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 ISMR, 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 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
- Develop machine learning models capable of accurately predicting long-term precipitation patterns, extending beyond traditional weather forecasting timelines.

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

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)

    El Niño-Southern Oscillation (NINO)

    North Atlantic Oscillation (NAO)

    Pacific Decadal Oscillation (PDO)

    Southern Oscillation Index (SOI)

- 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

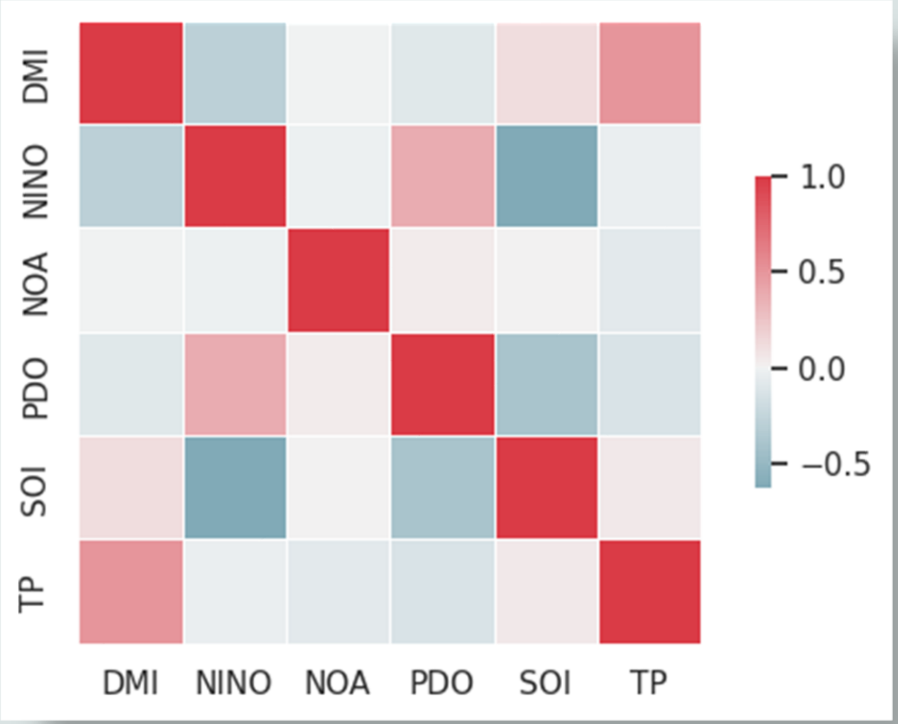
- The primary dataset utilized for this study is sourced from the ERA-5 (ECMWF Reanalysis 5) database.
- The data collection spans from January 1, 1951, to October 1, 2023, ensuring a comprehensive representation of climatic conditions over the study period.

Few lines of Data

Date	DMI	NINO	NOA	PDO	SOI	TP
1951-01-01	2.170837	25.24	0.08	-1.19	1.5	0.001842
1951-02-01	2.882813	25.71	0.7	-1.52	0.9	0.000237
1951-03-01	1.829071	26.9	-1.02	-1.72	-0.1	0.005755
1951-04-01	1.841797	27.58	-0.22	-1.35	-0.3	0.002562
1951-05-01	-0.14844	27.92	-0.59	-1.29	-0.7	0.003036
1951-06-01	2.009735	27.73	-1.64	-1.77	0.2	0.002159
1951-07-01	2.682587	27.6	1.37	-0.23	-1	0.005198
1951-08-01	3.439484	27.02	-0.22	-1.76	-0.2	0.003336
1951-09-01	1.586914	27.23	-1.36	-0.78	-1.1	0.004581

