

# Forecasting of Southwest Indian Summer Monsoon Rainfall Using Machine Learning Techniques

A DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR  
THE AWARD OF DEGREE

OF

MASTER OF TECHNOLOGY

IN

Computer Science and Engineering

*Submitted By:*

**Aman Kumar**

**Roll. No. 19203003**

*Under the supervision of:*

**Dr. Samayveer Singh**



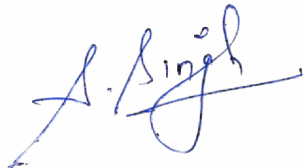
Department of Computer Science and Engineering  
Dr. B.R. Ambedkar National Institute of Technology  
Jalandhar - 144011 (Punjab) India  
August 2021

## CANDIDATE'S DECLARATION

I hereby declare that the work presented in the dissertation titled '**Forecasting of Southwest Indian Summer Monsoon Rainfall Using Machine Learning Techniques**' submitted towards the partial fulfilment of the requirements for the award of the degree of **Master of Technology** from the Department of Computer Science and Engineering, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, India, is an authentic record of my own work carried out in the period of August 2020 to August 2021 under the sincere guidance of Dr. Samayveer Singh. It is also stated that no earlier submission of the subject matter of the work demonstrated in this dissertation has been made for the award of any other degree in this or any other University/Institute.

*Aman Kumar*  
Aman Kumar  
Roll No.: 19203003

This is to certify that above statement made by the candidate is correct and true to best of my knowledge.



Dr. Samayveer Singh (Supervisor)  
Assistant Professor,  
Department of Computer Science and Engineering,  
Dr. B R Ambedkar NIT Jalandhar,

The M.Tech (Dissertation) Viva-Voce examination of Aman Kumar, Roll No. 19203003, has been held on \_\_\_\_\_ and accepted.

\_\_\_\_\_  
External Examiner

\_\_\_\_\_  
Supervisor


\_\_\_\_\_  
HOD, CSE

# ACKNOWLEDGEMENTS

First and Foremost, I would like to express my gratitude to my supervisors Dr. Samayveer Singh, Department of Computer Science and Engineering, Dr B R Ambedkar National Institute of Technology, Jalandhar, for the useful comments, remarks and engagement throughout the duration of my research work. I cannot thank him enough for his tremendous support and help. He motivated and encouraged me throughout this work. I consider myself extremely fortunate to have a chance to work under his supervision.

I also wish to thank wholeheartedly all the faculty members of the Department of Computer Science and Engineering for the invaluable knowledge they have imparted on me and for teaching the principles in most exciting and enjoyable way. I also extend my thanks to the technical and administrative staff of the department for maintaining an excellent working facility.

I would like to thank my family for their continuous support and blessings throughout the entire process. I would also like to extend thanks to my friends for the useful discussions, constant support and encouragement during whole period of the work.

  
Aman Kumar  
Roll No. 19203003

# ABSTRACT

A significant part of the Indian economy is still reliant on agriculture, which is directly linked to rainfall. Southwest monsoon precipitation over the nation is mostly consistent. The country's average seasonal rainfall is approximately 88 cm (1961-2010 period mean value), with a variance coefficient of about 10%. A year above (less than) 110% (90%) rainfall is known as excess (drought) year. Currently, IMD (Indian Meteorological Department) is using EMR (Ensemble Multiple Regression) and PPR (Projection Pursuit Regression) different potential climate parameters relationships with ISMR and various machine learning models. A sliding window is used for forecasting ISMR and also to ML models using 10 climate parameter.

An auto-encoder is an architecture belonging to the class of artificial neural networks, we used this architecture to acquire the intricate features of data, and utilized it for the reduction in dimensionality of the data. To recreate the SLP (sea surface pressure) input parameter this ANN model is used and to get the 6 major correlated SLP vectors for ML model training and forecasting using 16-year sliding window. The vectors of actual observed ISMR and model forecasted rainfall were found correlated with 0.829 Pearson correlation coefficient.

An analysis of percentage deviation of ISMR data from seasonal mean rainfall for the period (1991-2019) which shows recent deficit, normal, and excess monsoon rainfall years. Monthly statistical analysis of historical Indian rainfall data is done on the basis of 36 meteorological subdivisions, 7 homogeneous regions and the country as a whole.

# Contents

<b>Candidate's Declaration</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Publication</b>	<b>x</b>
<b>List of Abbreviations</b>	<b>xi</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 Monsoon . . . . .	1
1.2.1 Inter-seasonal variability of Indian summer monsoon . . . . .	2
1.2.2 Inter-annual variability of Indian summer monsoon . . . . .	2
1.3 Indian Meteorological Department . . . . .	3
1.4 Distribution of Rainfall . . . . .	4
1.5 Performance Measures . . . . .	5
1.5.1 Root Mean Squared Error (RMSE) . . . . .	5
1.5.2 Residual Standard Error (RSE) . . . . .	5
1.5.3 Mean Absolute Error (MAE) . . . . .	5
1.6 Motivation . . . . .	6
1.7 Research Objectives . . . . .	6
1.8 Dissertation Organization . . . . .	7
<b>Chapter 2 Literature Survey</b>	<b>8</b>
2.1 Overview . . . . .	8

2.2	EL Nino-Southern Osillation (ENSO) . . . . .	8
2.2.1	EL NINO . . . . .	8
2.2.2	LA NINA . . . . .	9
2.2.3	Neutral . . . . .	9
2.3	Indian Ocean Dipole (IOD) Links With The ISMR . . . . .	9
2.3.1	Positive Phase . . . . .	11
2.3.2	Negative Phase . . . . .	11
2.4	Literature Survey . . . . .	11
2.5	Different Monsoon Rainfall Forecasting Methods Used By Climate Researchers . . . . .	15
2.6	How IMD Forecast Indian Summer Monsoon . . . . .	18
2.7	Research Gaps . . . . .	19
2.8	Chapter Summary . . . . .	19
<b>Chapter 3 Statistical Analysis of Historical Rainfall's Data</b>		<b>20</b>
3.1	Overview . . . . .	20
3.2	All India Rainfall . . . . .	20
3.3	7 Homogeneous Regions Rainfall . . . . .	20
3.4	Analysis of the Rainfall of Meteorological Subdivisions . . . . .	25
3.5	Chapter Summary . . . . .	31
<b>Chapter 4 Data Extraction &amp; Prediction Methodology</b>		<b>32</b>
4.1	Overview . . . . .	32
4.2	Data Extraction Sources . . . . .	32
4.2.1	Anomaly Caluculation . . . . .	35
4.3	Machine Learning . . . . .	35
4.3.1	Support Vector Regressor . . . . .	35
4.3.2	Multiple Linear Regression . . . . .	36
4.3.3	Ridge Regression . . . . .	37
4.3.4	Stacked Auto-encoder . . . . .	37
4.4	Rainfall Forecasting Approach . . . . .	38
4.4.1	Approach - I . . . . .	38
4.4.2	Approach - II . . . . .	39

4.5 Chapter Summary . . . . .	41
<b>Chapter 5 Forecasting Results of Different Models and Their Comparative Analysis</b>	<b>42</b>
5.1 Overview . . . . .	42
5.2 Linkage Between Predictors and the ISMR Monthly Data . . . . .	42
5.3 Approach-I Results . . . . .	44
5.4 Approach-II Results . . . . .	46
5.5 Chapter Summary . . . . .	53
<b>Chapter 6 Conclusion and Future Scope</b>	<b>54</b>
<b>References</b>	<b>55</b>

# List of Figures

Fig. 1.1	Normal dates of starting monsoon rainfall in various regions.[Source: <a href="https://imdpune.gov.in/">https://imdpune.gov.in/</a> ] . . . . .	3
Fig. 2.1	Mechanism of all the three phases of El Niño–Southern Oscillation i)Positive ii)Neutral iii) Negative [Source :NOAA/PMEL/TAO Project Office, Dr. Michael J. McPhaden, Director] . . . . .	8
Fig. 2.2	Fluctuation of Indian Ocean Dipole index over the period of 1990-2019 [Source: <a href="http://enformtk.u-aizu.ac.jp">http://enformtk.u-aizu.ac.jp</a> ] . . . . .	10
Fig. 2.3	Sea surface temperature currents of Eastern and Western Indian Ocean mean value over [1981-2010] [Source: <a href="https://fews.net/indian-ocean-dipole">https://fews.net/indian-ocean-dipole</a> ] . . . . .	10
Fig. 2.4	ISMR percent deviation (PD) from mean rainfall, compared with Jan- Mar SST's PD and OMT's PD in south West Indian Ocean region for the period [1993-1997] . . . . .	13
Fig. 2.5	Geographical locations of 8 climate predictors used by IMD for LRF forecasting of ISMR [Source: <a href="https://imdpune.gov.in">https://imdpune.gov.in</a> ] . . . . .	18
Fig. 3.1	ISMR Percent Deviation from long term mean(88CM) . . . . .	21
Fig. 3.2	Long term Mean and Std of ISMR . . . . .	21
Fig. 3.3	Western Himalaya . . . . .	23
Fig. 3.4	North-West India . . . . .	23
Fig. 3.5	Central North-East India . . . . .	23
Fig. 3.6	East Coast . . . . .	24
Fig. 3.7	North-Central . . . . .	24
Fig. 3.8	Peninsula . . . . .	24
Fig. 3.9	West Coast . . . . .	25
Fig. 4.1	Showing SVM plane and SVM Margin . . . . .	36
Fig. 4.2	Architecture of Auto-encoder used to recreate the input vector . . . . .	38



Fig. 4.3	Steps involved in training Simple regression model using 10 predictors . . .	39
Fig. 4.4	Steps involved in training Simple regression model using 10 predictors . . .	40
Fig. 4.5	Layered architecture of auto-encoder model . . . . .	40
Fig. 5.1	30 Years running PCC between ISMR and climate predictors for the period [1980-2018] . . . . .	43
Fig. 5.2	ISMR forecasting using SVR . . . . .	45
Fig. 5.3	ISMR forecasting using MLR . . . . .	45
Fig. 5.4	ISMR forecasting using Ridge Regression . . . . .	45
Fig. 5.5	North-Central Indian rainfall forecasting using SVR . . . . .	46
Fig. 5.6	North-Central Indian rainfall forecasting using MLR . . . . .	46
Fig. 5.7	North-Central Indian rainfall forecasting using Ridge Regression . . . . .	46
Fig. 5.8	Indian East coastal region rainfall forecasting using SVR . . . . .	47
Fig. 5.9	Indian East coastal region rainfall forecasting using MLR . . . . .	47
Fig. 5.10	Indian East coastal region rainfall forecasting using Ridge Regression . . .	47
Fig. 5.11	Western Himalaya region rainfall forecasting using SVR . . . . .	48
Fig. 5.12	Western Himalaya region rainfall forecasting using MLR . . . . .	48
Fig. 5.13	Western Himalaya region rainfall forecasting using Ridge Regression . . .	48
Fig. 5.14	North-east region rainfall forecasting using SVR . . . . .	49
Fig. 5.15	North-east region rainfall forecasting using MLR . . . . .	49
Fig. 5.16	North-east region rainfall forecasting using Ridge Regression . . . . .	49
Fig. 5.17	North-west region rainfall forecasting using SVR . . . . .	50
Fig. 5.18	North-west region rainfall forecasting using MLR . . . . .	50
Fig. 5.19	North-west region rainfall forecasting using Ridge Regression . . . . .	50
Fig. 5.20	Interior Peninsula region rainfall forecasting using SVR . . . . .	51
Fig. 5.21	Interior Peninsula region rainfall forecasting using MLR . . . . .	51
Fig. 5.22	Interior Peninsula region rainfall forecasting using Ridge Regression . . .	51
Fig. 5.23	West coast region rainfall forecasting using SVR . . . . .	52
Fig. 5.24	West coast region rainfall forecasting using MLR . . . . .	52
Fig. 5.25	West coast region rainfall forecasting using Ridge Regression . . . . .	52
Fig. 5.26	ISMR forecasting using Approach-II . . . . .	52

