An Internship Report on

"Rainfall Pattern Prediction Using CHIRPS High-Resolution Data: An AI/ML and Data Science Project"

Submitted in partial fulfilment of requirement for the award of the degree

Bachelor of Engineering

Of

Visvesvaraya Technological University

By

Name of Student: J SRUJAN VISHWAKARMA USN: 1CR21AD023

Carried out under

Student Programme for Advancement in Research Knowledge (SPARK) at

Council of Scientific & Industrial Research

FOURTH PARADIGM INSTITUTE (CSIR-4PI)

NAL BELUR CAMPUS, BANGALORE - 560 037

Under the guidance of

Dr. Gyanendranath Mohapatra

Principal Scientist,

CSIR-Fourth Paradigm Institute,

NAL Belur Campus

CMR INSTITUTE OF TECHNOLOGY

Department of Computer Science

NAAC A++ Accredited

AECS Layout, Bangalore-56003

DECLARATIONS

I am J Srujan Vishwakarma, a 4^{th} -year student of Artificial Intelligence & Data Science Engineering, CMR Institute of Technology- 560 037, hereby confirm the successful completion of my four-month internship at CSIR-Fourth Paradigm Institute (CSRIR 4PI). This report serves as a fulfilment of the 15 weeks internship, conducted from September 11, 2024, to December 28, 2024.

NAME: J Srujan Vishwakarma

Date: 28-12-2024

Place: Bangalore

USN: 1CR21AD023

ACKNOWLEDGEMENT

This Internship is a result of accumulated guidance, direction, and support of several important persons. I take this opportunity to express my gratitude to all who have helped me to complete the Internship.

I sincerely acknowledge the patience, encouragement, timely help, and guidance of **G N Mohapatra**, Principal Scientist, and Associate Professor AcSIR, CSIR-Fourth Paradigm Institute, to complete my internship within the stipulated time successfully

I express my sincere thanks to **Dr. Sanjay Jain**, Principal, CMRIT, for providing me with adequate facilities to undertake this Internship. I would like to thank **Dr. Shanthi M B**, Head of Dept AI&DS, for providing me with an opportunity to carry out the Internship and for his valuable guidance and support.

I extend my gratitude to **Dr. Sridevi Jade**, Head and Outstanding Scientist at the CSIR-Fourth Paradigm Institute (CSIR-4PI), and **Dr. Anil Kumar**, Chief Scientist at CSIR-4PI, Bangalore, for providing me with the incredible opportunity to work at this esteemed institute.

I would like to thank all the other teaching and non-teaching staff of the Artificial Intelligence & Data Science Department, who had directly or indirectly, helped me in the completion of the project.

Last but not least, I would like to thank my parents and friends who provided us with valuable suggestions to improve our projects.

J SRUJAN VISHWAKARMA

1CR21AD023

Table of Contents

Abstract

This study provides a comprehensive exploration of rainfall patterns across India, shedding light on their critical implications for agriculture, water resource management, and disaster preparedness. By leveraging high-resolution CHIRPS data, the analysis offers an in-depth understanding of the dynamics of the Indian monsoon system, including the Indian Summer Monsoon Rainfall (ISMR) and the contrasting characteristics of the Southwest and Northeast monsoons. Regions across India are systematically classified based on distinct rainfall patterns, offering valuable insights into regional variations.

Advanced methodologies such as regression analysis, correlation, and trend analysis are applied to rainfall data, enabling accurate predictions and better understanding of long-term climatic shifts. Statistical analyses focus on monthly, yearly, and seasonal precipitation trends, providing comparisons between CHIRPS and IMD datasets to ensure data accuracy and reliability.

The study also presents a decade-wise analysis of rainfall trends across diverse regions, highlighting key shifts in precipitation patterns over time. By integrating these findings, the research underscores the role of rainfall trends in shaping India's socio-economic landscape, offering actionable insights for policymakers and stakeholders in agriculture, climate resilience, and sustainable development.

CHAPTER 1

Introduction

1.1 Indian Rainfall

Rainfall is the result of the gravitational descent of atmospheric moisture in the form of water, occurring when a portion of the atmosphere becomes saturated with water vapor, leading to condensation and precipitation. Three main types of rainfall are convectional, orographic, and cyclonic/frontal. Convectional rain is prevalent in equatorial regions due to daily ground heating, while orographic rain results from air ascent over highlands, causing condensation on the windward side. Cyclonic or frontal rain is associated with the upward movement of air due to wind convergence, often linked to cyclones or frontal systems. These distinct types of rainfall contribute to varied precipitation patterns globally, exemplified by phenomena such as the monsoon rains in India and the orographic rain experienced in the Western Ghats.

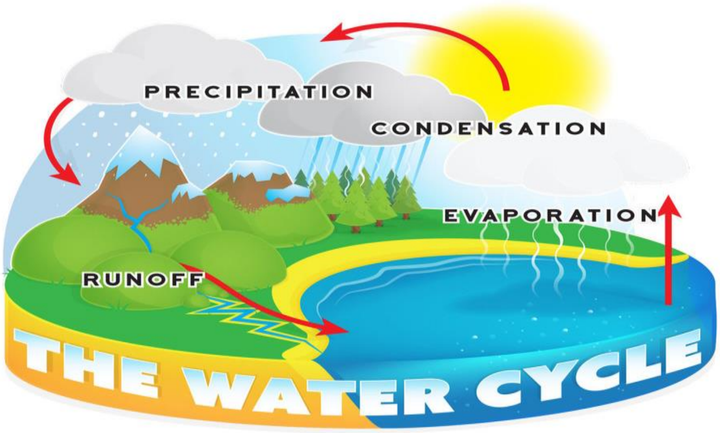

Fig. Water Cycle of Earth system

In the context of India, rainfall holds critical implications for agriculture, which heavily relies on rainwater management and surface or groundwater withdrawal. Monsoon rains significantly influence Indian agriculture, impacting green water availability (stored in soil) and playing a vital role in rainfed agriculture, particularly in arid and semi-arid regions. Challenges and opportunities arise, with climate change affecting rainfall patterns and requiring adaptive strategies. Conservation agriculture, climate-smart agro-technologies, and watershed-based approaches become essential for sustainable water management, emphasizing the importance of balancing blue and green water resources to ensure resilient farming and rural livelihoods.

1.2 Monsoon period

The Indian monsoon, a cornerstone of the region's climate system, plays a pivotal role in shaping the socio-economic and ecological framework of the Indian subcontinent. This dynamic phenomenon, marked by seasonal wind reversals, ushers in life-sustaining rainfall that profoundly influences agriculture, water resources, trade, and even cultural traditions. During the warmer months, the monsoon winds shift from the northeast to the southwest, delivering substantial rainfall primarily from June to September, which is critical for replenishing water reserves, supporting agricultural cycles, and driving economic productivity. Beyond its economic significance, the monsoon also holds a deep cultural resonance, influencing festivals, traditions, and rituals across the region.

India's monsoon is categorized into five meteorological divisions, each experiencing unique rainfall patterns and impacts. While the northeast monsoon shapes rainfall in Tamil Nadu and Kerala, regions such as Coastal Andhra Pradesh, Rayalaseema, and South Interior Karnataka rely heavily on monsoonal precipitation. These divisions reflect the diversity of the monsoon's effects, from agricultural dependencies to water resource management challenges. Understanding these distinct patterns is essential for predicting monsoon behavior, mitigating risks in agriculture and water management, and ensuring climate resilience. The monsoon's intricate interplay with regional livelihoods underscores its unparalleled importance, not just as a weather phenomenon but as a driver of life, culture, and sustainability in India.