Methodology

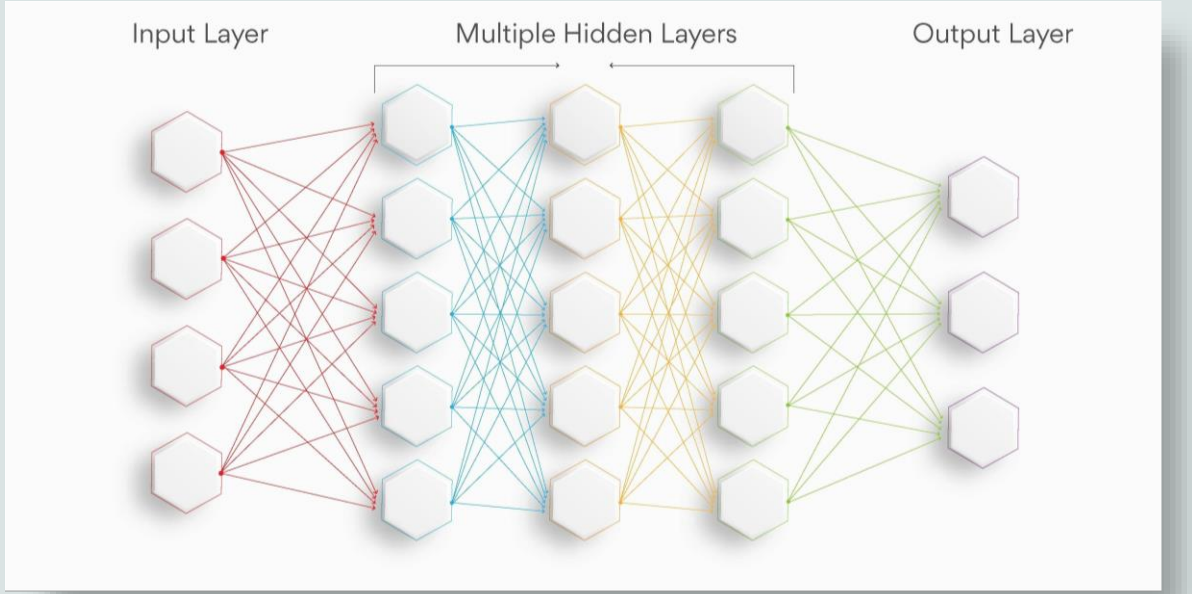
- Confusion Matrix



- TP - 1.000000
- DMI - 0.502877
- PDO - 0.123481
- NOA - 0.077731
- SOI - 0.049904
- NINO - 0.033829

Methodology

- Multivariate Long Short-Term Memory (LSTM) Architecture
- LSTM is a type of recurrent neural network (RNN) designed to capture long-term dependencies and patterns in sequential data. Unlike traditional RNNs, LSTMs have memory cells that can store and retrieve information over extended sequences, making them particularly effective for time series prediction tasks.



Methodology

- Parameters that effect the error and model performance are :

