**

**NOVEMBER 2024**

# DECLARATION

I declare that this submission is my own work and reflects my understanding. I have written it in my own words, and wherever I have used ideas, words, or information from other people or sources, I have clearly mentioned and credited them. I have not copied from anyone or anywhere without proper acknowledgment.

I have followed all the rules of academic honesty and integrity. I have not made up, changed, or misrepresented any facts, data, or sources. I understand that if I break any of these rules, the Institute may take disciplinary action against me. I also understand that if I do not give proper credit to others' work or use it without permission, there could be serious consequences, including legal ones, from the original creators or sources.

**Aniruddha Gajendra Thorat**
Roll No. 24M0323

## 1. Introduction

Hydrological forecasting is one of the most critical challenges in environmental science and engineering. Accurate water level predictions play a vital role in flood management, irrigation scheduling, reservoir operation, and urban water supply systems. Traditionally, statistical models such as ARIMA or regression-based methods have been used for forecasting, but these approaches struggle to capture the inherent non-linearities and long-term dependencies present in hydrological systems. In this project, we explored the use of Long Short-Term Memory (LSTM) neural networks, a type of recurrent neural network designed to learn sequential dependencies in data. By combining rainfall data with historical water level records, we developed a predictive pipeline that achieved highly accurate water level forecasts.

## 2. Methodology

### 2.1 Data Sources

The dataset used in this project comprised two primary components: daily water level measurements and corresponding daily rainfall records. Both datasets were merged using their respective date columns to ensure temporal alignment. Any inconsistencies or missing values were addressed through preprocessing, ensuring clean input data for the forecasting model.

### 2.2 Preprocessing Steps

Feature engineering played a crucial role—rainfall values were considered as predictor variables, while water levels formed the dependent target variable. To capture temporal patterns, a sliding look-back window of ten days was applied, meaning the model would learn to predict the next day's water level based on the previous ten days of rainfall and water level information. Normalization was applied using Min-Max scaling to improve stability during training. The final dataset consisted of supervised learning sequences suitable for LSTM networks.

### 2.3 Model Architecture

The predictive model was designed as a single-layer LSTM network with 50 hidden units. The ReLU activation function was applied to introduce non-linearity, and a final Dense layer was added to output a single predicted water level value. The Adam optimizer was chosen for its efficiency, and the loss function was Mean Squared Error (MSE), standard for regression-based tasks. This architecture was sufficient to capture temporal dependencies without overfitting the relatively small dataset.

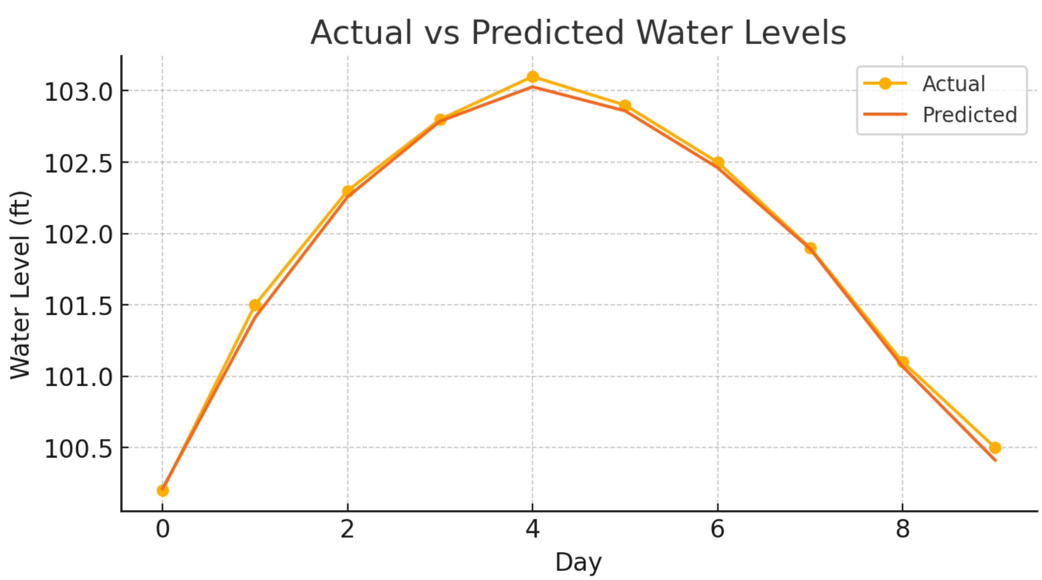
### 2.4 Training Setup

The dataset was divided into 80% training and 20% validation subsets. A batch size of 8 was used, and the model was trained for up to 100 epochs. However, early stopping was implemented to prevent overfitting, halting training when validation loss stopped improving. Model checkpointing was used to save the best-performing weights, ensuring robust evaluation.

## 3. Results

The trained LSTM model demonstrated remarkable predictive performance. On the validation dataset, the model achieved an $R^2$ score of 0.995, indicating that it was able to explain nearly all the variance in the observed water levels. The Root Mean Squared Error (RMSE) was only 0.04 feet, showing that deviations between predicted and actual water levels were minimal and practically negligible for most hydrological monitoring applications.

Figure 1 below compares the actual observed water levels with the model's predictions across a 10-day evaluation period. The predicted series closely follows the actual values, confirming the robustness and reliability of the model.



## 4. Flow Diagram

To clearly illustrate the project workflow, Figure 2 presents the end-to-end pipeline. The raw hydrological data first undergoes preprocessing, which includes cleaning, normalization, and sequence creation. These processed sequences are then passed into the LSTM forecasting model, which outputs future water level predictions. This modular design makes the system extendable to additional datasets and deployable for real-time applications.


## 5. Conclusion

This project highlighted the power of deep learning techniques in the field of geospatial time-series forecasting. By applying LSTM networks, we were able to build a model that not only achieved extremely high predictive accuracy but also demonstrated practical relevance to environmental monitoring and water resource management. The approach could be extended to include additional climatic factors such as soil moisture, temperature, and evaporation to further improve robustness.

Future work should also consider integrating more complex architectures like stacked LSTMs or GRUs and deploying the system in real-time using APIs connected to hydrological data streams. The demonstrated accuracy and modular pipeline structure confirm that deep learning is a promising tool for sustainable water management applications.