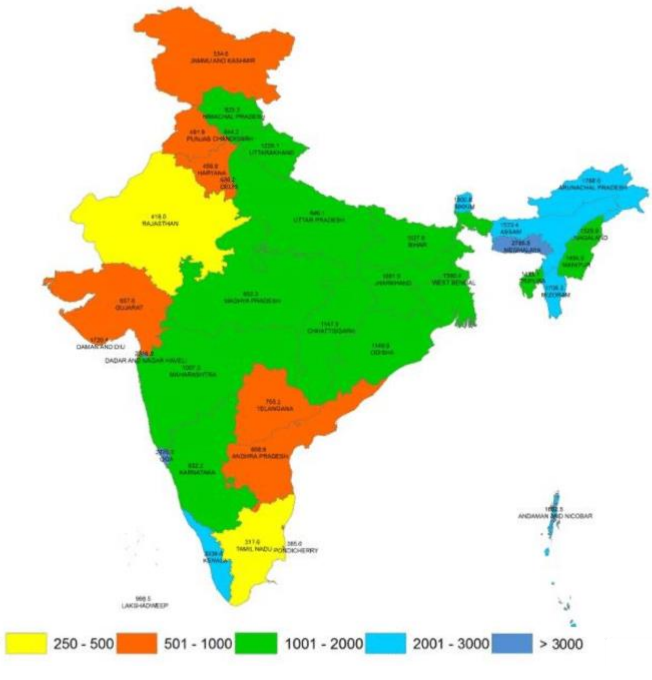

Fig. Average rainfall in India during the monsoon period

1.3 Importance of rainfall data

The Indian summer monsoon rainfall (ISMR) holds paramount importance for India, profoundly influencing various aspects of the nation's well-being and development. With approximately 80% of the annual precipitation occurring during the June to September period, the stability, variability, and extremes of ISMR play a pivotal role in shaping India's agricultural output, economic performance, and overall societal well-being. Deviations from the long-term mean can have profound effects on agricultural productivity and the Gross Domestic Product (GDP), emphasizing the critical role of the summer monsoon in sustaining livelihoods and economic stability.

Beyond its impact on agriculture and the economy, the ISMR is a key determinant of water availability for irrigation, drinking, and other purposes, directly influencing the livelihoods of millions in rural areas. The accurate forecasting of ISMR is indispensable for informed policy decisions across sectors such as agriculture, energy, water resources, health, and disaster management. Additionally, understanding the variability of ISMR is crucial for climate resilience planning, allowing the development of adaptive strategies based on historical data and trends. In essence, the monitoring and analysis of Indian summer monsoon rainfall are fundamental for fostering sustainable development, disaster preparedness, and ensuring the well-being of the diverse population across the Indian subcontinent.

1.4 South-west vs North-east monsoon seasons

The Southwest Monsoon (SWM) and the Northeast Monsoon (NEM) are integral components of India's climate system, each with distinct characteristics and implications. The Southwest or Summer Monsoon, occurring from June to September, brings moist winds from the southwest, resulting in abundant rainfall across the western coast, central India, and the northern plains. This season is critical for agriculture, supporting the cultivation of kharif crops and significantly impacting the nation's economy. SWM rainfall also plays a vital role in reservoir filling and holds cultural significance with associated festivals.

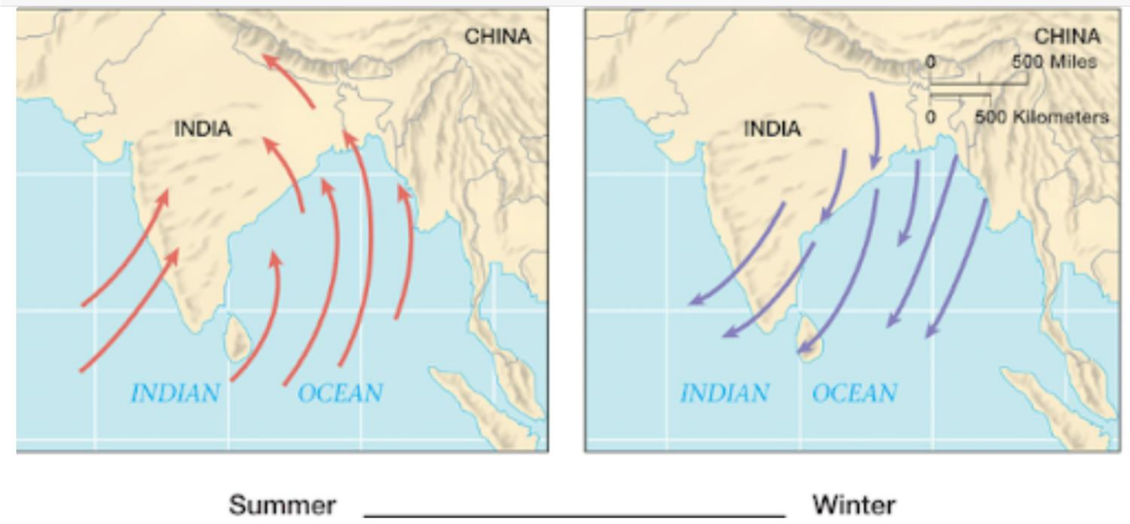

Fig. SWM (left) and NEM (right)

On the other hand, the Northeast or Winter Monsoon (NEM) takes place from October to December, with winds blowing from the northeast to the southwest. This monsoon predominantly affects the southern states of Tamil Nadu, Andhra Pradesh, Kerala, and parts of Telangana and Karnataka. Tamil Nadu, in particular, receives 48% of its annual rainfall from the NEM, making it crucial for agriculture and reservoir management in the state. The NEM is associated with cyclones in the North Indian Ocean and is often referred to as the winter or retreating monsoon due to its wind direction reversal after the SWM. In summary, these distinct monsoons contribute significantly to India's climate, agricultural practices, and water resources, with each playing a unique and crucial role in different regions.

1.5 Classification of India for Trend Analysis

India is geographically divided into five distinct regions, each characterized by unique rainfall patterns: North, South, East, West, and Central. In North India, encompassing states such as Jammu and Kashmir, Himachal Pradesh, Punjab, and Haryana, the influence of both the Southwest and Northeast monsoons results in diverse rainfall patterns. Additionally, winter rainfall may occur due to western disturbances. South India, including Kerala, Karnataka, Tamil Nadu, and Andhra Pradesh, is predominantly influenced by the Southwest Monsoon, with significant rainfall during the monsoon season. The Northeast Monsoon also plays a vital role, particularly in Tamil Nadu and parts of Andhra Pradesh, contributing to post-monsoon rainfall.