# List of Tables

Table 1.1	Rainfall categories based upon % departure from Mean . . . . .	1
Table 2.1	A List of climate predictor used by IMD in both forecasting phase (Apr-Jun), with their Pearson correlation coefficient (PCC) [Source: <a href="https://imdpune.gov.in">https://imdpune.gov.in</a> ] . . . . .	16
Table 3.1	Statistical Analysis of the June month rainfall for the long period [1961-2010] . . . . .	26
Table 3.2	Statistical Analysis of the July month rainfall for the long period [1961-2010] . . . . .	27
Table 3.3	Statistical Analysis of the August month rainfall for the long period [1961-2010] . . . . .	28
Table 3.4	Statistical Analysis of the September month rainfall of various meteorological sub-divisions for the long period [1961-2010] . . . . .	29
Table 3.5	Statistical Analysis of the monsoon rainfall witnessed in various meteorological sub-divisions for the long period [1961-2010] . . . . .	30
Table 5.1	Correlation Matrix of Indian Regions Rainfall and Climate predictors. R1:Peninsula, R2:North East, R3:North-West, R4:West-coastal, R5: North-central, R6: East-Coast, R7:Western Himalaya . . . . .	43
Table 5.2	List of predictors used for Approach-1 . . . . .	44
Table 5.3	Observed rainfall VS forecated rainfall PCC analysis . . . . .	47

# LIST OF PUBLICATION

1. A. Kumar and S. Singh, "A Review on Indian Summer Monsoon Rainfall Prediction Using Machine Learning Techniques," 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), pp. 524-528, 2021, doi: 10.1109/ICSCCC51823.2021.9478104.

# LIST OF ABBREVIATIONS

**ISMR** Indian Summer Monsoon Rainfall

**IMD** Indian Meteorological Department

**AZM** Atlantic Zonal Mode

**AO** Arctic Oscillation

**SOI** Southern Oscillation Index

**EASP** East Asia Surface Pressure

**NASP** North Atlantic Surface Pressure

**PDO** Pacific Decadal Oscillation

**SVR** Support Vector Regressor

**PCC** Pearson Correlation Coefficient

**RMSE** Root Mean Squared Error

**ANN** Artificial Neural Network

**MLR** Multiple Linear Regression

**NCPZW** North Central Pacific Zonal Wind

**NAOSST** North Atlantic Ocean Sea Surface Temperature

**SEIO SST** South-east Indian Ocean Sea Surface Temperature

**WWV** Warm Water Volume

# CHAPTER 1

## INTRODUCTION

### 1.1 Overview

In India, the growth of the economy is directly impacted by agricultural products which further dependent on the monsoon. The contribution of the agricultural to national economy is declining every year due to low or excessive rainfall. In this chapter, we are going to discuss about Indian summer monsoon, factors affecting precipitation during summer monsoon, monsoon signals and monsoon forecasting methodology by IMD.

### 1.2 Monsoon

The precise definition of planetary-scale monsoon is a much debated issue. The monsoon is defined as a seasonal shift in wind direction, being derived from the Arabic word “mausim”, meaning season. The Arabian sea is generally considered to be in that area of the world where the name “monsoon” was first used to signify a seasonal wind regime in which surface winds blow persistently from one general direction in summer and just as persistently from a markedly different direction in winter. Monsoons can be viewed as three-dimensional circulations associated with the global distribution of land and sea.

Regions of the surface cool and heating the planet at various paces relying on the imbibing of incoming solar energy and its intensity. The water endue a significantly bigger heat ability than land and mountains that makes it more effective to accumulate more energy than the land surface or mountains and so keeps heat for longer periods of time. In the meantime of summer, due to its smaller specific heat and lower depth, the

Table 1.1: Rainfall categories based upon % departure from Mean

Category	Rainfall-Range (Percentage)
Drought	Less than 90%
Below-Normal	90% - 97%
Near-Normal	98% - 102%
Above-Normal	103% - 110%
Excess	Above 110%

land is heated more quickly than the neighboring ocean.

To preserve the equilibrium of the climate, the heats are often moved to the deficit from excess locations, and this is done through an event called the "Land-Sea Breeze" in the case of a soil water differential. On a bigger scale when there is a contrast of land-water, such as an ocean-circumcised continent, the hot air volumes increase, the density decrease in the vertical amounts, and the pressure levels grow in the summer heat build upon the land. High-pressure dense air dominates the surface of the ocean. Surface air from sea to ground, which results in a pressure gradient, will blow and build up the circulation of the summer monsoon.

The major factors on which the monsoon precipitation depends are given below:

- Land and water are exposed to sunlight in different period.
- The change in flow of Inter Tropical Convergence Zone (ITCZ).
- Experiencing the high-pressure over east of Madagascar.
- Tibetan plateau becomes extremely hot.
- Over the monsoon season, the jet stream travels north of the Himalayas, and the tropical-easterly jet stream flies over the Indian peninsula.

### **1.2.1 Inter-seasonal variability of Indian summer monsoon**

Several studies demonstrate that the distribution of precipitation over India changes significantly every day, whereas mainstream rains occur in periods under the presence of favorable circumstances. The total seasonal strength of the monsoon rainfall can therefore be associated with active phases and downtime. This irregular rain behavior is linked to almost 3-7 days, 10-20 days, and 30-60 days hierarchy. The period of 3 to 7 days is connected with the monsoon trough oscillation, while the period of 10-20 days is related to the synoptic-scale convective systems developed over the warm Bay of Bengal which causes considerably heavy precipitation. Active monsoon season spells are defined by the cluster sequence of these systems, although during break times no disruption occurs.

### **1.2.2 Inter-annual variability of Indian summer monsoon**

The highest quantity of rainfall takes place throughout India, known as Monsoon, from June through September. Agriculture is India's key foundation, and monsoon forecasting

helps to take loss preventive measurements. Heavy precipitation causes concerns with flooding and water crisis owing to precipitation of the shortfall. Regional prediction of monsoon precipitation helps to reduce the loss of men, money, and machines. During the June-September season (JJAS) the Southwest mountain winds produce an ample amount of moisture and rainfall across the Indian area.

Regional precipitation can change regularly, whereas the monsoon rainfall in South-West India is more or less consistent, and it is around 75% of the total precipitation over the country. Thus economic growth is directed by seasonal monsoon rains in most of the nations of South Asia, where agriculture is the primary source of jobs and employment for the bulk of citizens. Normally mean seasonal rainfall go up-down with a coefficient of variation of 10% over the country as a whole, based on this fluctuation rainfall categorization is shown in Table 1.1.

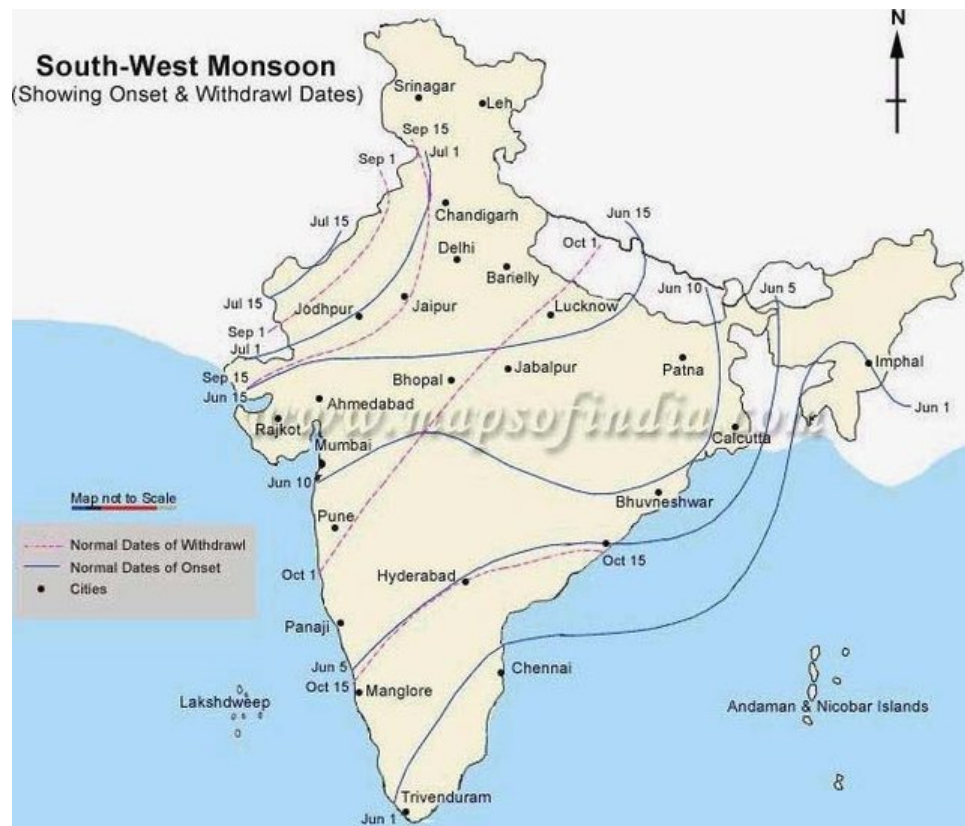


Figure 1.1: Normal dates of starting monsoon rainfall in various regions.[Source: <https://imd pune.gov.in/>]

### 1.3 Indian Meteorological Department

The India Meteorological Department (IMD) is a government of India department of the Ministry of Earth Sciences. It is the primary body in charge of meteorological

monitoring, forecasts, and seismology. The IMD is based in Delhi and runs multiple monitoring stations throughout India and Antarctica. Kolkata, Nagpur, Chennai, Mumbai, Guwahati, and New Delhi have regional offices. The IMD is also one of the World Meteorological Organization's six regional specialised meteorological centres. It is in the authority of anticipating, identifying, and disseminating tropical cyclone alerts in the Northern Indian Ocean region, listing the Bay of Bengal, the Malacca Straits, the Persian Gulf, and the Arabian sea. For more than a century, the India Meteorological Department (IMD) has issued long-range forecasts (LRF) based on the statistical methodologies for southwest monsoon rainfall across India (ISMR). Most of the statistical dynamical systems, including the IMD operational models, were uncapacious for forecasting the recent poor monsoon periods of 2002 and 2004.

Since, its early roots in 1875, IMD has gradually increased its facilities for meteorological observations, connectivity, prediction, and climate products, as well as its technological innovation. IMD has always employed cutting-edge technologies. Weather telegrams were widely used during the telegraph era for gathering observational data and conveying advisories. Later, IMD became India's first agency to have a message-switching computer to accommodate worldwide data interchange.

IMD was awarded one of the country's first-ever electronic computers for research areas in meteorology. INSAT, India's geostationary satellite, was the world's first developing nation to get its geostationary satellite for uninterrupted environmental observation of this region of the globe, notably for hurricane monitoring.

## 1.4 Distribution of Rainfall

The average annual rainfall in India is about 88 cm, but it has inherent spatial variations. Figure 1.1 shows the potential starting dates of the ISMR.

