Long-term Precipitation Prediction using Machine Learning

Jonnalagadda Naveen Reddy

20ME31021

Under the Supervision of:

Prof. Rajib Maity

Civil Engineering Department

Mechanical Engineering Department

Indian Institute of Technology Kharagpur

Contents

Introduction

- The Indian summer monsoon (ISM) stands as a pivotal climatic phenomenon crucial for the well-being of Indian society, economy, and ecosystems, this contributes a substantial 75% to the total annual rainfall.
- The fluctuations in Indian Summer Monsoon Rainfall, whether heightened or diminished, wield a profound impact on agricultural productivity and the overall Indian economy.
- ISM plays a vital role in sustaining ecological balance and biodiversity across diverse ecosystems in India. The timing and intensity of the monsoon directly influence the health of rivers, lakes, and other water bodies, impacting the flora and fauna that depend on these aquatic habitats.

Motivation

- The ability to enhance forecasts of Indian Summer Monsoon Rainfall (ISMR) at extended lead times is of paramount importance.
- Improved predictive capabilities would empower effective water management strategies, enable more proactive responses to flood and drought scenarios, and ultimately aid in averting humanitarian crises.
- In essence, a more nuanced understanding and accurate forecasting of ISMR stand as crucial tools for sustainable planning and resilience against the unpredictable vagaries of the Indian summer monsoon.

Objectives

- Long-Term Precipitation Prediction
- To develop machine learning model capable of accurately predicting long-term precipitation patterns, extending beyond traditional weather forecasting timelines.

- Classification into Drought, Floods, and Normal Conditions
- To extend the prediction models to classify the predicted total precipitation values into categories such as drought, floods, and normal conditions.

Literature Review

- The long-term precipitation prediction intricately relies on several key climate indices that serve as crucial determinants of atmospheric conditions, they are:

  Dipole Mode Index (DMI) = $\boldsymbol{SST}_{WesternIndianOcean} - \boldsymbol{SST}_{EasternIndianOcean}$

  El Niño-Southern Oscillation (NINO)

  North Atlantic Oscillation (NAO) $= \dfrac{(\boldsymbol{P}_{IcelandicLow} - \boldsymbol{P}_{AzoresHigh})}{Normalization\ factor}$

  Pacific Decadal Oscillation (PDO)

  Southern Oscillation Index (SOI) $= \dfrac{(\boldsymbol{P}_{Tahiti} - \boldsymbol{P}_{Darwin})}{Normalization\ factor}$

- These indices collectively play pivotal roles in shaping the climatic dynamics influencing precipitation patterns.

Study Area and Data

Study Area

- The study focuses on the Hyderabad region, encompassing the geographical coordinates of approximately 18.25° N latitude and 78.75° E longitude.

Data Collection

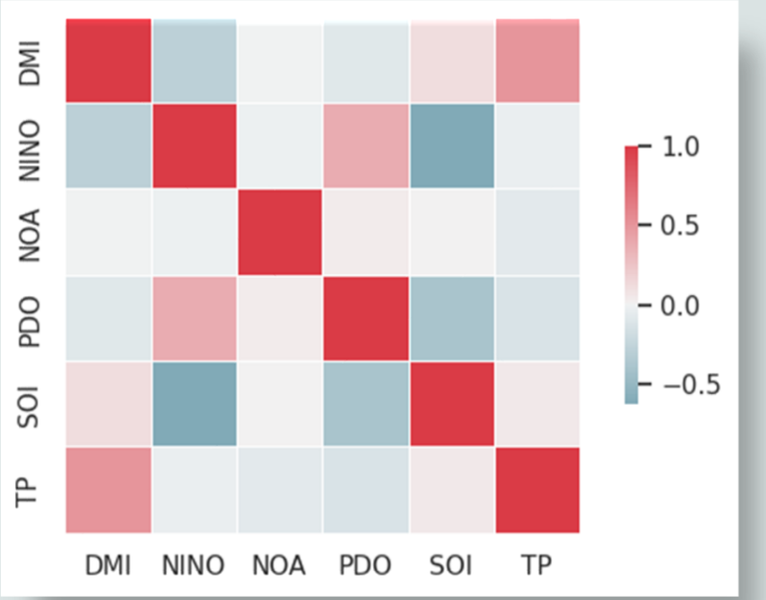
- The primary dataset utilized for this study is sourced from the ERA-5 (European Centre for Medium-Range Weather Forecasts Reanalysis 5) database.
- The data collection spans from January 1, 1951 to September 1, 2023 ensuring a comprehensive representation of climatic conditions over the study period.