Moving to East India, which comprises states like West Bengal, Odisha, Bihar, and Jharkhand, abundant rainfall is observed during the Southwest Monsoon, with the Bay of Bengal influencing monsoon winds. Some areas also experience post-monsoon rainfall from the Northeast Monsoon. In West India, including Maharashtra, Gujarat, and Rajasthan, rainfall is primarily received from the Southwest Monsoon, with variations in distribution influenced by the Western Ghats along the west coast. Central India, encompassing states like Madhya Pradesh and Chhattisgarh, serves as a transition zone, experiencing a mix of rainfall patterns influenced by both the Southwest and Northeast monsoons. This region plays a crucial role in bridging the arid northwest and the more humid eastern parts of the country. The regional classifications underscore the diversity in India's climate, with each area contributing to the overall complexity of the country's rainfall dynamics.

CHAPTER 2

2. Objectives

2.1 Comparison of High-Resolution CHIRPS and IMD-Observed Data:

CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) is a global satellite rainfall estimation system that combines infrared and microwave satellite observations with ground-based station data. It provides high-resolution precipitation estimates, making it useful for water resource management, flood mitigation, and drought warning.

On the other hand, the IMD (Indian Meteorological Department) gridded rainfall data relies on ground-based rain gauges across India. While widely used, it may have limitations due to sparse station coverage and interpolation methods. In a study evaluating streamflow forecasting in the Nethravathi Basin, Karnataka, India, CHIRPS outperformed IMD data, emphasizing the importance of context and purpose in choosing the appropriate dataset.

2.2 Trend Analysis of Rainfall Pattern in India's Summer Monsoon Season

In India's summer monsoon season, there has been a notable trend in rainfall patterns. Over the years, variations in monsoon precipitation have impacted agriculture, water availability, and overall ecosystem health. Researchers have observed shifts in monsoon onset dates, intensity, and spatial distribution. These trends are crucial for sustainable water management, crop planning, and disaster preparedness. Understanding the changing monsoon dynamics is essential for adapting to climate variability and ensuring food security in the region.

2.3 Pre-monsoon, Monsoon, and Post-Monsoon Analysis and Comparison of CHIRPS and IMD Data with Missing Learning Technique

India's climate and agriculture are significantly influenced by the pre-monsoon**,** monsoon, and post-monsoon seasons. The pre-monsoon season, from March to May, is characterized by rising temperatures and dry conditions. During this period, data from CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) and IMD (Indian Meteorological Department) can assist in assessing soil moisture, water availability, and early agricultural planning. The monsoon season, which spans from June to September, is dominated by the southwest monsoon, contributing over 75% of India's annual rainfall. CHIRPS provides high-resolution satellite estimates, while IMD relies on ground-based rain gauges. Researchers study monsoon onset, intensity, and spatial distribution to understand its impact on agriculture and the economy.

The post-monsoon season extends from October to December, bringing relief after the monsoon with cooler temperatures and reduced rainfall. CHIRPS and IMD data help monitor groundwater recharge, river flow, and crop harvesting during this phase. While both datasets have strengths, addressing missing data is crucial. Techniques like deep learning models can handle missing values effectively, improving accuracy in rainfall predictions and water resource management. In summary, understanding these seasonal variations and leveraging data from CHIRPS and IMD aids in sustainable planning, disaster preparedness, and ensuring food security in India.

CHAPTER 3

3. Literature Survey:

3.1 Spatio-temporal rainfall variability over different meteorological subdivisions in India: analysis using different machine learning techniques, Gyanendranath Mohapatra, Rakesh1, Smrati Purwar, and A. P. Dimri, CSIR-Fourth Paradigm Institute, Windtunel Road, Bangalore, Karnataka, India.

Brief: The analysis of rainfall trends across different regions of India reveals notable patterns over the past century. Group 6, encompassing Kerala and the northeast part of the country, exhibited a discernible decline in annual and southwest monsoon season rainfall, while the pre-monsoon season indicated a negative trend specifically in this group. Conversely, post-monsoon rainfall displayed a negative trend in many regions, except for coastal Karnataka and Konkan. The findings suggest a concerning weakening of Northeast Monsoon (NEM) rainfall activity, particularly in peninsular and northeast India. Of particular significance is the consistent decrease in rainfall observed in Group 1, covering much of east-central India, predominantly dependent on rainfed agriculture. Decadal variability in rainfall did not show a clear trend except for Group 6, where a positive trend was noted until the mid-20th century, followed by predominantly negative trends. Furthermore, the pre-monsoon period experienced a negative departure in the last two decades (2001–2017) across most parts of the country.

These observed trends in rainfall patterns carry implications for various sectors, especially agriculture in rainfed regions. The negative trends in post-monsoon rainfall and the weakening of NEM in certain areas underscore the need for ongoing monitoring and adaptation strategies to mitigate potential impacts on water resources and agricultural practices in these regions. In conclusion, the observed trends in rainfall patterns across India, including declining annual and southwest monsoon season rainfall in specific regions and negative trends in post-monsoon periods, highlight potential challenges for water resources and rainfed agriculture, necessitating ongoing monitoring and adaptive strategies.

3.2 Characteristics of extreme rainfall in different gridded datasets over India during 1983–2015, Suman Bhattacharyya, S. Sreekesh, Andrew King, 2021

Brief: This study undertakes a comprehensive assessment of twelve gridded rainfall datasets, categorized into gauge-based, satellite-derived, reanalysis, and merged products, to analyze the spatial and temporal variations of extreme rainfall events in India from 1983 to 2015. The evaluation includes comparing these datasets against a high-resolution, gauge-based dataset from the India Meteorological Department (IMD). The analysis involves calculating eleven extreme climate indices, spanning magnitude, frequency, and duration, based on the World Meteorological Organization's Expert Team on Climate Change Detection and Indices (ETCCDI) and IMD definitions. Results indicate significant uncertainties and underestimation of higher extreme events in most datasets compared to the IMD reference data, with GPCC and MSWEP exhibiting better performance in capturing the magnitude, duration, and frequency of extreme rainfall events in India.

The study highlights the challenges and variations in representing extreme rainfall events across different datasets, emphasizing the importance of dataset selection for reliable assessments. It notes consistent underestimation over the northern Himalayan region and underscores the limitations of gridded datasets in capturing spatial patterns of observed trends in extreme rainfall indices when compared to high-resolution gauge-based datasets. These findings contribute valuable insights for researchers and practitioners relying on gridded rainfall datasets for hydro-climatological studies in India, emphasizing the need for cautious interpretation and potential improvements in these datasets to better represent extreme rainfall characteristics.

3.3 Drought Identification and Trend Analysis Using Long-Term CHIRPS Satellite Precipitation Product in Bundelkhand, India, Varsha Pandey, Prashant K Srivastava, ORCID, Sudhir K Singh, George P., Rajesh Kumar Mall

Brief: In conclusion, the study utilized high-resolution satellite-derived CHIRPS data to assess meteorological drought in the Bundelkhand region of Uttar Pradesh, India. The Standardized Precipitation Index (SPI) was calculated at different time scales (1-, 3-, 6-, and 12-months), and drought events were identified using run theory. The Mann-Kendall (MK) test was employed for trend analysis at annual and seasonal scales.

The results revealed that the Bundelkhand region experienced an average of nine severe drought events over the 38-year study period, with the most intense drought recorded in the

Jalaun district during 1983–1985. Significant decreasing trends were observed in the SPI1 during the post-monsoon season, indicating an increasing severity of meteorological drought in the area. The findings highlight the need for effective drought management plans to address water crisis, and food security, and improve the well-being of the inhabitants in the Bundelkhand region.