**Areas of high rainfall :** The heaviest monsoon rainfall happens near the west coast, in the Western Ghats, and in the sub-Himalayan highlands of the northeast and Meghalaya's hills. Rainfall in this area reaches 200 cm. Rainfall reaches 1,000 cm in certain areas of the Khasi and Jaintia hills. Rainfall in the Brahmaputra valley and surrounding hills is less than 200 cm.

**Areas of medium rainfall :** Rainfall occurs in east Tamil Nadu, the southern region of Gujarat, the northeastern Peninsula covering Odisha, the northern Ganga plain, Bihar,



Jharkhand, eastern Madhya Pradesh, along the sub-Himalayas and the Cachar Valley, and Manipur, vary from 100 to 200 cm.

**Areas of Low Rainfall :** Precipitation in the following regions; Delhi, Haryana, western Uttar Pradesh, eastern Rajasthan, the Deccan Plateau, Punjab, Gujarat, and in Jammu Kashmir vary within 50 and 100 cm. Inadequate Rainfall Areas: Parts of the Peninsula, Karnataka, particularly in Andhra Pradesh, and Maharashtra, as well as Ladakh and much of western Rajasthan, receive rainfall of less than 50 cm.

## 1.5 Performance Measures

Various ML models provide different results so to full-fill the need of comparative analysis, some performance matrices are discussed below:

### 1.5.1 Root Mean Squared Error (RMSE)

It is the average error made by the model while estimating the result of an observation. The RMSE is the square root of the mean squared error (MSE), which is the average squared difference between the observed real outcome values and the anticipated values by the model. The model performs better when the RMSE is low.

$$RMSE = \sqrt{\text{mean}(\text{observed} - \text{predicted})^2} \quad (1.1)$$

### 1.5.2 Residual Standard Error (RSE)

The model sigma, commonly known as the RMSE adjusted for the number of predictors in the model, is a version of the RMSE. The model performs better when the RSE is low. The difference between RMSE and RSE is negligible, especially for large multi-variate data sets.

### 1.5.3 Mean Absolute Error (MAE)

The MAE, like the RMSE, estimates the forecast error. It is the average absolute difference between the observed and anticipated outcomes. When compared to RMSE, MAE is less susceptible to outliers.

$$MAE = \text{mean}(\text{abs}(\text{observed} - \text{predicted})) \quad (1.2)$$

## 1.6 Motivation

As a result, the economic success of most South Asian nations, where agriculture employs the bulk of the residents, is directly associated with the seasonal monsoon rains. Variations in rainfall owing to monsoon whims can cause major droughts or, in many cases, severe floods over India, resulting in substantial socio-economic consequences. As a result, understanding, and forecasting the monsoon is critical.

- Every year, the distribution of rainfall causes droughts or floods in various regions of the nation.
- Soil erosion is a problem across large areas of India as a result of an unexpected monsoon burst.
- India's agricultural success is heavily reliant on timely and evenly distributed rainfall. Agriculture suffers if it fails, especially in areas where irrigation systems have not been established yet.

Because traditional monsoon precipitation forecasting methods do not sufficiently anticipate the quantitative analysis of precipitation throughout the country, substantial research on novel forecasting methods is being done. As a result of these factors, we decided to investigate the monsoon rainfall process and attempt to fix some current concerns and obstacles.

## 1.7 Research Objectives

The main objective of the dissertation is to understand and predict the southwest monsoon variability using oceanic and atmospheric parameters over Tropical ocean. The dissertation objectives have been formulated as follows:

- To check the significance (using Pearson Correlation Coefficient) of currently used climate predictors with Indian summer monsoon rainfall and to search new climate predictors in order to improve statistical forecasting models.
- To develop machine learning model for forecasting the Southwest Indian Summer monsoon rainfall using best suited training and testing time series.

## 1.8 Dissertation Organization

The structure of the rest of the dissertation is organized as follows:

- In Chapter 2, a literature review of some existing monsoon rainfall forecasting methods based on different approaches and models of machine learning. Additionally, to provide an insight into various strengthening and weakening relationships predictors of Monsoon rainfall and a comparative analysis of previous work.
- In Chapter 3, a monthly statistical analysis of historical Indian rainfall data on the basis of 36 meteorological subdivisions, 7 homogeneous regions and the country as a whole.
- In Chapter 4, we discuss the data extraction methodology and the machine learning techniques used for forecasting purpose. We briefly describe the two forecasting approaches for training the ML models.
- In Chapter 5, tables and graphs plots depicts the findings of machine learning tools and the comparative analysis of used methodology.
- In Chapter 6, the concluding remarks are presented. Also, the future scope in the existing research work has been discussed.

# CHAPTER 2

## LITERATURE SURVEY

### 2.1 Overview

Climate peccancy is a natural occurrence that has a direct influence on the ISMR in terms of timing, distribution over the entire country, and severity. The goal of this chapter is to examine some previous research done in the domains of predicting Indian summer monsoon precipitation, strengthening and weakening linkages of climate variables, and determining the most associated climate variables.

### 2.2 EL Nino-Southern Osillation (ENSO)

Though ENSO is a single climatic event, it can exist in three states or phases. Because ENSO is a linked climatic event of the two opposing phases, “El NiNo” and “La Nia” necessitate changes in both the ocean and the atmosphere. “Neutral” is located in the center of the continuum. Because this current emerges around Christmas in December, the name EI-Nino means “Child Christ”. In Peru (Southern Hemisphere), December is a summer month. In India, EI-Nino is used to forecast long-term monsoon rainfall. In 1990-91, there was a wild EI-Nino event, and the start of the southwest monsoon was delayed by five to twelve days throughout much of the nation. Figure 2.1 shows the mechanism of all the three phase of El Niño–Southern Oscillation .

#### 2.2.1 EL NINO

In the central and eastern tropical Pacific Ocean, there is a warming of the ocean surface, or above-average sea surface temperatures (SST). Rainfall tends to decrease across

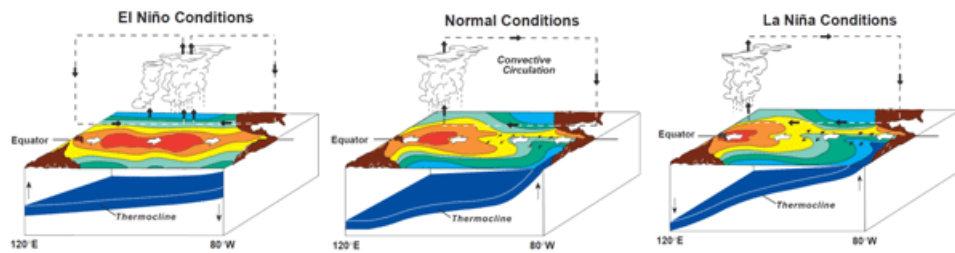


Figure 2.1: Mechanism of all the three phases of El Niño–Southern Oscillation  
i)Positive ii)Neutral iii) Negative [Source :NOAA/PMEL/TAO Project Office, Dr.  
Michael J. McPhaden, Director]

Indonesia, whereas it rises across the tropical Pacific Ocean. Low-level surface winds, which generally blow from east to west near the equator (“easterly winds”), diminish or, in certain situations, reverse direction (from west to east or “westerly winds”). Capotondi et al. (2015), C. Wang et al. (2017), Yeh et al. (2018) does research work to identify, describe, and understand these “El Niño types,” spurring debates on whether there are indeed two distinct modes of variability, or whether ENSO can be more aptly described as a diverse continuum. Santoso et al. (2017) provide a review of ENSO, its nature and dynamics, and through analysis of various observed key variables, we outline the processes that characterize its extremes.

### 2.2.2 LA NINA

In the central and eastern tropical Pacific Ocean, there is a cooling of the ocean surface, or below-average sea surface temperatures (SST). Rainfall tends to rise across Indonesia, whereas it falls across the central tropical Pacific Ocean. The regular easterly winds around the equator intensify.

### 2.2.3 Neutral

Neither El Niño or La Niña. Often tropical Pacific SSTs are generally close to average. However, there are some instances when the ocean can look like it is in an El Niño or La Niña state, but the atmosphere is not playing along (or vice versa).

## 2.3 Indian Ocean Dipole (IOD) Links With The ISMR

Neither El NINO nor La NINO exists if SSTs in the tropical Pacific are frequently near to average. However, there are times when the water appears to be in an El Nio or La Nia condition, but the atmosphere is not cooperating (or vice versa). Figure 2.2 depicts the variation in IOD index during 1990-2020. The difference between sea surface temperature of two locations (or poles, hence a dipole) - a western pole in the Arabian sea (Western Indian Ocean) and an eastern pole south of Indonesia (Eastern Indian Ocean). The IOD has a great influence on the climate of Australia and other nations in the Indian ocean basin, and it is a substantial contributor to rainfall variability in this region. The IOD is a linked ocean-atmosphere phenomenon that occurs in the equatorial Indian Ocean, comparable to ENSO. The IOD is hypothesized to be linked to ENSO occurrences by an extension of the Walker circulation to the west and accompanying Indonesian throughflow (the flow of warm tropical ocean water from the Pacific into the Indian ocean). In the

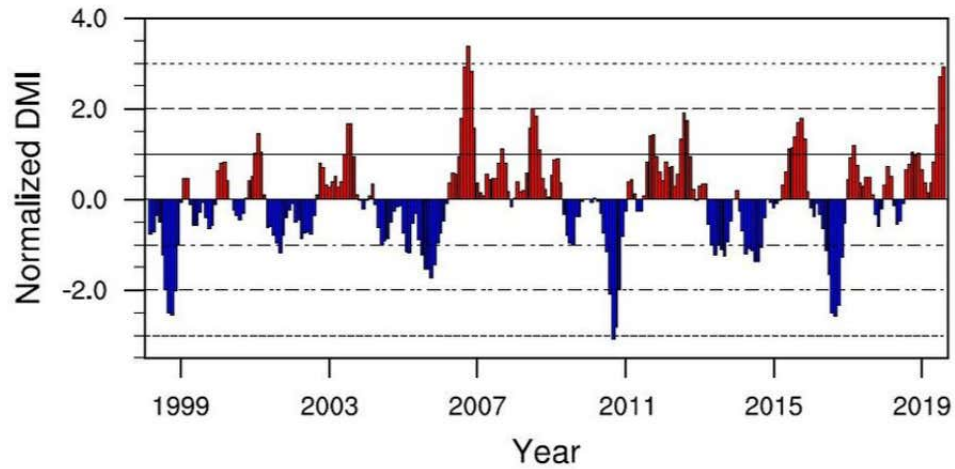


Figure 2.2: Fluctuation of Indian Ocean Dipole index over the period of 1990-2019  
[Source:<http://enformtk.u-aizu.ac.jp>]

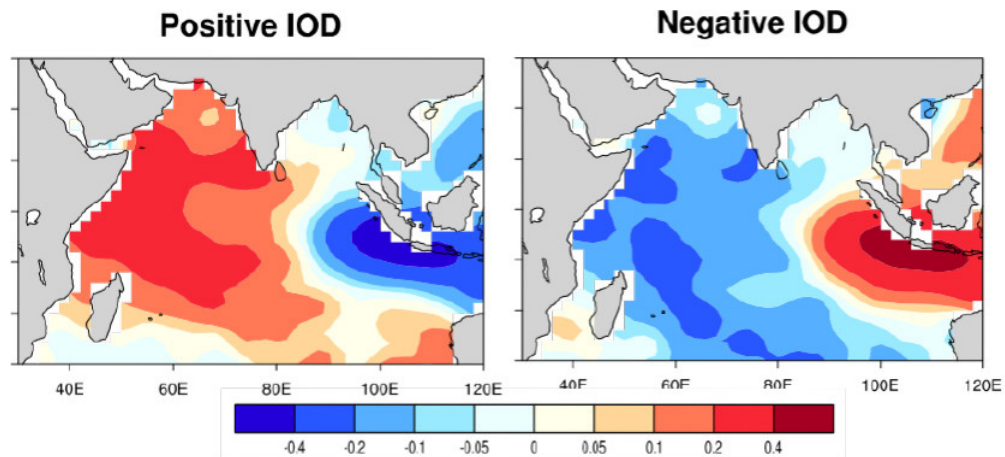


Figure 2.3: Sea surface temperature currents of Eastern and Western Indian Ocean mean value over [1981-2010] [Source:<https://fews.net/indian-ocean-dipole>]

2019 boreal autumn, a very positive IOD event occurred, causing significant climatic repercussions throughout the Indian Ocean basin. In this study, Lu and Ren (2020), G. Wang et al. (2020) the purpose is to determine the cause of the 2019 IOD event and the mechanisms that contribute to it. Since May 2019, we have seen a surprising strengthening of the Australian high and a reduction of sea level pressure over the South China Sea/Philippine Sea. Figure 2.3 visualize the warm (Positive phase) and colder (Negative phase) phase of Indian ocean sea surface temperature.