    Batch Size

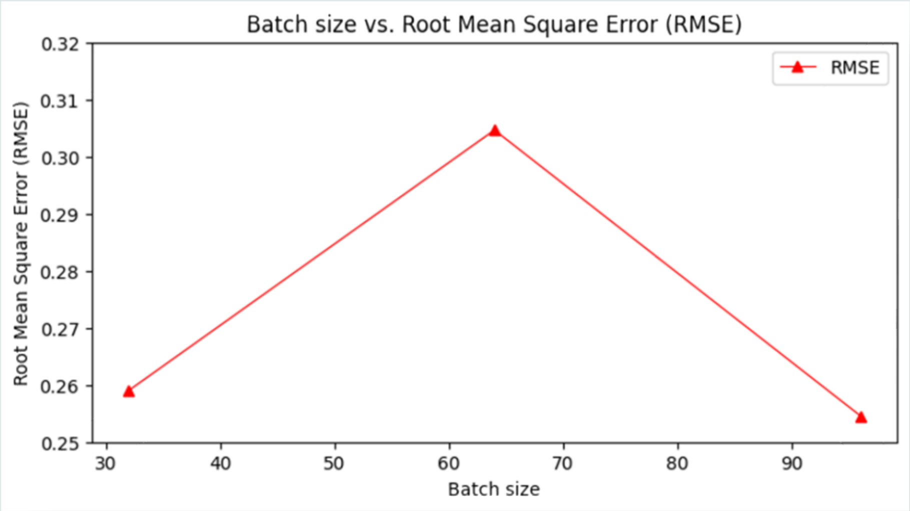
    Window Length

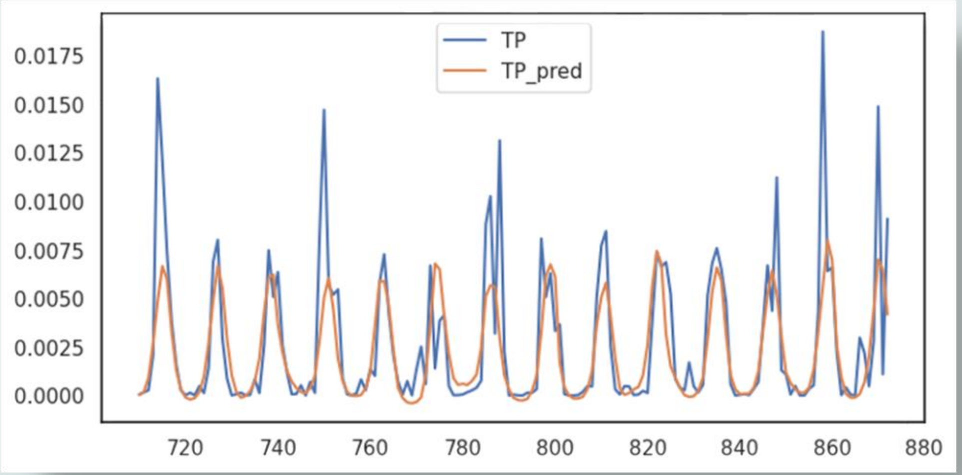
    Number of Neurons

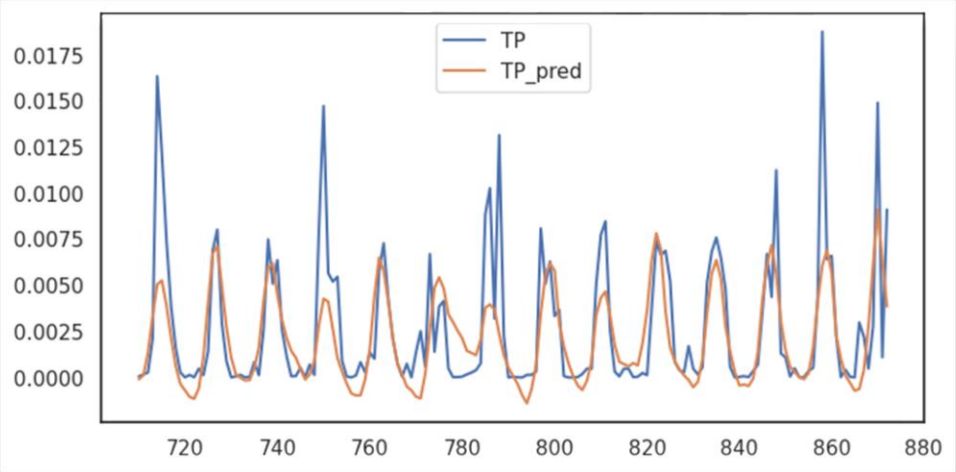
    Dropout Rate

    Epochs

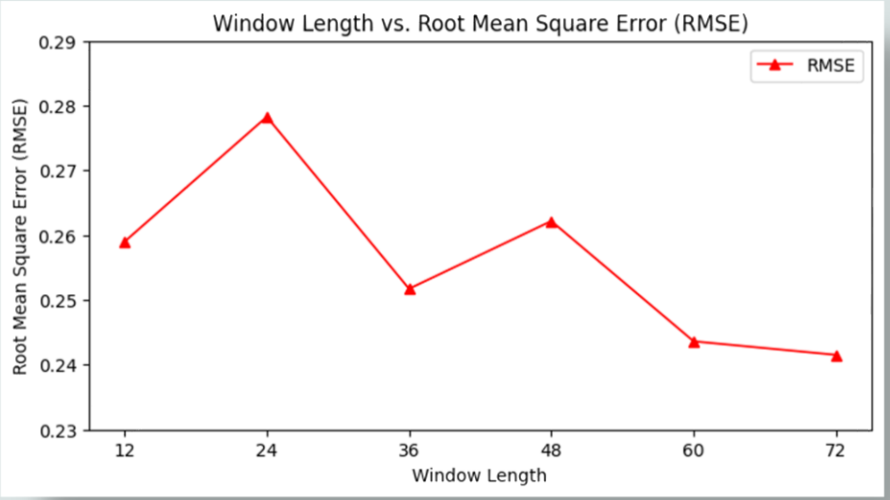
Parameter Tuning

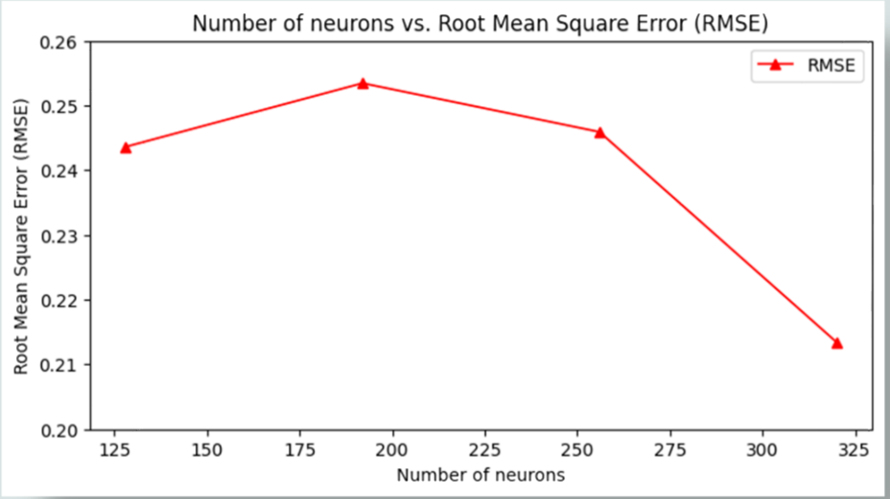