The study contributes to the understanding of spatio-temporal drought patterns in a region known for its recurring drought episodes. The use of high-resolution satellite data and advanced drought indices like SPI enhances the accuracy and reliability of drought monitoring. The findings can guide policymakers in developing intervention strategies for water resource management, agriculture, and pastoral plans to mitigate the impact of drought in the Bundelkhand region.

3.4 Performance Analysis of IMD High-Resolution Gridded Rainfall (0.25° × 0.25°) and Satellite Estimates for Detecting Cloudburst Events over the Northwest Himalayas, Pravat Jena, Sourabh Garg, Sarita Azad, 2020.

Brief: This study focuses on evaluating the performance of high-resolution satellite datasets, such as CHIRPS and PERSIANN-CCS, along with IMD gridded data, in monitoring cloudburst events over the northwest Himalayas (NWH) region. The critical concern is the reliability of IMD gridded data, which is the only observed source available for detecting and predicting such events at ungauged locations. The study emphasizes the need to assess the accuracy of IMD data and compares it with satellite-derived datasets, particularly for cloudburst events, which have not been previously evaluated over the NWH region.

The identification of cloudburst events relies on critical rainfall thresholds calculated as percentiles from IMD data. The paper discusses the challenges associated with the IMD gridded dataset, such as low station density in complex terrains, and highlights the importance of improving interpolation techniques and station maintenance to enhance data quality in remote areas. The evaluation extends to seven high-resolution satellite products, and a new metric called improved Probability of Detection (IPOD) is introduced to account for temporal lags between observed and satellite estimates during cloudburst events.

The findings suggest that CHIRPS, a fine-scale satellite product, demonstrates a higher probability of detecting cloudburst events (60.5%–78.6%) using the IPOD metric, making it a more suitable option for monitoring extreme rainfall events over the NWH region. The study underscores the significance of evaluating different datasets for accurate monitoring and prediction of cloudbursts in complex terrains like the Himalayas.

CHAPTER 4

Data and Methodologies

4.1 Data Source

• The CHIRPS dataset provides high-resolution, long-term precipitation data.

• The dataset for this study includes daily rainfall values from 2012 to 2023.

• Additional information is obtained from the India Meteorological Department (IMD) website: www.imd.gov.in..

4.2 Data Overview

- Variables: Rainfall (in mm), Latitude, Longitude, Time (daily).
- Format: NetCDF (.nc).
- Coverage: Spatial coverage over India with high-resolution grids.

```
         LONGITUDE  LATITUDE       TIME  RAINFALL
0             66.5       6.5 2023-01-01       NaN
1             66.5       6.5 2023-01-02       NaN
2             66.5       6.5 2023-01-03       NaN
3             66.5       6.5 2023-01-04       NaN
4             66.5       6.5 2023-01-05       NaN
...            ...       ...        ...       ...
6356470      100.0      38.5 2023-12-27       NaN
6356471      100.0      38.5 2023-12-28       NaN
6356472      100.0      38.5 2023-12-29       NaN
6356473      100.0      38.5 2023-12-30       NaN
6356474      100.0      38.5 2023-12-31       NaN

[6356475 rows x 4 columns]
```

Fig: Data Structured in .nc file

4.3 Techniques and Tools

- **Data Extraction**: Python libraries such as xarray and pandas were used to extract and preprocess the data.

Code:

```
import xarray as xr
import pandas as pd
file_path = 'path_to_nc_file.nc'
dataset = xr.open_dataset(file_path)
df = dataset.to_dataframe().reset_index()
df['TIME'] = pd.to_datetime(df['TIME'])
df['MONTH'] = df['TIME'].dt.month
df['YEAR'] = df['TIME'].dt.year
data = df[['YEAR', 'MONTH', 'LATITUDE', 'LONGITUDE', 'RAINFALL']]
data = data.dropna(subset=['RAINFALL'])
```

- **Visualization**: Matplotlib, Seaborn, and Cartopy were employed to plot rainfall maps and trends.

Code:

```
import matplotlib.pyplot as plt
import cartopy.crs as ccrs
plt.figure(figsize=(10, 8))
ax = plt.axes(projection=ccrs.PlateCarree())
ax.coastlines()
sc = plt.scatter(
    data['LONGITUDE'], data['LATITUDE'], c=data['RAINFALL'], cmap='Blues', s=10, transform=ccrs.PlateCarree()
)
plt.colorbar(sc, label='Rainfall (mm)')
plt.title('Rainfall Distribution')
plt.show()
```

- **Machine Learning**: Random Forest Regression was utilized for predictive modeling.

Code:

```
import xarray as xr
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
# Function to process a single .nc file
def process_nc_file(file_path):
    dataset = xr.open_dataset(file_path)
    df = dataset.to_dataframe().reset_index()
    df['TIME'] = pd.to_datetime(df['TIME'])
    df['MONTH'] = df['TIME'].dt.month
    df['YEAR'] = df['TIME'].dt.year
    df = df[['YEAR', 'MONTH', 'LATITUDE', 'LONGITUDE', 'RAINFALL']]
    return df
# Paths to the .nc files for years 2014 to 2023
file_paths = [
    "F:\\CSIR project\\RF25_ind2014_rfp25.nc",
    "F:\\CSIR project\\RF25_ind2015_rfp25.nc",
    "F:\\CSIR project\\RF25_ind2016_rfp25.nc",
    "F:\\CSIR project\\RF25_ind2017_rfp25.nc",
    "F:\\CSIR project\\RF25_ind2018_rfp25.nc",
    "F:\\CSIR project\\RF25_ind2019_rfp25.nc",
    "F:\\CSIR project\\RF25_ind2020_rfp25.nc",
    "F:\\CSIR project\\RF25_ind2021_rfp25.nc",
```

```
    "F:\\CSIR project\\RF25_ind2022_rfp25.nc",
    "F:\\CSIR project\\RF25_ind2023_rfp25.nc",
]
# Process all files and combine into a single DataFrame
rainfall_data = pd.concat([process_nc_file(fp) for fp in file_paths], ignore_index=True)
# Filter for June, July, August, September (6, 7, 8, 9)
rainfall_data_filtered = rainfall_data[rainfall_data['MONTH'].isin([6, 7, 8, 9])]
# Group by year and month to get average rainfall
data = rainfall_data_filtered.groupby(['YEAR', 'MONTH']).agg({'RAINFALL': 'mean'}).reset_index()
# Handle missing values in the dataset
data = data.dropna(subset=['RAINFALL'])
# Prepare features (X) and target (y)
X = data[['YEAR', 'MONTH']]
y = data['RAINFALL']
# Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Train the Random Forest Regressor
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
# Evaluate the model
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
```

4.3 Regional Analysis

North India: North India, encompassing the Himalayan region and the Indo-Gangetic plains, experiences significant rainfall variability. This region is predominantly influenced by the southwest monsoon and western disturbances.

South India: South India receives rainfall from both the southwest monsoon and the northeast monsoon, with the Western Ghats playing a pivotal role in shaping rainfall distribution.

East India: East India, characterized by high humidity and dense forests, receives substantial rainfall due to the Bay of Bengal branch of the southwest monsoon.

West India: West India, dominated by arid and semi-arid regions, sees stark contrasts in rainfall between coastal areas and interior zones.

Central India: Central India forms the heart of the Indian subcontinent, experiencing moderate rainfall, crucial for its agrarian economy.