### 2.3.1 Positive Phase

Warmer sea surface temperatures in the western region of Indian ocean relative to easterly wind anomalies across the Indian ocean, as well as increased cloudiness in India's southwest, are expected during this phase more rain over southwest India and the top end.

### 2.3.2 Negative Phase

During this phase cooler sea surface temperatures in the western Indian ocean relative to the east winds become more westerly, bringing increased cloudiness to Australia's northwest. Less rainfall experienced over India. Lim and Hendon (2017) examine the causes and predictability of the strong negative IOD and its impact on the development of La Niña in 2016.

## 2.4 Literature Survey

Rajeevan et al. (1998) discovered that the combination of temperature fluctuation-patterns and deficient precipitation years had statistically significant variations. Temperature patterns over Northwest India between January and May have also been discovered to be positively associated with ISMR. These predictors were found to be great for LRF of ISMR.

Rajeevan et al. (1998) used June to Sept rainfall data which is an area-weighted sub-divisional rainfall aggregated from 36 meteorological stations and the Greenhouse Effect Detection Experiment (GEDEX) CD-ROM was used to collect data on global land air temperature. During the period 1901-1997, India has witnessed 9-excess and 11-deficit monsoon rainfall years. The temperature anomalies were sultry during Jan over northern Europe and North Asia. The temperature anomalies were cooler during Jan over central Asia. Sultry anomalies over NW India, Africa, and NW Europe, as well as dew anomalies over Central Asia, have been attributed to excess monsoon years. Temperature anomalies, during Jan, were found to be +ve correlated with NW India and NW Europe, and during May -ve correlated with central Asia.

Clark et al. (2000) analysed the relationship of SST and ISMR by regional SST and the effect of ENSO and decadal fluctuation on the stability of the relationship using the monthly grid of 40X40 and found a strong correlation of (0.53) summer precipitation and sea SST.

D. N. Kumar et al. (2007) studied effect of ENSO, EQUINOO, and ocean land temperature contrast were studied for Orissa regional rainfall. The ANN was combined with the GO optimizer to tackle the extremely non-linear and dynamic behavior of climate variables. The model shows a reasonable correlation between actual rainfall with actual regional rainfall.

Sahai et al. (2008) used an coupling of ML models to produce an effective probabilistic prediction of ISMR by using global sea surface temperature. The forecast system were more confident to predict monsoon. The AI technique GP is used to implement such complex behavior problems. The output of GP based prediction model was correlated with ISMR with 0.866 correlation coefficient.

Saha and Mitra (2019) conferred about the different predictors evolve in the process as time pass and climate change occurs. A new approach was used to detect new climate predictors and monsoon prediction using correlation of variables with ISMR. A newly developed model predicts ISMR with an error rate of 4.2%. Shahi et al. (2019) studies SST variations of various geographical regions and apply multiple regression (MLR) on these various predictors in different combinations to predict ISMR. Dutta et al. (2021) show that inlying fluctuation of Indian precipitaion considerably chaotic, is predictable as found to be slowly varying force (ENSO). They provide a new way to predict ISMR beyond the traditional estimates of potential predictability(PP). While climate change is a naturally occurring phenomenon, several research papers suggest that the rise in temperature in the twentieth century is connected to human developments. Loo et al. (2015) noticed that the % deviation from mean rainfall and cyclonic patterns observed globally are affected by climate change. Until the 1970s, a comparison of decadal variance in precipitation and temperature fluctuations showed general rises that were mainly variable. However, after the 1970s, global precipitation anomalies increased in lockstep with global temperature anomalies for the same time span. While precipitation is 70 percent below average, ISMR has undergone significant changes and has shifted westward. The distribution of ISMR is influenced by the AO, the Siberian High and western Pacific subtropical high, and the Tibetan Plateau.

Venugopal et al. (2018) studied the impact's of Indian Ocean heat content to assist the prediction system of ISMR and found OMT shows a better qualitative prediction capability as compared to the SST index. The prosperity tariff in prediction of



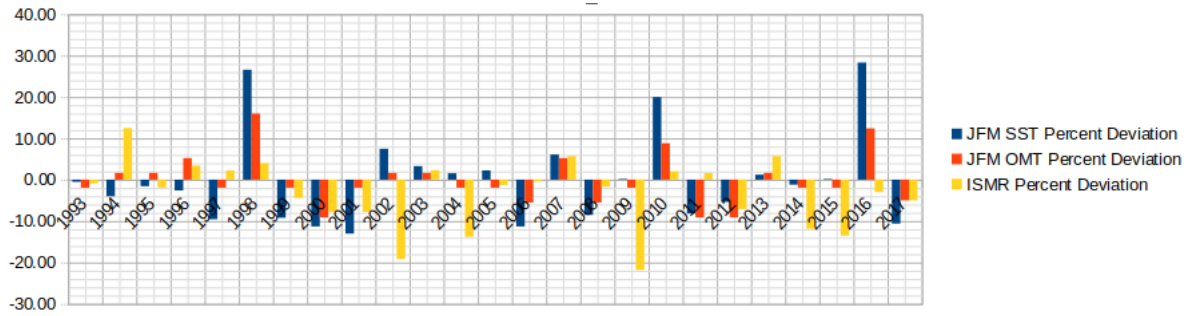


Figure 2.4: ISMR percent deviation (PD) from mean rainfall, compared with Jan-Mar SST's PD and OMT's PD in south West Indian Ocean region for the period [1993-1997]

above/below mean ISMR is 80% for ocean mean temperature compared to 60% for sea surface temperature. Climate indexes i.e. NINO 3.4, IOD, SOI are less predictive ability for ISMR. During the time, they used the spatial distribution of OHC D26° and the Jan-Mar climatic mean of SST and OMT of the Indian Ocean zone (1993-2017). comparative analysis of percentage deviation from mean of the observed values of ocean mean temperature, ISMR, and SST shown in Figure 2.4. OMT PD was not able to find out +ve/-ve ISMR only for five years (1995, 1997, 2001, 2011, and 2016); for SST there are 10 failures (1994, 1996, 1997, 2002, 2004, 2005, 2009, 2011, 2015, and 2016) out of 25 years.

The impact of Eurasian snow cover over monsoon precipitation always remains a matter of debate. Zhang et al. (2019) studied the well known reversionary kinship of central Eurasian snow cover and ISMR has disappeared since 1990 due to non-driving behavior of snow cover of the temperature of mid-tropospheric in summer, over Iranian Plateau region and neighborhood. The Eurasian snow cover may have lost its relationship with the Indian monsoon in a changing climate scenario. Vittal et al. (2020) found a relation between positive and negative Atlantic meridional mode (AMM) for Mar, Apr, and May with Indian summer rainfall. The positive (negative) phase of AMM weakening (strengthening) the summer monsoon precipitation overall India. They compared the skills of forecasting ISMR using AMM and found, new findings results was the better than North American Multimodel ensemble. A substantial calculation of the ENSO-ISMR links in a reconstructing climate is brittle significant for representation of the ISMR forecasting.

Pandey et al. (2020) found the short-period ENSO-ISMR links, shown by vital negative CC between the season (Jun, Jul, Aug, and Sep) and NINO 3.4 SST. ISMR does not affected plausibly with an augmentation in greenhouse gas (GHG) and simulated a thermal

environment, in order to forecast ISMR effectively, the global warming mode should be considered alongside the ENSO mode. Both short and long-period ENSO-ISMIR links are valuable, forecasting of ISMR must be recalculated taking into consideration the contributions from both. Although the ‘long-period’ relationship shifts from weak to substantially positive, GHG forcing occurs.

Dutta et al. (2021) tried to find out the correlation of cloud ice & sea ice with ISMR. The representation of ocean currents clouds, which involve ice phase cycles, in a coupled climate model greatly reduces ISMR variability in response to environmental predictor variables. The importance of cloud hydrometeors in ISM rainfall forecasting is shown by simulation findings using a coupled global climate model.

Zahan et al. (2021) studied that the oscillatory modes unbiased the long-term change in seasonal average precipitation over NEI derived from this nonlinear pattern, resulting in a climate sensitivity of  $3.2 \pm 1.65\%/K$ . However, even on multi-decadal and longer time scales, the NEIR appears to be out of step with that over the rest of India, which has huge ramifications for understanding past rainfall reconstructions from valleys in the north-east India.

Using 68 years (1951–2018) of gridded weather stations info, a detailed analysis of the Indian summer monsoonal rainfall (ISMIR) was conducted by V. Kumar et al. (2021), in the view of decadal shifts in continuous rainfall events and rainy-days. To find patterns over different climatic zones of India for the time sessions [1951–1984] and [1985–2018], Mann–Kendall’s methodology was applied to total rainfall amount, percentage of persistent rainfall events, and frequency of rainy days. In contrast to [1951–1984], they found a declining pattern for more than 4-days of regular rainfall events over the last 34 years [1985–2018]. Furthermore, since 1985, the rainfall had been moved to a lower number of continuous rainfall days with higher amount. Throughout the last 34 years, the total rainy-days during the crop’s sowing season has been decreased by a half day. Rainfall associates with soil moisture and evaporation up to 0.87 on a daily basis across India’s different land cover. Continuous rainfall tends to be a limiting factor in re-supplying moisture content at the top land surface levels.

Azhar et al. (2020) emphasised the relationship of the Antarctic sea-ice extent and the tropical-climate and diagnose the variability of sea-ice (20 deg- 90 deg East) extent inter-annual kinship with ISMR under the influence of the Mascarene high. The results were

amazing for the period 1979 – 2013, sea-ice extent of AMJ found to continue a valuable correlation with the indian summer monsoon precipitation over the Peninsular-Indian region. A combination of mean SLP, 850 hPa uwnd anomalies, and 500hPa geopotential-height shows a significant correlation between Mascarene high and sea-ice during high and low ice phases, unfolding that the enhancement (softening) of the Mascarene High, as well as an increase (decrease) in Indian summer monsoon rainfall, conform to the rising (steep) ice phase .