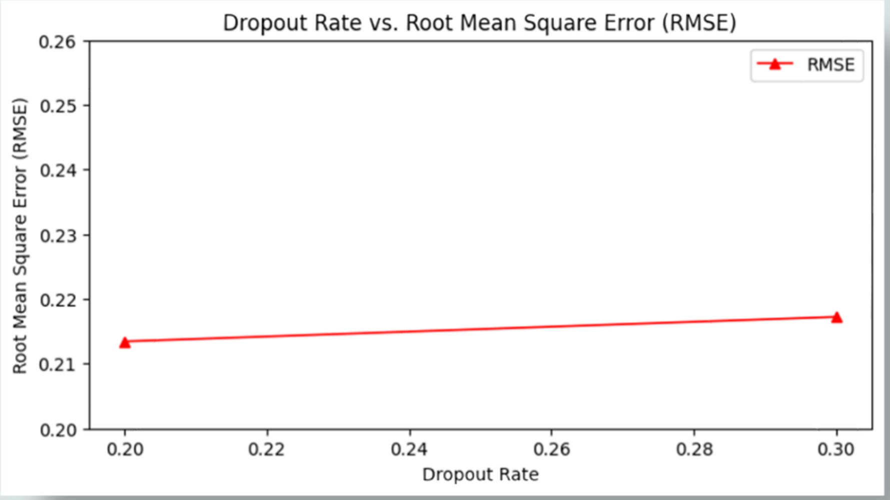


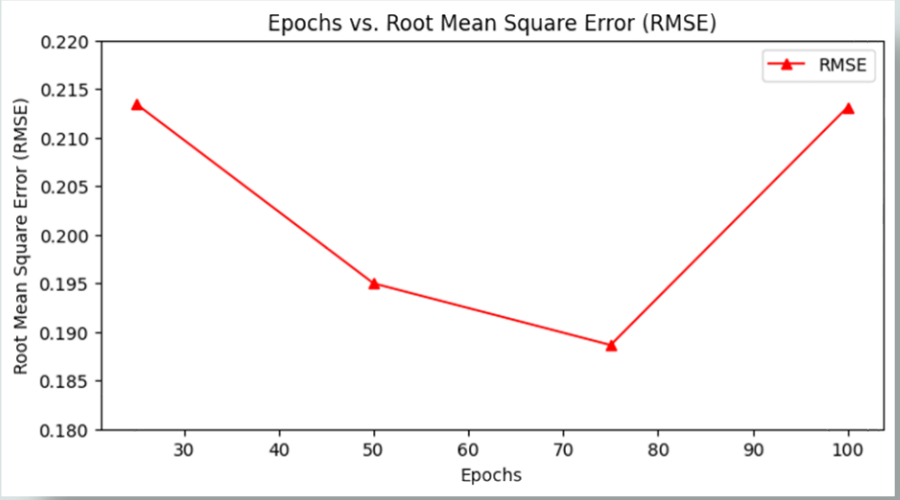


Parameter Tuning









Results and Discussions

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

  Batch Size - 32

  Window Length - 72

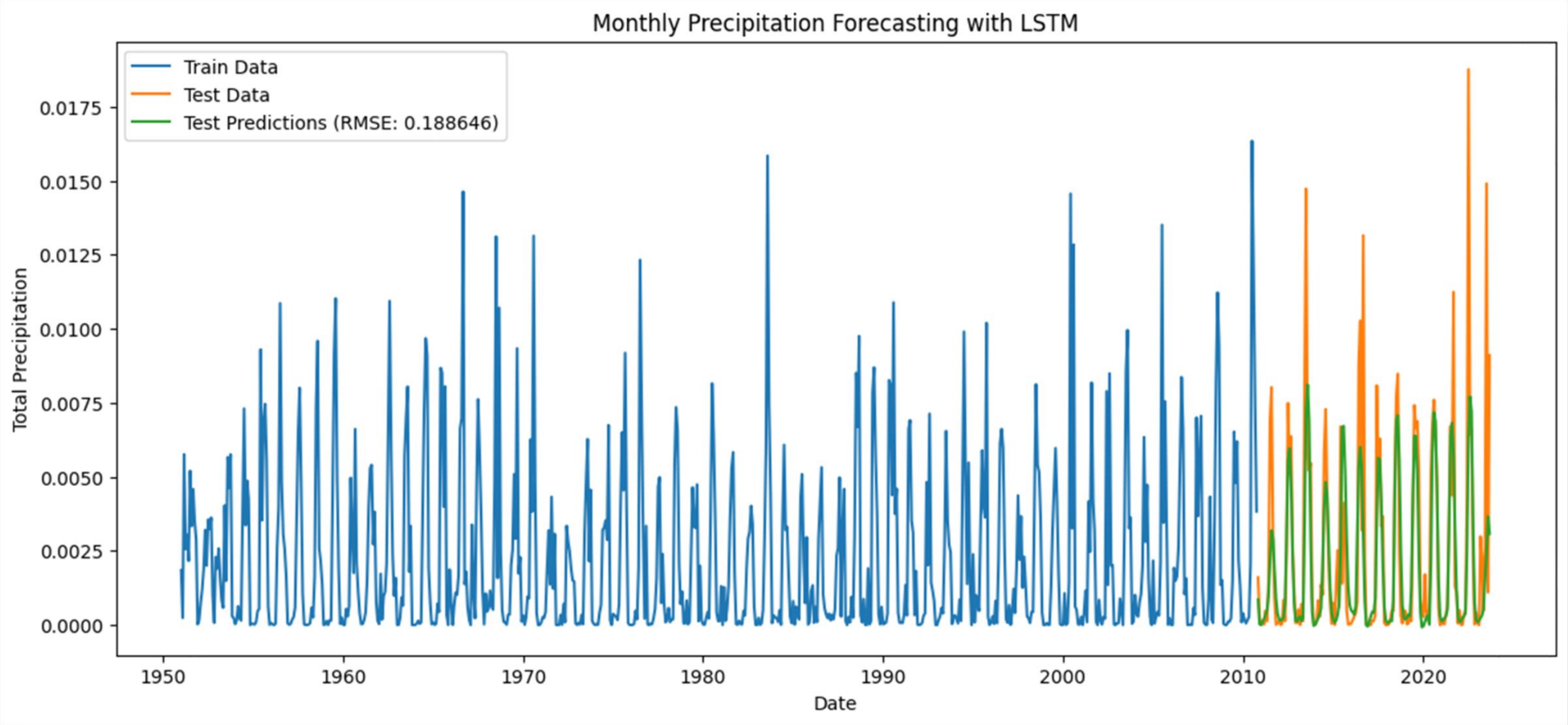
  Number of Neurons - 320