4.4 Methodology

Data Analysis

1. Data Cleaning

    - Handled missing values by applying advanced imputation techniques to ensure data completeness.
    - Identified and removed anomalies or outliers using statistical methods such as Z-score analysis and interquartile range (IQR).
    - Standardized the data for uniformity, making it suitable for analysis and machine learning.

2. Exploratory Data Analysis (EDA)

    - Investigated long-term rainfall trends and their seasonal variability across regions.
    - Examined the influence of geographical and climatic factors on rainfall distribution.

- Analyzed correlations between time (month, year), geographical coordinates, and rainfall intensities.

3. Visualization

    - Heatmaps: Illustrated rainfall distribution across different regions to identify hotspots and drought-prone areas.
    - Line Plots: Highlighted temporal changes in rainfall patterns over the years.
    - Bar Charts: Depicted monthly and seasonal variations in rainfall to provide an intuitive understanding of trends.
    - Interactive visualizations allowed deeper insights into spatial-temporal relationships in rainfall data.

Machine Learning Approach

1. Feature Selection

- Selected relevant features such as year, month, latitude, and longitude to build an effective predictive model.
- Incorporated derived features like seasonal indices to capture hidden patterns in rainfall variability.

2. Model Development

- Used Random Forest Regression, a robust ensemble-based algorithm, to predict rainfall based on historical data.
- Tuned hyperparameters like the number of trees and maximum depth to enhance model performance and reduce overfitting.
- Considered alternative models (e.g., Gradient Boosting, XGBoost) for comparative analysis.

3. Evaluation Metrics

- Mean Squared Error (MSE): Measured prediction accuracy and quantified errors in rainfall forecasts.

- R-squared Value: Assessed the proportion of variance in rainfall explained by the model.
- Conducted k-fold cross-validation to ensure reliability and generalizability of the model

Tools and Technologies Used

1. Python Libraries

- xarray: Efficiently handled large-scale multi-dimensional data from NetCDF files.
- pandas: Processed tabular data and performed aggregations.
- numpy: Conducted numerical operations and statistical calculations.
- matplotlib & seaborn: Created high-quality static and interactive visualizations.
- scikit-learn: Built and evaluated machine learning models for predictive analysis.

2. Software

- Jupyter Notebook: Provided an interactive environment for coding, visualizing, and documenting the workflow.
- Spyder (Optional): Used for debugging and testing the scripts in a development-friendly IDE.

3. Mapping Tools

- Cartopy: Visualized spatial patterns, creating maps to represent rainfall trends across India.
- GeoPandas: Integrated geospatial data for further spatial analysis.
- Shapefiles: Added regional boundaries and administrative divisions to enhance map details.

CHAPTER 5

5. Result and Analysis

5.1 Rainfall Trends (2012-2023)

This study aims to compute the mean monthly precipitation values for the years 2012 to 2023, highlighting the distinctive rainfall patterns across India. By analyzing the average rainfall for each month, the study provides a detailed understanding of the typical precipitation distribution throughout these years. This analysis is vital for critical sectors such as agriculture, water resource management, and disaster preparedness.