Sample Data

Date	DMI (°C)	NINO (°C)	NOA	PDO	SOI	TP (mm)
1951-01-01	2.170837	25.24	0.08	-1.19	1.5	55.2739
1951-02-01	2.882813	25.71	0.7	-1.52	0.9	7.121963
1951-03-01	1.829071	26.9	-1.02	-1.72	-0.1	172.6387
1951-04-01	1.841797	27.58	-0.22	-1.35	-0.3	76.84635
1951-05-01	-0.14844	27.92	-0.59	-1.29	-0.7	91.09028
1951-06-01	2.009735	27.73	-1.64	-1.77	0.2	64.77641
1951-07-01	2.682587	27.6	1.37	-0.23	-1.0	155.9257
1951-08-01	3.439484	27.02	-0.22	-1.76	-0.2	100.0813
1951-09-01	1.586914	27.23	-1.36	-0.78	-1.1	137.442
1951-10-01	0.96875	27.2	1.87	-0.09	-1	104.7145
1951-11-01	1.15521	27.25	-0.39	-0.31	-0.8	86.6932
1951-12-01	1.85446	26.91	1.32	-1.45	-0.7	0.62956

Methodology

- Correlation Matrix



Feature	Correlation coefficient with total Precipitation
DMI	0.502877
PDO	0.123481
NOA	0.077731
SOI	0.049904
NINO	0.033829

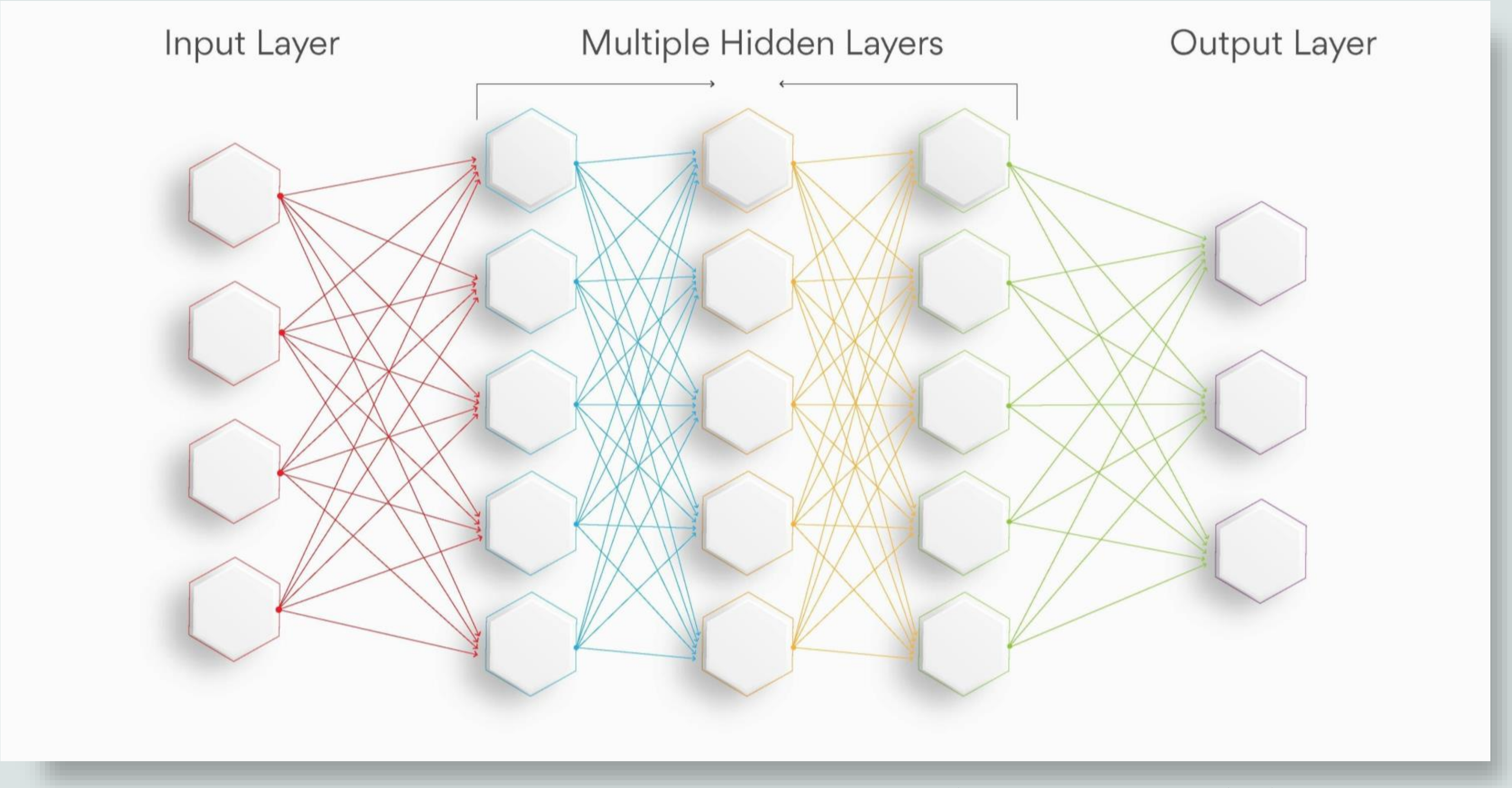
$$\text{Pearson Correlation factor } (r) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Where x_i, y_i are values of i^{th} datapoint of feature X, Y