  Dropout Rate - 0.2

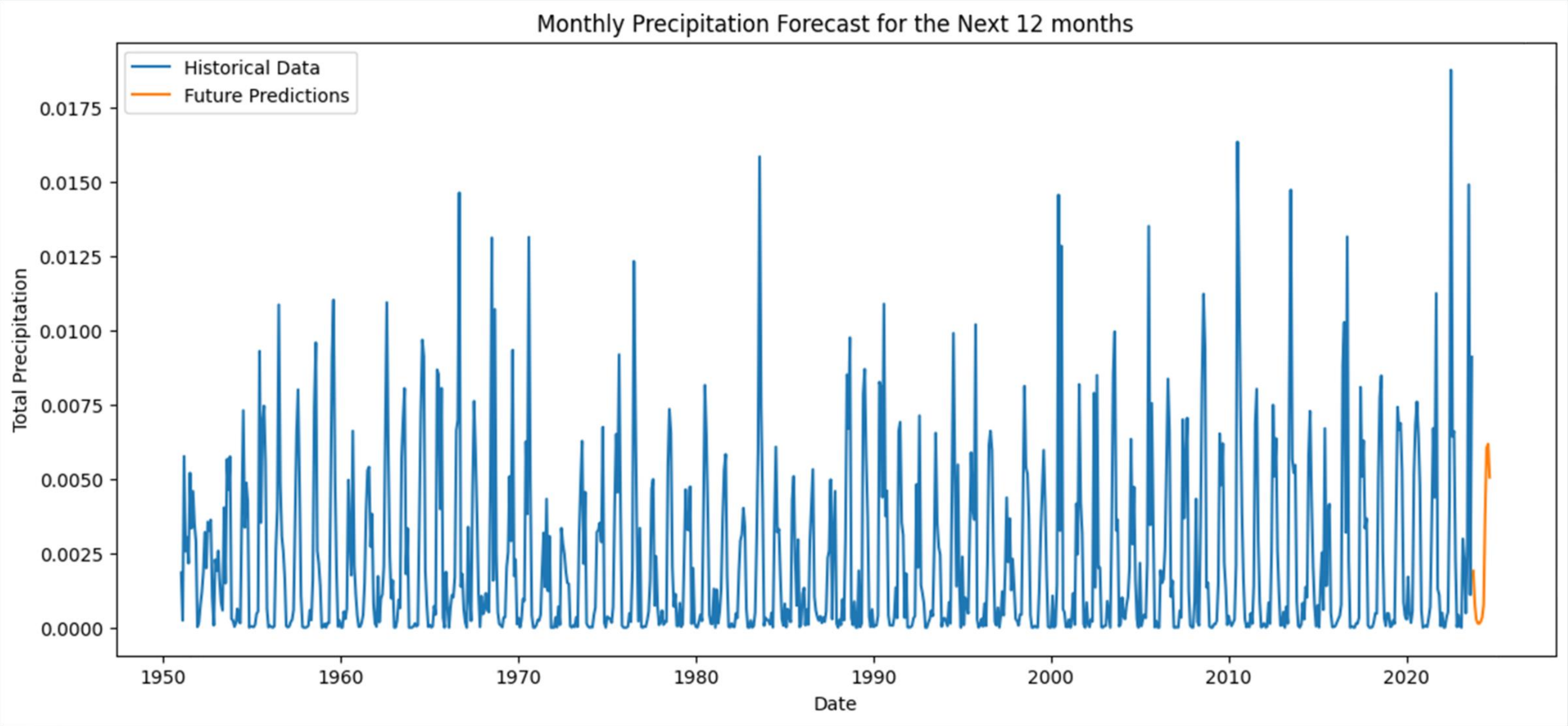
  Epochs – 75

- Root Mean Square Error (RMSE) value for the model using above parameters is approximately 0.19

Results and Discussions



Results and Discussions



Results and Discussions

Date	Total Precipitation
2023-10-01	0.001898
2023-11-01	0.000799
2023-12-01	0.000282
2024-01-01	0.000139
2024-02-01	0.000135
2024-03-01	0.000197
2024-04-01	0.000343
2024-05-01	0.000773
2024-06-01	0.003876
2024-07-01	0.006038
2024-08-01	0.006172
2024-09-01	0.005050

Conclusions

Future Work

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 !!