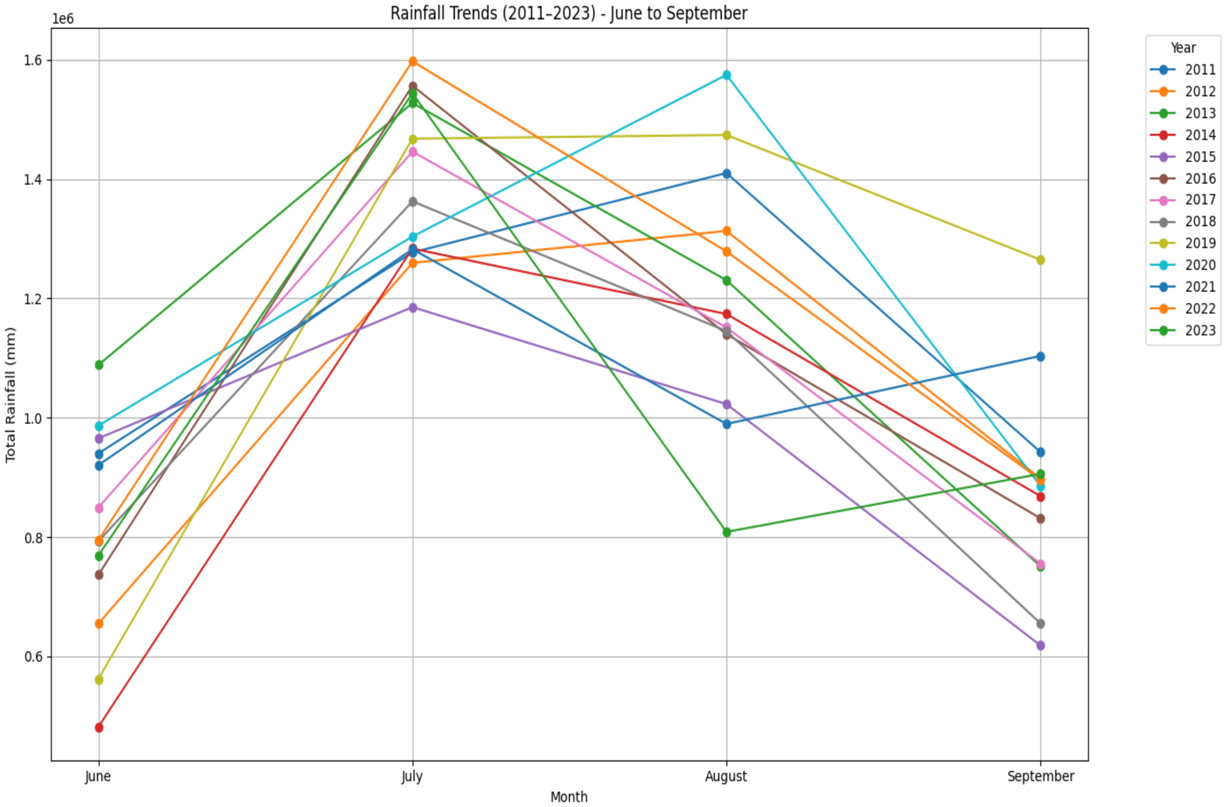

Fig : Rainfall Trends(2011-2023) - June to September

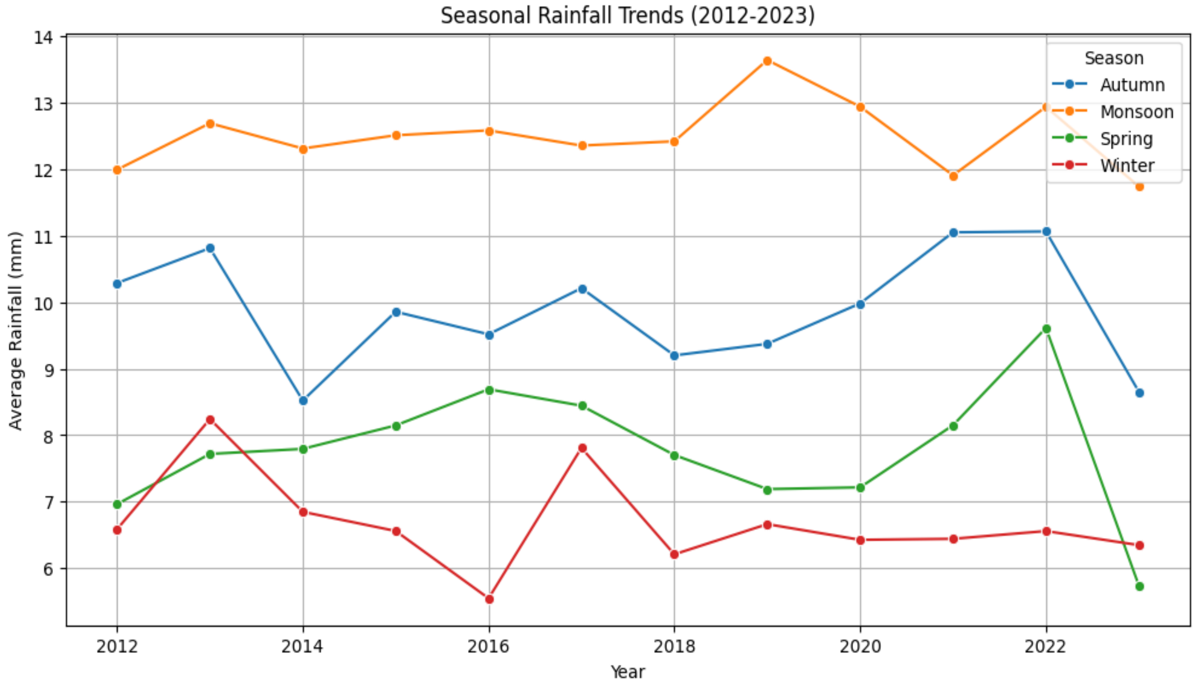

Fig : Seasonal Rainfall Trends (2012 -2023)

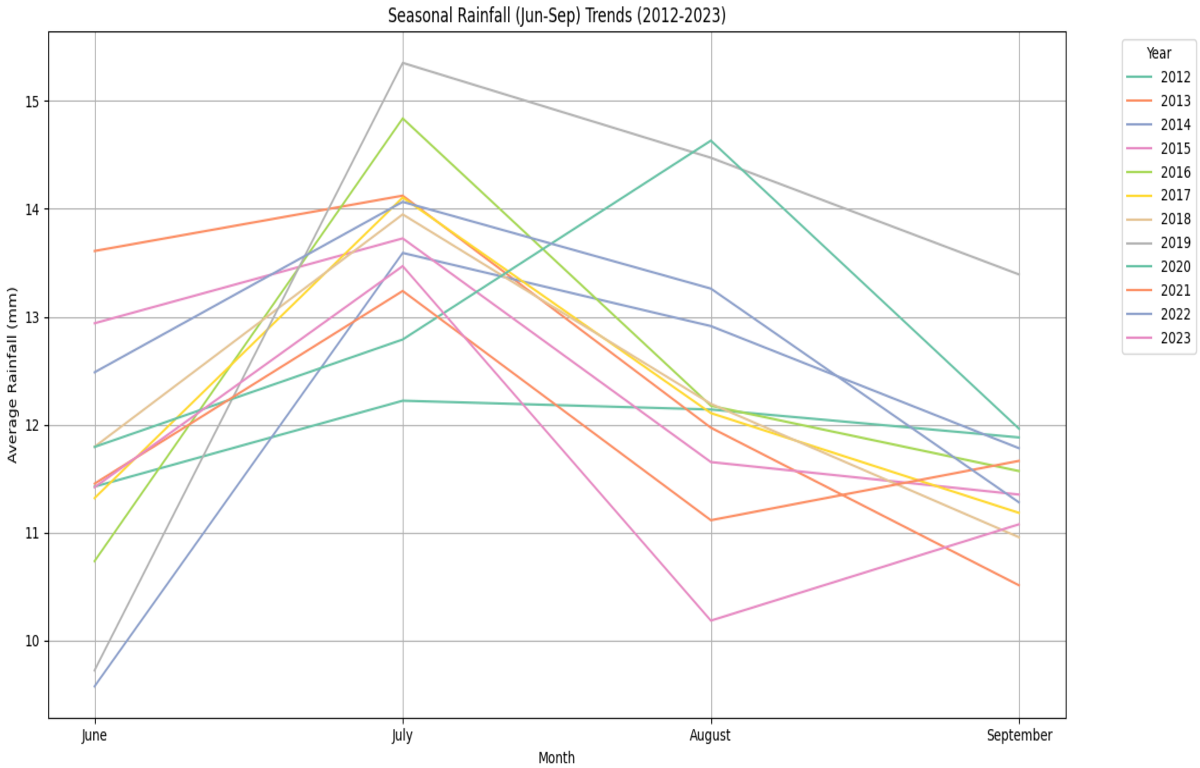

Fig : Seasonal Rainfall (June-sep) Trends (2012-2023)

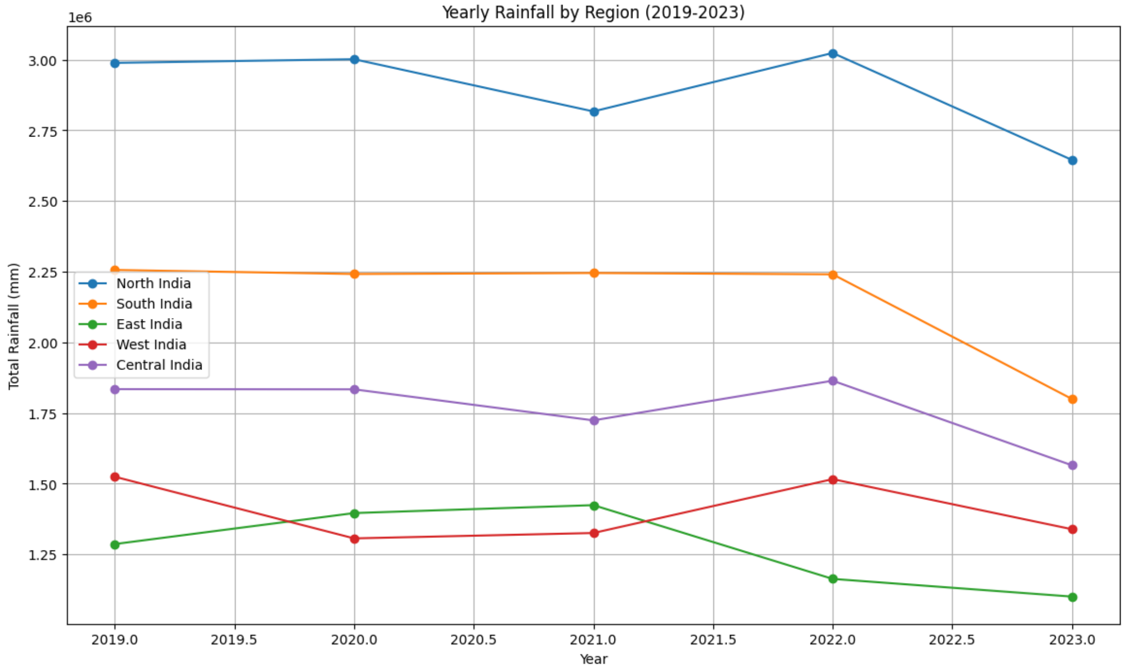

Fig : Yearly Rainfall by Reagion (2019 -2023)

- North and East India exhibit the highest rainfall during the monsoon months (June-September).
- Central and South India show moderate rainfall with noticeable inter-annual variability.