ISMR forecasting performed well in various seasonal forecasting systems, according to Jain et al. (2019) In forecasting Indian monsoon rains, the average of multi-model prediction results in different seasons using eight prediction methods produced statistically interesting results. Few models had a maximum correlation score of 0.6, which is close to the MME average, while others had very poor abilities. They looked at the effects of the geographically mean on ISMR forecasting results and found that as the spatial consistency of monsoon rainfall shifts improves, forecasts that are averaged over a wider region may provide better qualitiveness than control observations.

The climatic relationship between the atlantic zonal mode (AZM) and ISMR is particularly prominent in years when there are no ENSO events, which is why understanding the AZM and its relationship with ISMR is important for improving ISMR forecasting in coupled Ocean-Atmospheric models. Sabeerali et al. (2019) go through the results of hindcast simulations of 9-months of a coupled ocean-atmospheric model national centers for environmental prediction-coupled forecast system model version 2 (NCEP CFSv2) in evaluating the vaticinate ability of AZM. It was experienced by simulation that, for the February month vaticinate skill in CFSv2 hindcasts of the AZM was less and experience an unrealistic connection among ISMR. The combination of the ENSO generated ISMR from FebIC hindcasts and AZM generated ISMR from MayIC hindcasts can make considerable enhancement in the prediction ability of ISMR in CFSv2.

## 2.5 Different Monsoon Rainfall Forecasting Methods Used By Climate Researchers

Sahai et al. (2008) try to develop a Multi-model large ensemble forecasting system on the climate predictors like SST of different geographical locations for a long lead forecasting of the ISMR. The monthly SST data of years [1921-1998] was considered

Table 2.1: A List of climate predictor used by IMD in both forecasting phase (Apr Jun), with their Pearson correlation coefficient (PCC) [Source: <https://imdpune.gov.in>]

No.	Predictor	Correlation Coefficient	Prediction phase
1	Europe Land Surface Air Temperature Anomaly (Jan)	0.42	Apr
2	Eq. Pacific Warm Water Volume Anomaly (Feb+Mar)	-0.35	Apr
3	SST Gradient Between NW Pacific and NW Atlantic (Dec + Jan)	0.48	Apr & Jun
4	Eq. SE India Ocean SST (Feb)	0.51	Apr & Jun
5	East Asia MSLP (Feb + Mar)	0.51	Apr & Jun
6	NINO 3.4 SST ( MAM + (MAM - DJF) )	-0.45	Jun
7	North Atlantic Mean SLP (May)	-0.48	Jun
8	North Central Pacific Zonal Wind Gradient 850 hPA (May)	-0.57	Jun

as training period and years [1999-2007] as testing period. The RMSE & correlation experienced between predicted and actual amount of rainfall was 7.2% 0.62 respectively.

Genetic programming approach was used by Kashid and Maity (2012) on SST of region NINO 3.4 as ENSO index & EQWIN anomaly of Indian ocean as EQUINOO index. The model intends to medium range prediction of the ISMR for all India as a whole. Predictors data of years [1950-1975] was used as development period, years [1976-1990] as validation period, and years [1991-2010] testing period of the models. The RMSE & correlation witnessed between predicted and actual amount of rainfall was 7.79 & 0.70 respectively.

Regression Model was trained on SST, Eurasian Snow cover, NW Europe temperature, East Asia pressure, Wind Pattern. The new 8-parameter and 10-parameter were constructed over 38 years of data [1958-1995]. The data of the period[1996-2002] were used for the validation of models. The results of both of the new regression models was extremely better than that time existing 16-parameter model but none of these three were able to detect the major drought year 2002. Rajeevan et al. (2004) observe that 8-parameter LDA model showed 68% correct classification whereas the 10-parameter LDA gives 78% accurate results.

An experiment was implemented by Singh et al. (2018) on Hybridization of fuzzy and rough set techniques for patterns and behaviours using historical data of Indian summer monsoon rainfall. The developed method categories information in four

segregated zone which are as fuzzy positive zone, completely fuzzy zone, gray fuzzy zone, and fuzzy negative zone. Further, a substantial classification and decision-making perspective was developed, known as a four-way decision-making approach based on rough set and fuzzy set. The model was able to classified 100% correctly.

Artificial neural network was desingend by Singh and Borah (2013) on Indian summer monsoon rainfall. A feed-forward neural network with back-propagation was developed over the whole data [1871-1994]. Model has 25 units in the Input layer, first hidden layer have 2 neurons, second neural have 4 units, and output layer having 1 output unit. A sliding window of 5 years data of JJAS and seasonal mean rainfall was used as input to predict next year's ISMR. The rainfall data-set was divided into a training set [1871-1960] and a test set [1956-1994]. Good results were experienced.

Random vector functional link (RVFL), Single layer feed-forward neural network(SLFN), and Regularised Version of online sequential random vector functional-link (ROS-RVFL) techniques implemented by Dash et al. (2018) using the ISMR time-series data-set Seasonal (JJAS) rainfall used for the input to develop the model. 5-yr, 6-yr, 7-yr, 8-yr, 9-yr, and 10-year's sliding window was used to forecast the next( $t+1$ ) year. RVFL, SLFN and, ROS-RVFL network architecture were compared based on RMSE and Pearson's correlation and best results was 23.8949 & 0.997 using ROS-RVFL.

Charaniya and Dudul (2012) train a Neural network model with gamma memory on Weekly Indian ocean dipole indices(IOD). They attempt to predict ISMR by building a time lag neural network trained on one month's worth of IOD data. For the prediction of rainfall sequences with delay taps equal to 3 and gamma memory coefficient = 0.5, an FTLNN model with gamma memory and conjugate gradient back-propagation learning algorithm was created. The simulation covers the years 1995 to 2007. RMSE and PCC was 0.278 & 0.943.

Saha et al. (2021) develope ensemble regression model on the girded data-set of sea-level pressure, sea surface temperature, and U-WIND at a 200 hPa pressure level. An autoencoder is a model that can learn a non-linear function by iteratively teaching the model to obtain the data's complex features and using them to reduce the data's dimensionality. From 1948 to 2000, the preparation cycle was used, and the test time-series was used from 2001 to 2014. An ensemble regression tree with a bagging

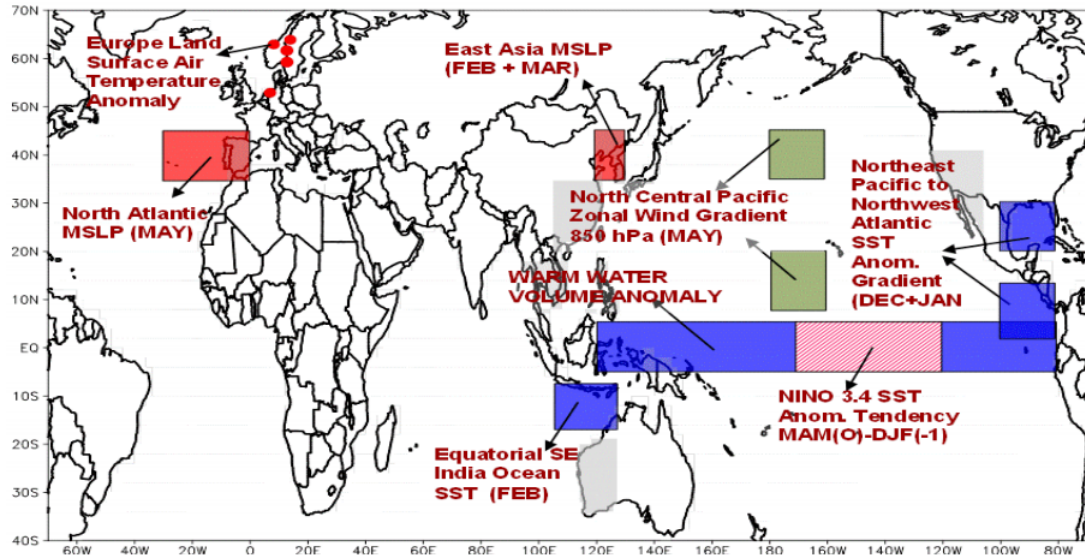


Figure 2.5: Geographical locations of 8 climate predictors used by IMD for LRF forecasting of ISMR [Source: <https://indpune.gov.in>]

algorithm was used to construct the prediction models on the predictors extracted from **auto-encoder trained** over data of different months. The best results found by researcher was **RMSE (5.0%) PCC of 0.82** .

## 2.6 How IMD Forecast Indian Summer Monsoon

Prediction of monsoon is done by IMD in different spatial and temporal scales. It varies from country as a whole to district wise in spatial scales and from a seasonal forecast to nowcast in temporal scale. Seasonal forecast for monsoon rainfall issued based on long range forecast (LRF) in the month of April for the entire country. This forecast update in the month of May for the entire country and also for its broad homogenous regions. LRF of monsoon will give a general picture of seasonal and monthly rainfall for the coming monsoon season.

Statistical models developed for LRF of southwest monsoon seasonal (June–September) rainfall. These models were developed to facilitate the IMD’s present two-stage operational forecast strategy. Models based on the ensemble multiple linear regression (EMR) and projection pursuit regression (PPR) techniques were developed to forecast the ISMR. According to this new strategy, IMD’s operational forecasts for the seasonal ISMR for the country as whole are issued in two stages. The first stage forecast is issued in mid April and an update or second stage forecast is issued by the end of June.

## 2.7 Research Gaps

Currently IMD is using statistical (Conventional) model Ensemble multiple linear regression [EMR] and Projection pursuit regression [PPR]. Statistical models does not have the ability to learn patterns among the data efficiently. Here we are trying to develop a new dynamic model using machine learning technique.

## 2.8 Chapter Summary

A literature study of different has been conducted to know more about how rain cycle starts, potential climate parameters influencing the rate of rainfall and distribution over the country, and the machine learning models used so far by different researchers over the long time span and the comparative analysis of their findings. The relationship of SST, SLP, Air temperature and zonal winds with the Indian summer monsoon rainfall.

# CHAPTER 3

## STATISTICAL ANALYSIS OF HISTORICAL RAINFALL'S DATA

### 3.1 Overview

Rain gauges are used by meteorologists and hydrologists to collect and calculate the amount of liquid precipitation. Plumviometer, udometer, and ombrometer are all names for rain gauges. These gadgets are used to determine the amount of precipitation. Currently, India has around 8,000 rain gauges. In India, two historical rainfall data sets are useful and widely utilized. The Indian Institute of Tropical Meteorology in Pune, performs the collection of 306 rain gauges the area-weighted average monsoon season precipitation for all of India is widely known, a trustworthy summer monsoon precipitation index in the Indian region ([www.tropmet.res.in](http://www.tropmet.res.in)). Since 1871, a long time series of this indicator has revealed several fascinating characteristics of the monsoon's inter-annual and decadal changes, as well as its regional and worldwide telecommunications. In this chapter, we are going to discuss the variation of the observed amount of rainfall in the summer monsoon over the 7 homogeneous regions, the 36 meteorological subdivisions, and the country as a whole.

### 3.2 All India Rainfall

All India summer monsoon rainfall is a area- weighted sum of observed rainfall over the country as a whole. Figure 3.2 shows the monthly, annually and monsoon rainfall observed in the period [1970-2020]. Figure 3.1 depicts the percentage deviation of summer monsoon rainfall from LPA (88CM) in period 1991-2019.