And $\bar{x}$, $\bar{y}$ are mean values for features X, Y.

Methodology

❖ Multivariate Long Short-Term Memory (LSTM) Architecture



**Source: https://images.app.goo.gl/2yk8SBVuBKrjxzEG6*

Methodology

- Parameters that affect the error and model performance are:

  Batch Size - refers to the number of training examples utilized in one iteration.

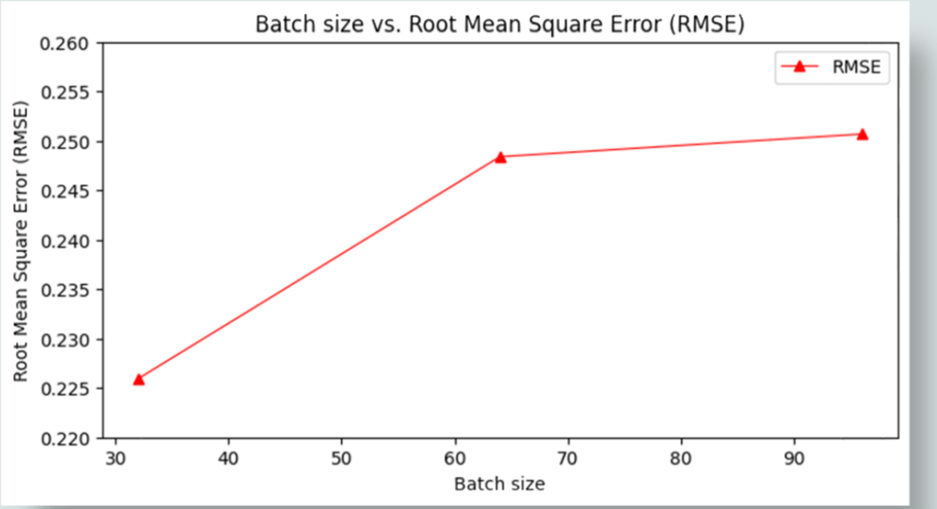
  Window Length - the number of time steps considered as input to the model.