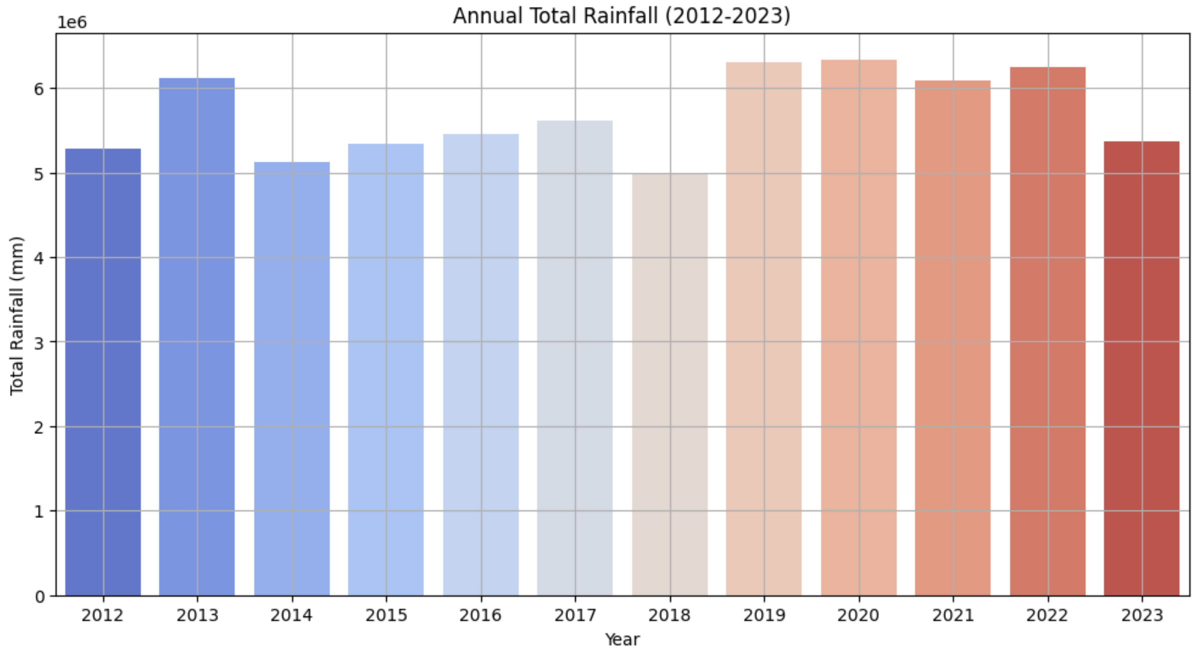

Fig : Annual Total Rainfall (2012-2023)

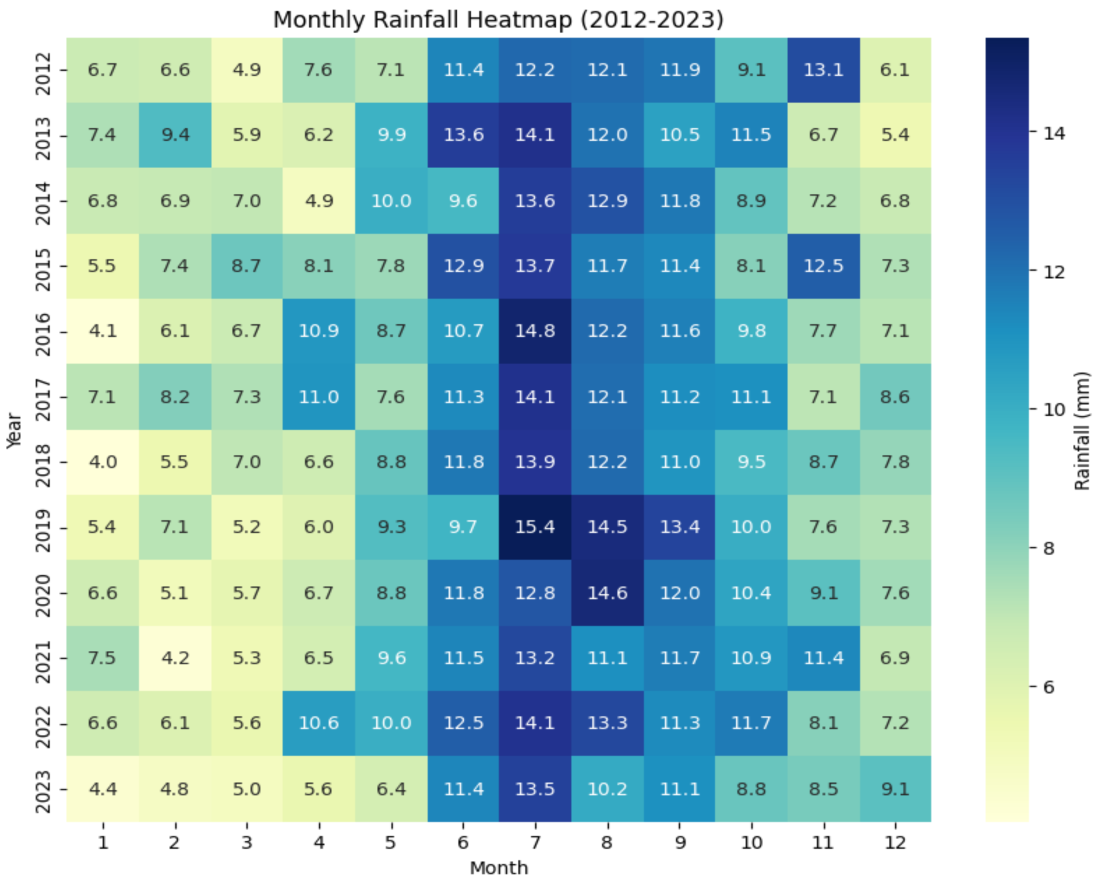

Fig : Monthly Rainfall Heatmap (2012 -2023)