### 3.3 7 Homogeneous Regions Rainfall

In this section, we observed the 50 year running mean of 7 homogeneous regions of India which include northwest (Punjab, Gujarat, Rajasthan, Maharashtra and Goa), western Himalaya (Jammu Kashmir, Himachal Pradesh and Uttaranchal), north central (Haryana, Uttar Pradesh, Uttarakhand, Madhya Pradesh and Chhatisgarh), northeast (Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and



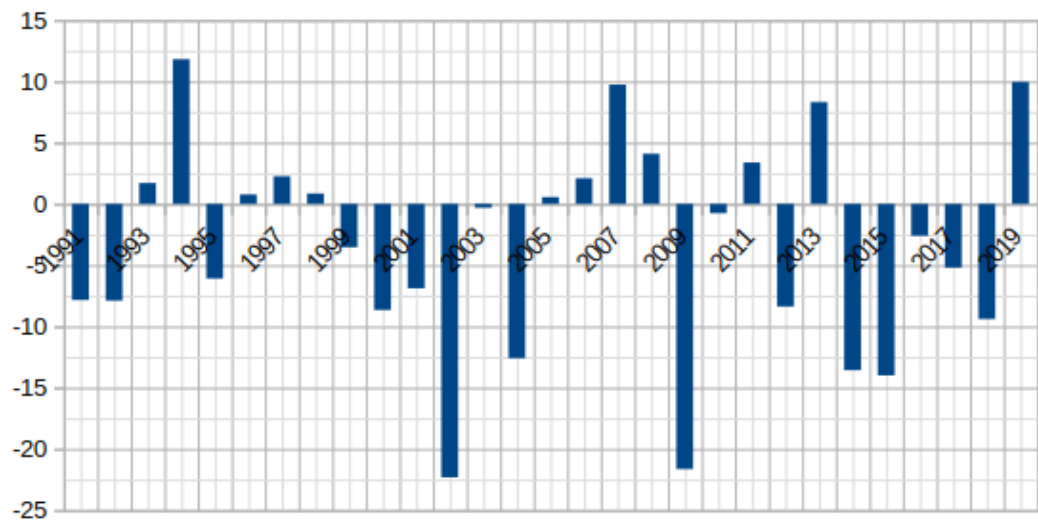


Figure 3.1: ISMR Percent Deviation from long term mean(88CM)

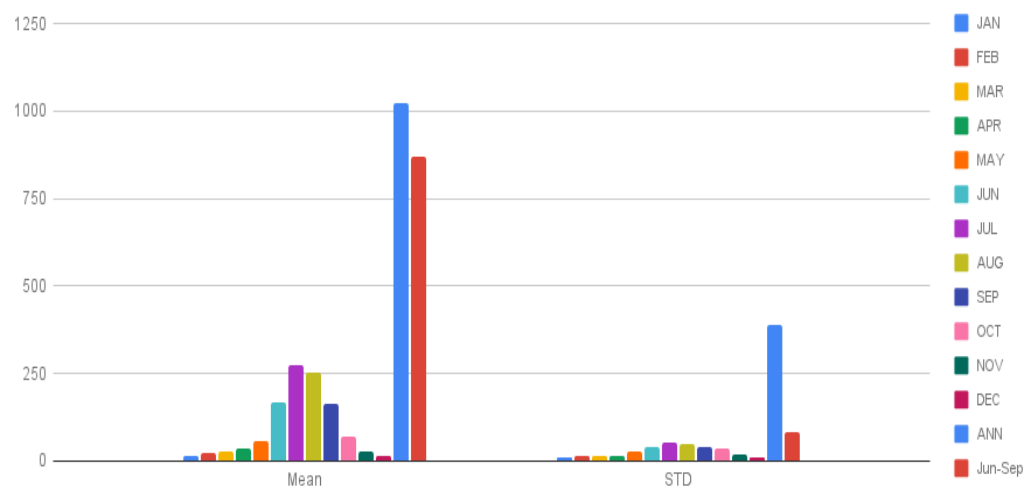


Figure 3.2: Long term Mean and Std of ISMR

Tripura), peninsula (Vidarbha, Chattisgrah, Marathwada, Central Maharastra, South interior Karnatka, Tamilnadu), east coast (Orisaa, Coastal Andhra Predesh, Pondicherry, Rayalaseema) and west coast (Coastal Karnatka, Lakshdweep, Konkan and Goa).

$$Running\ Mean(\mu) = \frac{\sum(X^{Y_m-50} + X^{Y_m-49} + .. + X^{Y_m})}{50} \quad (3.1)$$

$$Running\ STD(\sigma) = \frac{\sum(X_{\sigma}^{Y_m-50} + X_{\sigma}^{Y_m-49} + .. + X_{\sigma}^{Y_m})}{50} \quad (3.2)$$

where  $X^{Y_m}$  is the observed rainfall of month 'm' of year 'y'

Figure [3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9] shows the 50 year running mean and standard deviation of June, July, August & September months for the period 1970-2012 to analyse the fluctuations experienced in the Western Himalaya, North-West India, Central North-East India, East coast, North central, Interior peninsula, West Coast.

- In the time span form 1970-2012 Western Himalaya get slight increment in running mean for Jun and Aug months whereas a slight decrements in Jul and Sep months. Standard deviation remain consistent with a very small fluctuations.
- In case of Northwest India Jun & Aug month mean rainfall get increment, where July & Sep get a lower mean score and STD witnessed minor fluctuations.
- July & Sep months experienced a stagnant value of (around 405MM & 275MM respectively) of 50 year running mean and STD in central Northeast India region. Where the observed running mean rainfall in Jun & Aug month decreased.
- Jun and July months running mean rainfall increase with same pace but Aug & Sep months running mean rainfall got decreased and standard deviation of all four months experience tiny fluctuations in East coast Indian regions.
- The observed rainfall in North central India region does not show major fluctuations in 50 year running mean and standard deviation. More or less, it is consistent.
- Fig. 3.8 shows drastic changes of mean and std of monsoon rainfall in the period [1970-2012] of Indian interior peninsula.
- A small hike apparent in Jun & August month rainfall whereas a pull down in Jul & Sep month rainfall experienced in West coast Indian region.