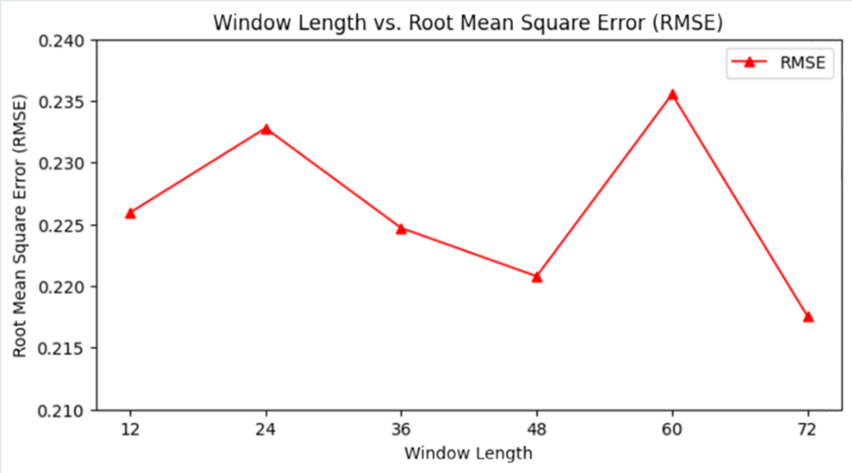
  Number of Neurons - determines the complexity of the model.

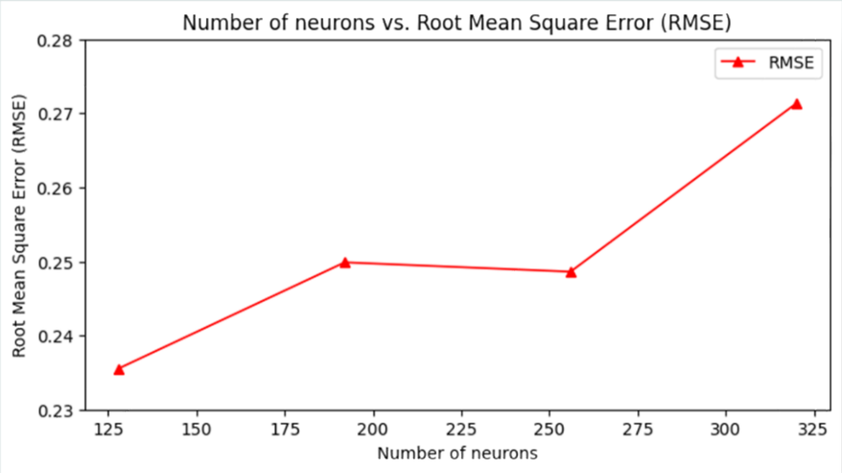
  Dropout Rate - regularisation technique to reduce overfitting.

  Epoch - one complete pass through the entire training dataset.