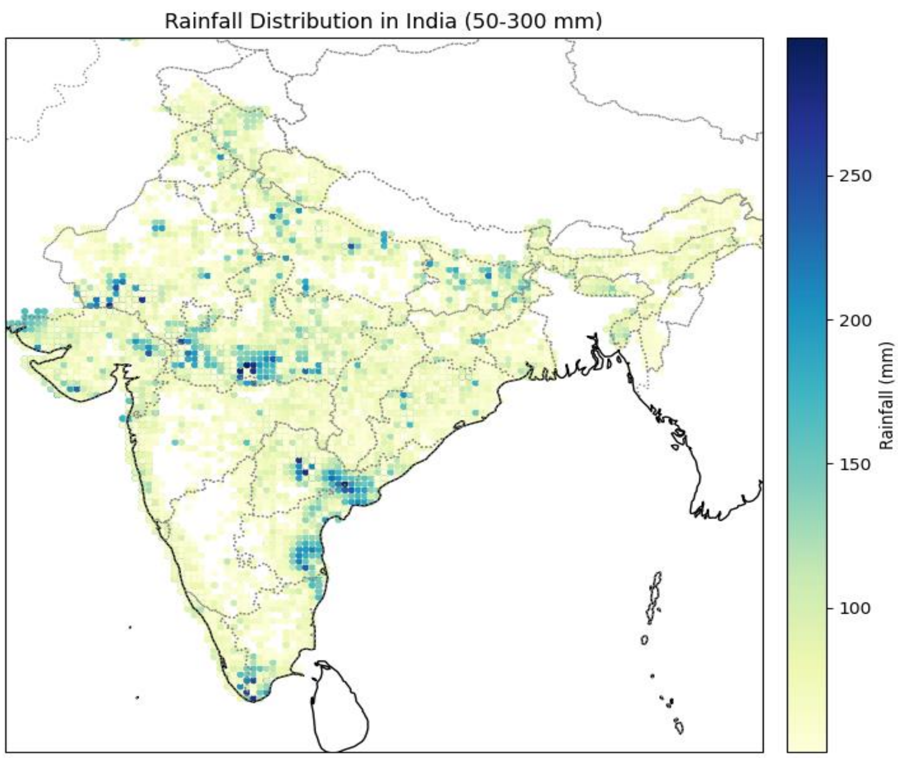

Fig: 2023 Rainfall Distribution

5.2 Predictive Analysis

- The predicted rainfall for 2025 indicates a peak in July, with precipitation levels significantly higher than in other months of the monsoon season. However, the model suggests a consistent decrease in rainfall from July to September, highlighting potential variability in the monsoon pattern.

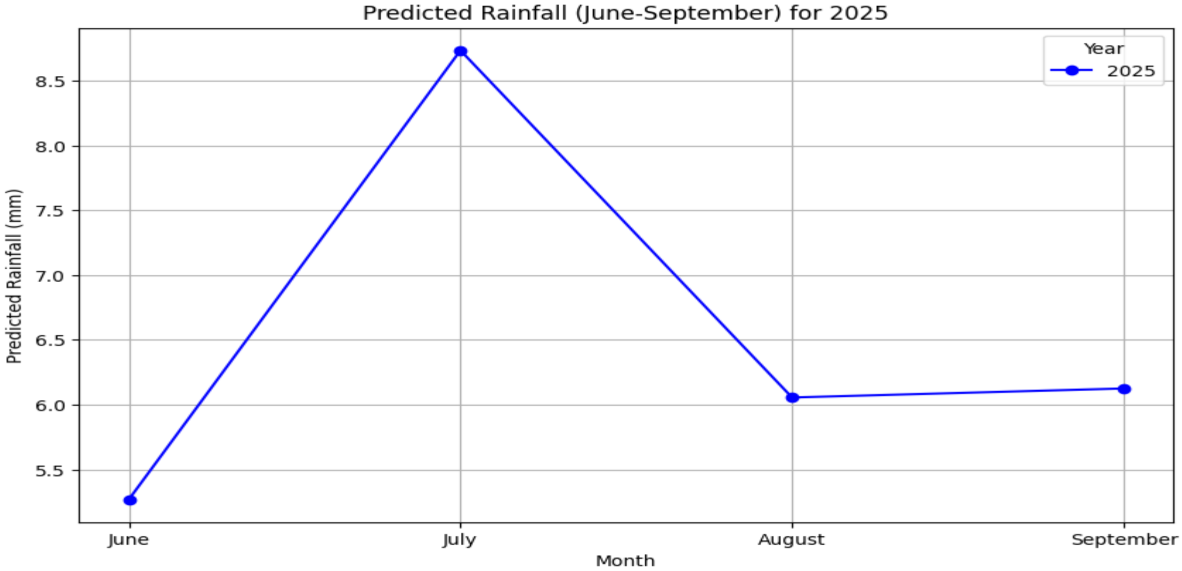

Fig : Predicted Rainfall (June -Sep) for 2025 by analyzing 2012 to 2023 Rainfall

Suggestions for Agriculture and Crop Cultivation:

1. Sowing Schedule:
   - Plan the sowing of crops that require significant water (e.g., paddy or maize) around late June or early July to take advantage of the peak rainfall in July.
   - For crops with lower water requirements, sowing can be done in late August or September.
2. Water Management:
   - Use rainwater harvesting techniques during July to store excess water for use in the drier months (August and September).
   - Promote the use of drip irrigation and mulching to conserve moisture and minimize water loss.

3. Crop Selection:
    - opt for water-intensive crops (like paddy or sugarcane) in areas with reliable irrigation systems during July.
    - In August and September, consider planting short-duration crops like millets or pulses that are better suited for low rainfall periods.
4. Fertilizer Application:
    - Time fertilizer application just before or during the July rainfall to maximize nutrient absorption.
    - Avoid excessive fertilization during the dry months to reduce the risk of soil nutrient depletion.
5. Intercropping and Crop Rotation:
    - Use intercropping methods with drought-tolerant crops to ensure productivity even in months with less rainfall.
    - Implement crop rotation strategies to maintain soil fertility, especially for regions relying on rainfall.
6. Soil Preparation:
    - Prepare the soil in May or early June to capture initial rainfall effectively in June.
    - Use organic manure or compost to enhance soil water retention for the drier months.
7. Livelihood Diversification:
    - Farmers in rainfed areas should consider alternative income sources, such as poultry, fisheries, or agroforestry, to mitigate risks associated with low rainfall in August and September.

6. Conclusion

- The data indicates that July remains the critical month for monsoon rainfall, emphasizing its importance for agriculture and hydrological systems.

- The declining trend in total rainfall, particularly in recent years, suggests potential challenges in sustaining water availability and highlights the need for proactive measures in climate adaptation and water conservation.

- The predictive model for 2025 aligns with historical patterns, reinforcing the importance of July for rainfall-dependent activities, while also suggesting caution for August and September due to relatively lower rainfall predictions

- **Seasonal Trends**: Monsoon consistently shows the highest average rainfall across the years (2012–2023), while Winter experiences the lowest. Seasonal variations are prominent, with Spring and Autumn displaying intermediate and stable rainfall patterns.
- **Monthly Variations (Heatmap)**: The heatmap highlights significant rainfall peaks during mid-year months (likely June to September), aligning with Monsoon season, while early-year and late-year months (e.g., January, December) exhibit comparatively low rainfall.
- **Yearly Rainfall Trends**: Total yearly rainfall remains relatively stable, with slight fluctuations. Notably, 2013 experienced the highest rainfall, whereas 2023 saw a decline, marking the lowest total rainfall over the years analyzed.

7. References

1. **India Meteorological Department (IMD)**. Official Website: www.imd.gov.in.
2. **India Meteorological Department (IMD), Pune**. CHIRPS Data: Rainfall data in NetCDF format.
3. **CSIR Fourth Paradigm Institute (CSIR-4PI)**. Studies on climate modeling and data analysis.
4. **National Remote Sensing Centre (NRSC)**. Resources on spatial data and rainfall mapping.
5. **World Meteorological Organization (WMO)**. Guidelines on rainfall prediction and climate models.
6. **Climate Hazards Group**. CHIRPS Documentation and Tools.
7. Funk, C., et al. *The Climate Hazards InfraRed Precipitation with Stations—A New Environmental Record for Monitoring Extremes*. *Scientific Data*, 2015.
8. Guhathakurta, P., & Rajeevan, M. *Trends in the Rainfall Pattern over India*. *Current Science*, 2008.