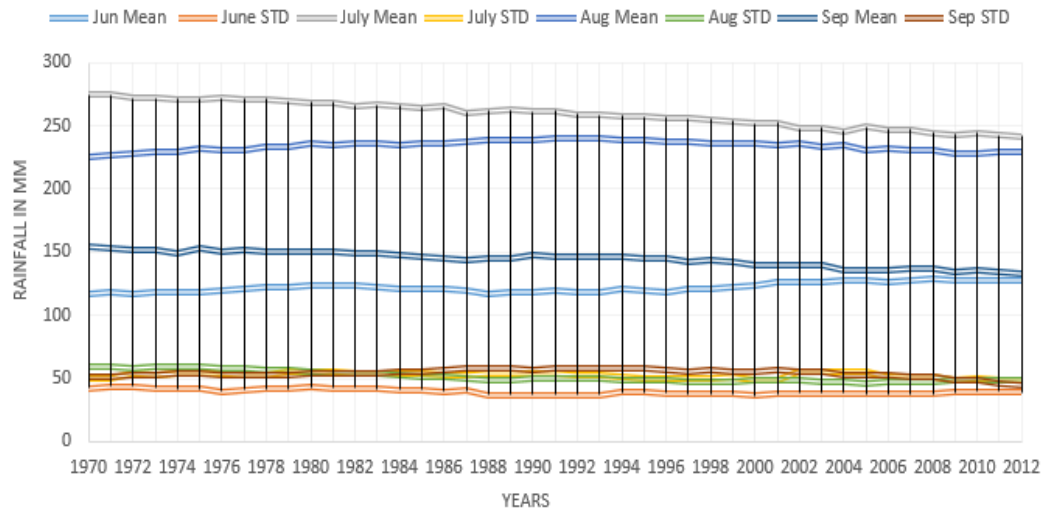


Figure 3.3: Western Himalaya

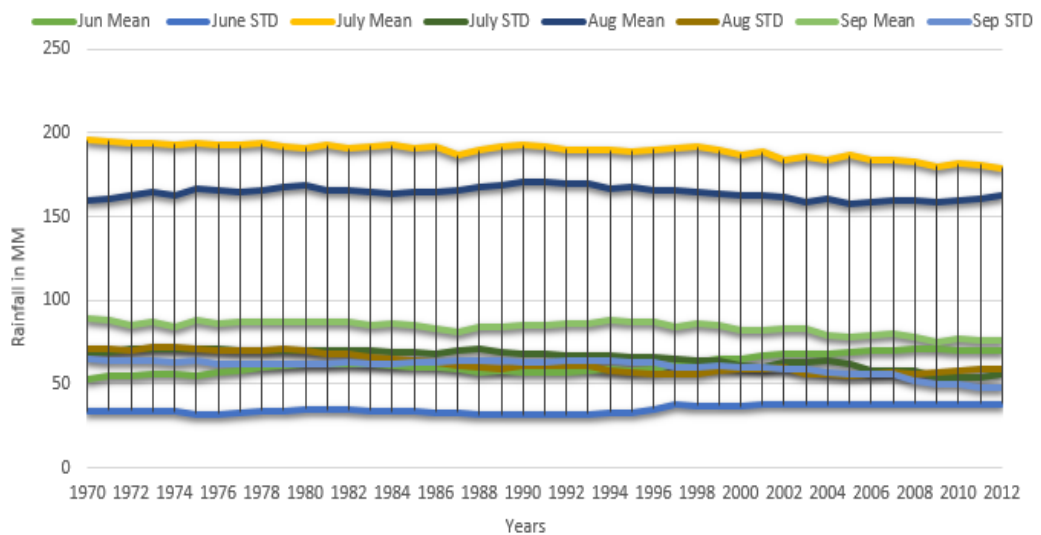


Figure 3.4: North-West India

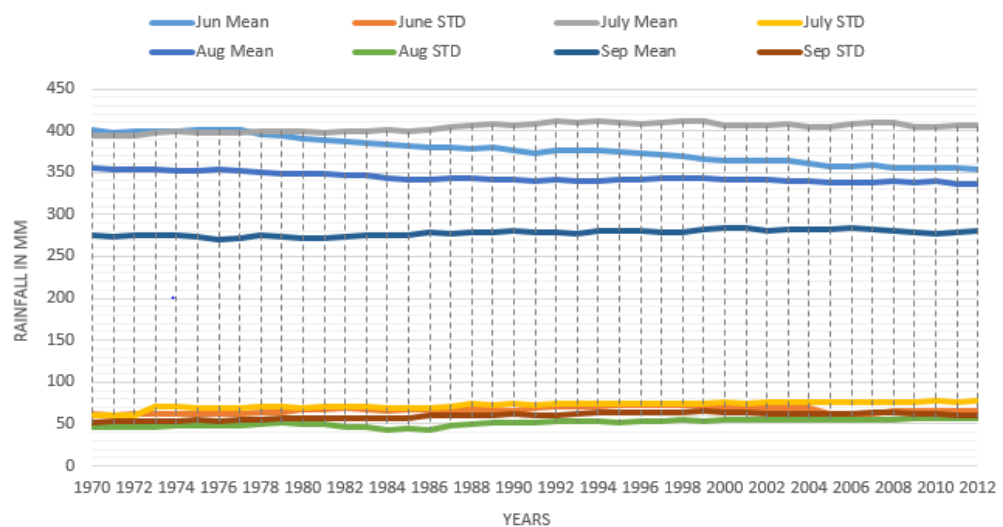


Figure 3.5: Central North-East India

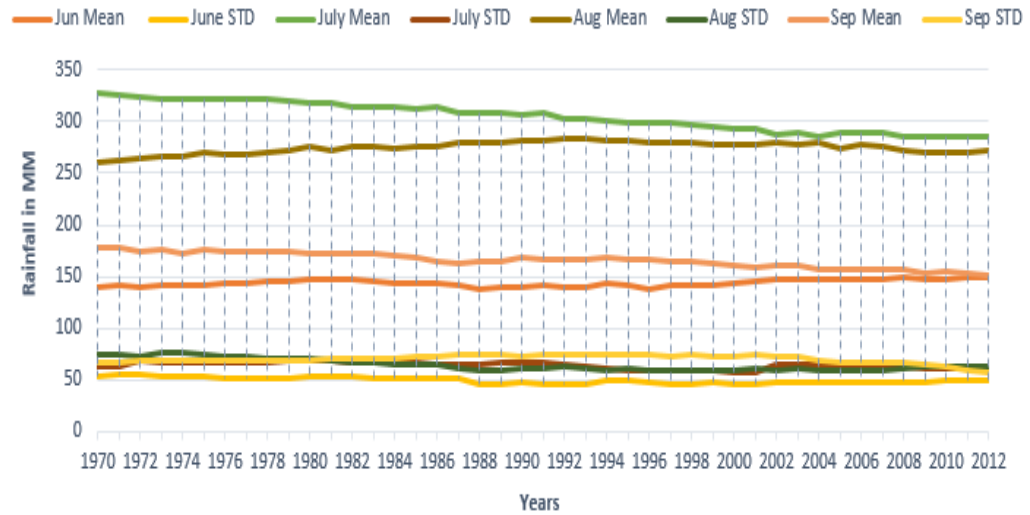


Figure 3.6: East Coast

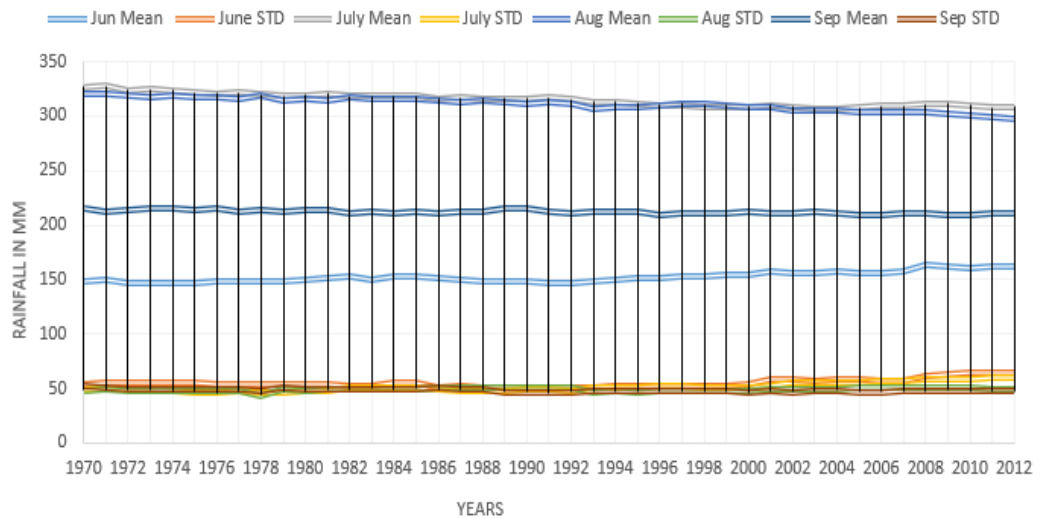


Figure 3.7: North-Central

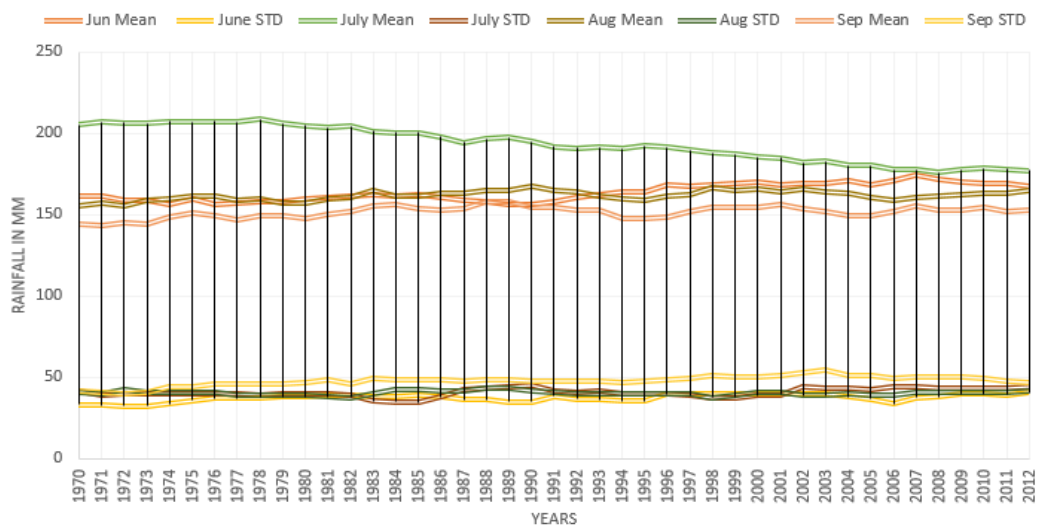


Figure 3.8: Peninsula

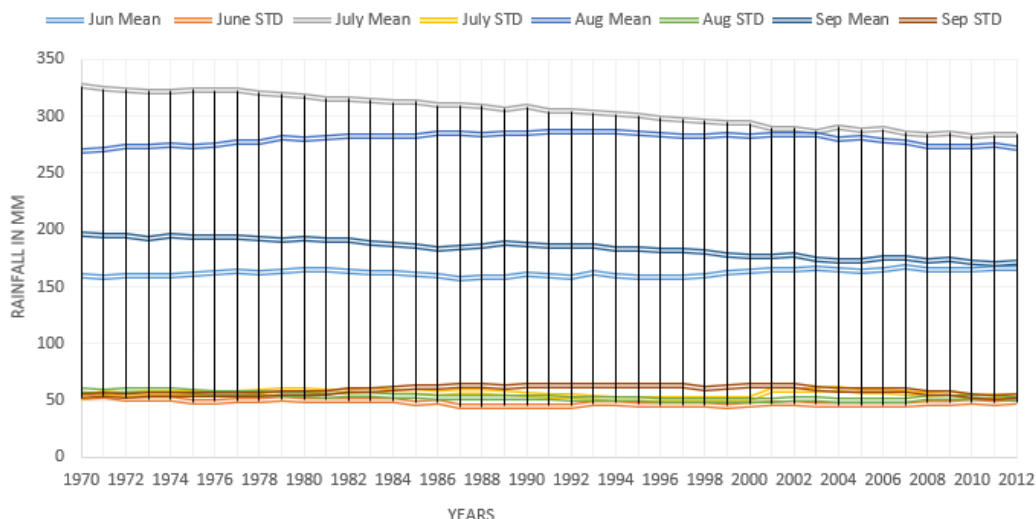


Figure 3.9: West Coast

### 3.4 Analysis of the Rainfall of Meteorological Subdivisions

Andaman Nicobar Islands, Naga Mani Mizo Tripura, Sub Himalayan West Bengal Sikkim, West Rajasthan, Vidarbha, Gangetic West Bengal, South Interior Karnataka, Orissa, Jharkhand, Bihar, East Uttar Pradesh, Konkan Goa, Madhya Maharashtra, West Uttar Pradesh, Uttarakhand, Haryana Delhi Chandigarh, Punjab, Himachal Pradesh, Jammu Kashmir, Assam Meghalaya, East Rajasthan, West Madhya Pradesh, East Madhya Pradesh, Gujarat Region, Saurashtra Kutch, Marathwada, Chhattisgarh, Coastal Andhra Pradesh, Telangana, Rayalseema, Tamil Nadu, Coastal Karnataka, North Interior Karnataka, Kerala and Lakshadweep are the 36 meteorological subdivisions of India. Table 3.1 shows the min, max, mean, the median and standard deviation of rainfall observed in these 36 subdivisions of June month. Table 3.2 shows the min, max, mean, median, and standard deviation of rainfall observed in these 36 subdivisions of July month. Table 3.3 shows the min, max, mean, median, and standard deviation of rainfall observed in these 36 subdivisions of August month. Table 3.4 shows the min, max, mean, median, and standard deviation of rainfall observed in these 36 subdivisions of June month. Table 3.5 shows the min, max, mean, median, and standard deviation of rainfall observed in these 36 subdivisions of [Jun-Sep] months.

Table 3.1: Statistical Analysis of the June month rainfall for the long period [1961-2010]

SUB-DIVISION	Min	Max	Mean	Median	STD
ANDAMAN & NICOBAR ISLANDS	148.80	706.30	407.52	408.05	117.72
ARUNACHAL PRADESH	239.40	1432.80	503.24	478.70	207.70
LAKSHADWEEP	125.60	604.30	331.45	320.25	105.47
KERALA	196.80	1096.10	608.11	597.80	185.91
SOUTH INTERIOR KARNATAKA	64.00	269.40	154.11	149.40	50.03
NORTH INTERIOR KARNATAKA	60.90	235.70	110.01	102.00	38.47
COASTAL KARNATAKA	405.40	1327.80	834.39	820.10	202.77
TAMIL NADU	21.20	128.10	52.53	49.85	22.18
RAYALSEEMA	24.00	270.70	70.02	57.65	44.36
TELANGANA	64.60	261.60	141.38	134.25	47.97
COASTAL ANDHRA PRADESH	52.00	300.10	121.79	116.05	49.49
CHHATTISGARH	45.30	413.40	194.76	178.50	84.18
VIDARBHA	64.40	338.60	169.39	156.25	68.46
MATATHWADA	55.80	250.10	136.00	123.15	54.24
MADHYA MAHARASHTRA	46.80	293.80	159.22	149.05	59.71
KONKAN & GOA	330.60	1011.90	695.59	686.70	169.24
SAURASHTRA & KUTCH	1.40	321.80	88.23	72.90	66.52
GUJARAT REGION	8.50	367.30	131.80	112.00	91.84
EAST MADHYA PRADESH	35.30	356.60	136.98	103.75	79.46
WEST MADHYA PRADESH	16.40	254.40	106.48	93.15	54.71
EAST RAJASTHAN	5.10	171.30	61.89	53.85	38.42
WEST RAJASTHAN	0.40	143.20	33.91	30.55	26.14
JAMMU & KASHMIR	24.80	182.00	73.17	67.25	32.79
HIMACHAL PRADESH	24.90	252.70	97.81	86.45	46.69
PUNJAB	1.60	155.50	51.64	42.85	32.42
HARYANA DELHI & CHANDIGARH	2.10	193.50	51.54	45.85	33.76
UTTARAKHAND	41.90	332.40	153.91	139.05	73.66
WEST UTTAR PRADESH	3.70	206.20	77.25	71.80	45.33
EAST UTTAR PRADESH	20.70	286.90	111.34	102.05	61.34
BIHAR	56.70	353.40	167.03	158.40	65.13
JHARKHAND	62.70	479.60	194.89	184.10	88.75
ORISSA	87.30	343.50	210.40	194.35	70.03
GANGETIC WEST BENGAL	69.70	597.10	255.28	255.20	100.13
SUB HIMALAYAN WEST BENGAL & SIKKIM	261.70	859.00	509.93	496.10	120.26
NAGA MANI MIZO TRIPURA	212.30	861.10	420.96	388.50	125.00
ASSAM & MEGHALAYA	273.10	780.50	488.92	498.50	105.66

Table 3.2: Statistical Analysis of the July month rainfall for the long period [1961-2010]