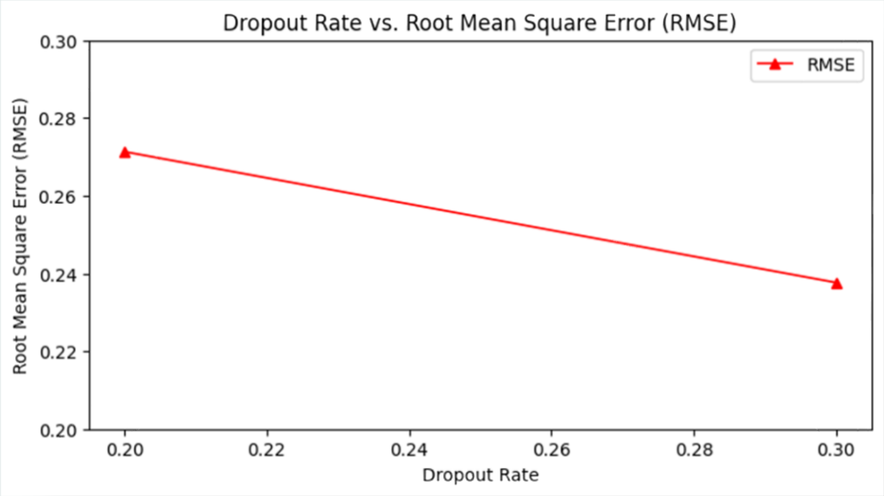
Parameter Tuning

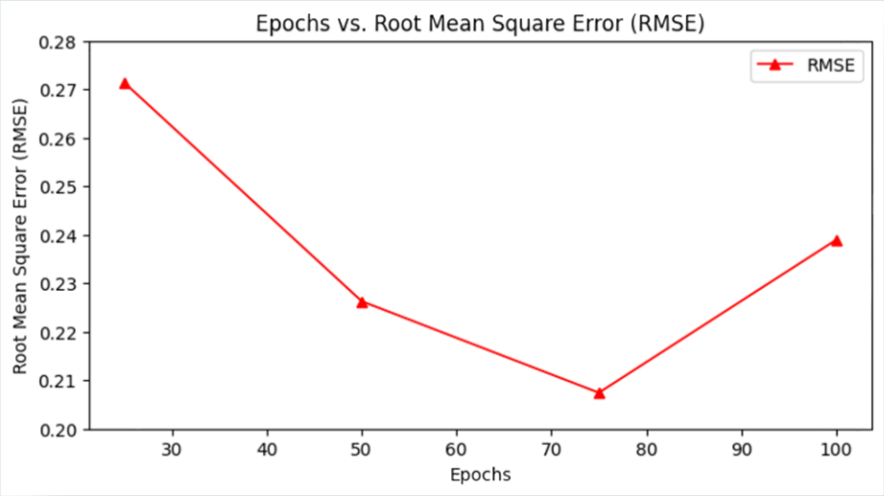






Parameter Tuning





Results and Discussion

- Parameters for minimum loss and high performance of the model are:

  Batch Size - 32

  Window Length - 72

  Number of Neurons - 128

  Dropout Rate - 0.3

  Epochs - 75

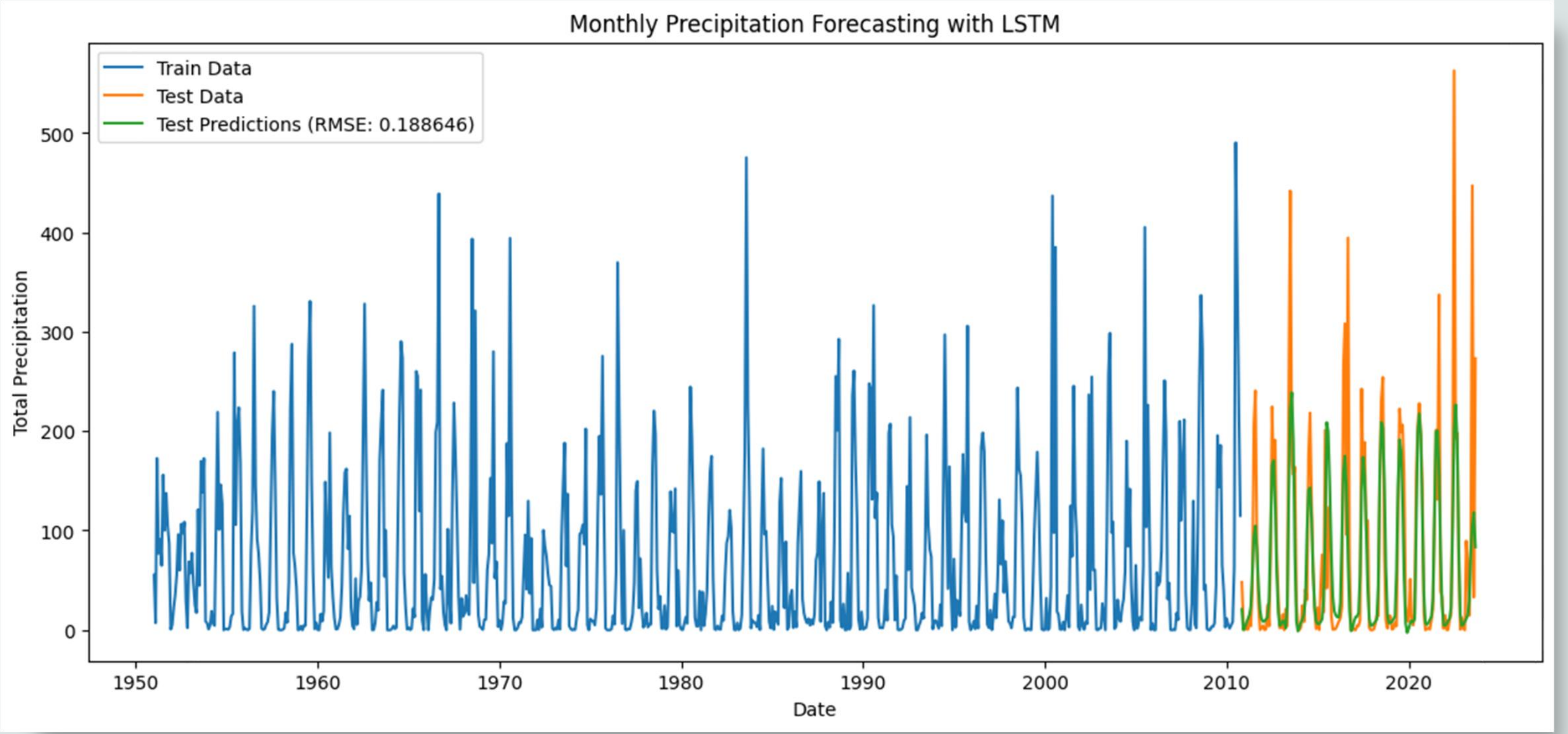
- Root Mean Square Error (RMSE) value for the model using above parameters is approximately 0.19 (~40 for unscaled values).

$$\text{RMSE} = \sqrt{\frac{\Sigma_{i=1}^{N}(A_i - P_i)^2}{N}}$$

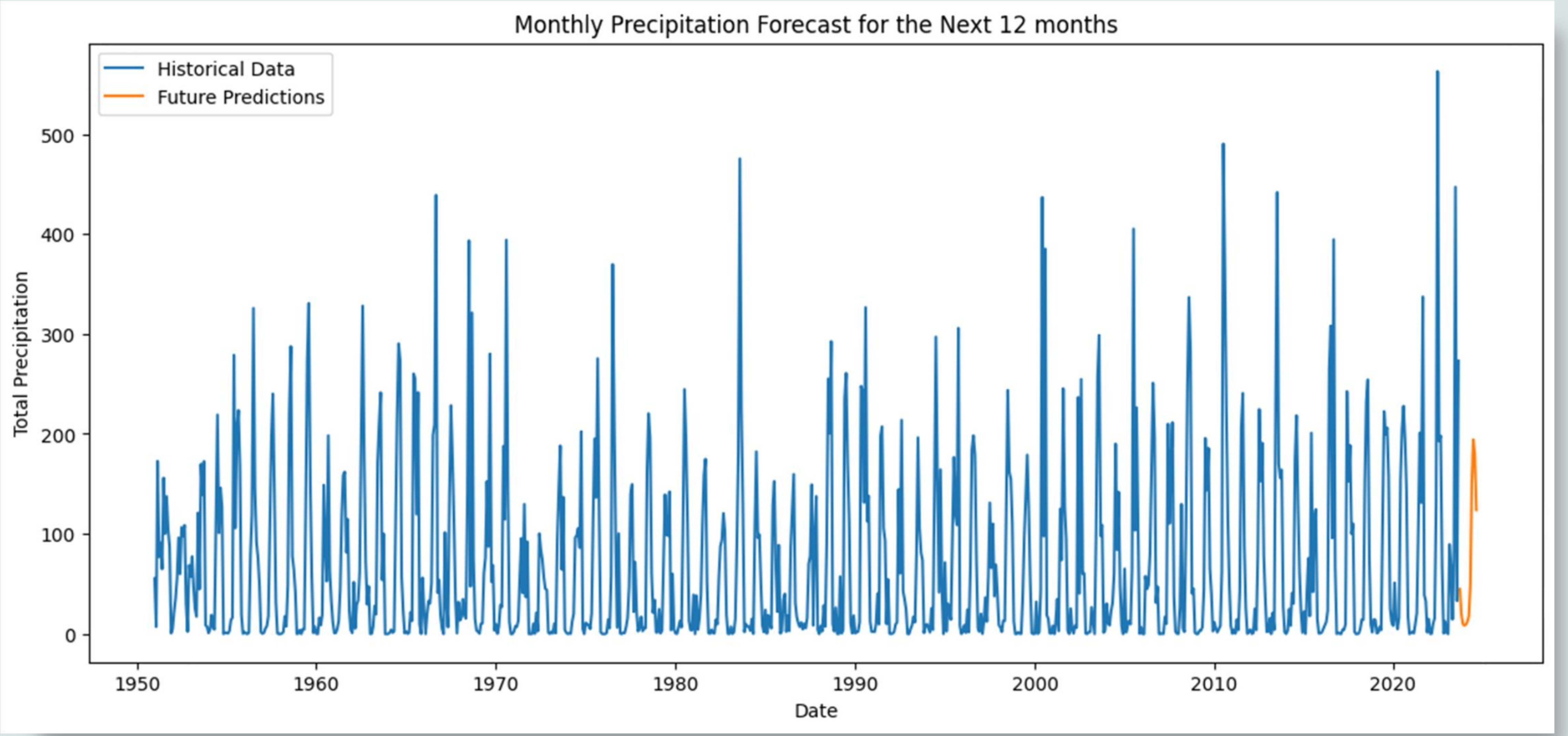
A_i : Actual value of i^{th} testing point

P_i : Predicted value of i^{th} testing point

Results and Discussion



Results and Discussion



Results and Discussion

Date	Total precipitation (mm)
2023-10-01	20.4081
2023-11-01	19.0373
2023-12-01	9.91922
2024-01-01	8.42815
2024-02-01	9.39634
2024-03-01	12.2064
2024-04-01	17.6274
2024-05-01	48.0225
2024-06-01	151.3613
2024-07-01	194.136
2024-08-01	179.086
2024-09-01	123.848

- Actual total precipitation for Oct 2023 is 17.0323 mm.
- For 2024 Monsoon: Predicted total precipitation is 648.431 mm
- Z-score:

$$\text{Z-score} = \frac{X - \mu}{\sigma}$$

where μ, σ are mean and standard deviation of the dataset

- The Z-score for 2024 Monsoon data is

(652.451 – 648.431)/201.31 = 0.02

- Normal conditions (lies in the range [-1, 1])

Conclusions

- This study introduces a Multivariate Long Short-Term Memory (LSTM) framework designed for forecasting long-term monthly precipitation.
- Leveraging time series precipitation data and relevant influencing indices, the model successfully predicts "Total Precipitation" for the upcoming 12 months, achieving a Root Mean Square Error (RMSE) value of 0.19.
- This result underscores the effectiveness of the proposed framework in accurately capturing and predicting complex precipitation patterns, offering valuable insights into long-term climatic trends.

Future Work

- Explainable Artificial Intelligence
- To explain the output predictions using Explainable boosting algorithm.
- To create an user interface where user can visualize predictions as well as explanations.

References

- Spatial variation in long-lead predictability of summer monsoon rainfall using a time-varying model and global climatic indices Riya Dutta, Rajib Maity.
- A multivariate EMD-LSTM model aided with Time Dependent Intrinsic Cross-Correlation for monthly rainfall prediction Kavya Johny, Maya L. Pai, Adarsh S.
- Long-Lead Statistical Forecasts of the Indian Summer Monsoon Rainfall Based on Causal Precursors G. Di Capua, M. Kretschmer, J. Runge, A. Alessandri, R. V. Donner, B. Van Den Hurk, R. Vellore, R. Krishnan and D. Coumou.
- Long-term precipitation prediction in different climate divisions of California using remotely sensed data and machine learning S. Majnooni, M. Reza Nikoo, B. Nematollahi, M. Fooladi, N. Alamdari, G. Al-Rawas and Amir H. Gandomi.

THANK YOU !!