SUB-DIVISION	Min	Max	Mean	Median	STD
ANDAMAN & NICOBAR ISLANDS	45.30	918.50	402.39	391.90	141.96
ARUNACHAL PRADESH	233.00	1506.10	546.36	524.05	208.61
LAKSHADWEEP	29.40	491.60	302.06	304.00	112.47
KERALA	221.00	1308.90	677.70	674.45	215.57
SOUTH INTERIOR KARNATAKA	92.60	492.70	238.81	238.95	67.82
NORTH INTERIOR KARNATAKA	53.70	272.80	136.63	131.55	49.04
COASTAL KARNATAKA	545.70	1884.90	1140.90	1098.15	286.36
TAMIL NADU	11.80	174.70	76.40	74.15	33.19
RAYALSEEMA	28.80	245.20	100.92	91.00	48.27
TELANGANA	91.70	656.20	251.75	246.25	102.02
COASTAL ANDHRA PRADESH	74.70	329.60	177.30	169.70	54.97
CHHATTISGARH	138.40	630.60	383.60	373.55	90.18
VIDARBHA	88.10	508.20	307.68	296.95	88.97
MATATHWADA	24.30	413.50	178.03	169.30	76.88
MADHYA MAHARASHTRA	95.60	416.40	251.32	250.60	71.92
KONKAN & GOA	452.80	1543.30	1057.82	1072.80	242.96
SAURASHTRA & KUTCH	8.00	495.80	190.91	167.85	120.29
GUJARAT REGION	45.80	584.40	333.09	306.80	127.52
EAST MADHYA PRADESH	84.70	597.20	336.34	337.25	99.82
WEST MADHYA PRADESH	46.90	553.10	282.54	274.80	88.58
EAST RAJASTHAN	13.50	394.90	214.55	219.45	71.87
WEST RAJASTHAN	2.40	215.40	98.67	88.80	51.81
JAMMU & KASHMIR	67.20	440.30	183.39	161.90	89.56
HIMACHAL PRADESH	81.80	477.10	252.57	246.60	83.97
PUNJAB	44.10	359.30	173.84	164.45	72.97
HARYANA DELHI & CHANDIGARH	19.40	370.70	156.03	144.95	75.96
UTTARAKHAND	157.50	522.90	369.55	384.50	91.28
WEST UTTAR PRADESH	29.60	440.60	242.67	246.40	83.86
EAST UTTAR PRADESH	101.70	553.10	279.38	267.15	91.61
BIHAR	155.50	580.10	335.02	339.00	107.03
JHARKHAND	170.20	503.10	316.72	300.80	81.22
ORISSA	170.20	586.00	337.59	325.30	81.43
GANGETIC WEST BENGAL	168.70	610.10	335.19	302.40	92.78
SUB HIMALAYAN WEST BENGAL & SIKKIM	357.30	959.10	678.94	714.10	146.14
NAGA MANI MIZO TRIPURA	255.30	976.70	420.13	388.55	140.23
ASSAM & MEGHALAYA	308.60	995.20	522.79	493.65	138.60

Table 3.3: Statistical Analysis of the August month rainfall for the long period [1961-2010]

SUB-DIVISION	Min	Max	Mean	Median	STD
ANDAMAN & NICOBAR ISLANDS	125.80	666.50	398.44	382.15	112.27
ARUNACHAL PRADESH	172.40	1405.90	411.09	352.05	232.65
LAKSHADWEEP	84.10	437.90	226.01	215.05	85.56
KERALA	202.10	678.30	414.54	390.90	111.04
SOUTH INTERIOR KARNATAKA	91.00	305.40	192.93	188.65	52.74
NORTH INTERIOR KARNATAKA	29.40	290.10	126.09	125.70	48.33
COASTAL KARNATAKA	278.30	1579.10	795.28	743.70	245.47
TAMIL NADU	18.70	169.80	92.69	88.75	36.38
RAYALSEEMA	13.60	241.50	113.64	109.15	53.65
TELANGANA	52.30	471.10	245.63	236.15	96.36
COASTAL ANDHRA PRADESH	45.50	336.20	183.11	174.80	61.16
CHHATTISGARH	208.30	519.70	370.39	378.10	79.64
VIDARBHA	165.90	565.50	303.17	283.05	84.76
MATATHWADA	49.70	380.40	182.80	164.50	85.94
MADHYA MAHARASHTRA	82.70	421.70	206.07	202.15	67.41
KONKAN & GOA	224.10	1414.70	732.60	706.75	252.45
SAURASHTRA & KUTCH	4.00	634.00	134.84	113.10	113.51
GUJARAT REGION	55.10	624.80	279.70	247.30	134.67
EAST MADHYA PRADESH	162.60	590.30	364.18	357.40	107.92
WEST MADHYA PRADESH	130.50	536.00	305.02	313.30	92.52
EAST RAJASTHAN	70.90	380.90	214.84	224.40	78.80
WEST RAJASTHAN	6.80	311.30	84.81	72.50	55.72
JAMMU & KASHMIR	49.10	441.50	173.38	163.15	81.90
HIMACHAL PRADESH	89.10	476.10	239.06	237.20	78.06
PUNJAB	17.40	375.50	162.22	152.45	77.49
HARYANA DELHI & CHANDIGARH	32.10	405.30	165.66	163.45	82.92
UTTARAKHAND	73.30	536.10	355.73	351.30	96.68
WEST UTTAR PRADESH	34.60	509.60	253.55	242.25	104.90
EAST UTTAR PRADESH	41.50	449.70	255.48	241.05	84.53
BIHAR	154.90	539.80	278.16	280.10	82.88
JHARKHAND	152.80	526.10	300.44	287.90	75.81
ORISSA	187.30	624.90	350.19	348.45	86.39
GANGETIC WEST BENGAL	167.10	573.40	315.64	313.00	91.67
SUB HIMALAYAN WEST BENGAL & SIKKIM	263.80	990.50	517.58	488.55	174.00
NAGA MANI MIZO TRIPURA	241.60	1347.20	385.38	360.05	165.39
ASSAM & MEGHALAYA	209.70	773.40	396.40	387.60	103.70



Table 3.4: Statistical Analysis of the September month rainfall of various meteorological sub-divisions for the long period [1961-2010]

SUB-DIVISION	Min	Max	Mean	Median	STD
ANDAMAN & NICOBAR ISLANDS	215.90	775.10	416.08	414.45	111.46
ARUNACHAL PRADESH	152.50	646.60	338.45	303.70	122.82
LAKSHADWEEP	36.60	457.50	171.65	163.00	89.31
KERALA	48.50	526.70	255.38	220.35	129.99
SOUTH INTERIOR KARNATAKA	51.90	254.60	147.94	147.30	49.21
NORTH INTERIOR KARNATAKA	33.10	313.90	146.80	135.70	69.17
COASTAL KARNATAKA	79.50	702.80	304.98	267.30	147.55
TAMIL NADU	26.10	208.30	118.49	117.10	40.38
RAYALSEEMA	38.00	272.50	132.33	131.05	53.16
TELANGANA	52.40	378.20	166.23	146.15	82.64
COASTAL ANDHRA PRADESH	71.70	398.10	181.64	162.65	71.71
CHHATTISGARH	98.70	401.20	207.62	205.90	73.68
VIDARBHA	23.00	326.20	157.18	143.75	78.53
MATATHWADA	38.50	422.30	163.82	147.40	87.58
MADHYA MAHARASHTRA	54.80	324.80	158.97	161.35	67.39
KONKAN & GOA	65.00	678.80	353.38	333.30	158.49
SAURASHTRA & KUTCH	2.10	233.10	72.75	57.15	67.30
GUJARAT REGION	5.70	387.00	139.83	129.35	104.37
EAST MADHYA PRADESH	50.40	522.70	191.58	174.20	104.73
WEST MADHYA PRADESH	25.10	559.40	159.59	142.70	100.55
EAST RAJASTHAN	7.70	337.20	89.81	87.50	61.97
WEST RAJASTHAN	0.60	143.40	37.00	21.60	35.14
JAMMU & KASHMIR	13.50	308.60	83.23	66.25	58.46
HIMACHAL PRADESH	16.30	360.20	119.16	115.50	71.24
PUNJAB	0.50	434.10	80.13	57.45	76.06
HARYANA DELHI & CHANDIGARH	0.10	194.50	80.22	59.95	55.19
UTTARAKHAND	36.40	447.00	196.98	181.30	97.92
WEST UTTAR PRADESH	19.70	356.60	146.54	125.05	83.68
EAST UTTAR PRADESH	23.10	365.50	183.94	176.00	82.41
BIHAR	79.20	379.70	209.24	215.65	73.78
JHARKHAND	87.90	395.90	232.84	221.50	82.00
ORISSA	113.20	483.20	233.12	235.30	74.80
GANGETIC WEST BENGAL	101.10	591.80	275.37	251.40	106.05
SUB HIMALAYAN WEST BENGAL & SIKKIM	174.60	689.50	412.87	418.20	102.88
NAGA MANI MIZO TRIPURA	129.60	494.60	290.93	281.95	80.17
ASSAM & MEGHALAYA	137.40	494.70	297.00	291.20	91.64

Table 3.5: Statistical Analysis of the monsoon rainfall witnessed in various meteorological sub-divisions for the long period [1961-2010]

SUB-DIVISION	Min	Max	Mean	Median	STD
ANDAMAN & NICOBAR ISLANDS	1108.00	2191.50	1624.43	1620.40	230.92
ARUNACHAL PRADESH	1111.80	3896.10	1799.15	1634.60	582.32
LAKSHADWEEP	537.00	1753.00	1031.17	1010.95	211.65
KERALA	1297.10	3229.30	1955.72	1899.45	382.13
SOUTH INTERIOR KARNATAKA	454.50	996.20	733.81	740.80	119.11
NORTH INTERIOR KARNATAKA	329.70	852.70	519.54	498.45	115.38
COASTAL KARNATAKA	2205.00	4536.90	3075.55	2975.25	472.87
TAMIL NADU	94.20	481.20	340.10	349.10	80.50
RAYALSEEMA	268.80	736.80	416.91	399.25	109.64
TELANGANA	495.40	1447.20	804.99	805.50	192.61
COASTAL ANDHRA PRADESH	435.00	954.00	663.84	636.50	133.50
CHHATTISGARH	808.80	1779.50	1156.36	1131.85	198.67
VIDARBHA	643.00	1305.50	937.41	937.85	181.60
MATATHWADA	307.60	1128.60	660.64	626.75	167.92
MADHYA MAHARASHTRA	472.80	1260.40	775.58	764.60	144.61
KONKAN & GOA	1826.20	3698.60	2839.39	2863.35	406.73
SAURASHTRA & KUTCH	85.50	1043.70	486.74	470.85	215.24
GUJARAT REGION	329.30	1558.80	884.43	878.00	287.73
EAST MADHYA PRADESH	575.30	1477.50	1029.09	996.45	205.30
WEST MADHYA PRADESH	568.50	1362.50	853.64	825.30	167.90
EAST RAJASTHAN	256.80	919.50	581.09	563.85	140.01
WEST RAJASTHAN	67.20	522.20	254.39	239.65	100.90
JAMMU & KASHMIR	193.10	964.40	515.61	506.10	161.58
HIMACHAL PRADESH	369.60	1160.30	708.60	708.05	178.44
PUNJAB	181.50	1067.20	467.83	456.50	158.81
HARYANA DELHI & CHANDIGARH	178.70	776.70	453.45	451.60	139.25
UTTARAKHAND	529.50	1558.50	1076.18	1089.75	229.91
WEST UTTAR PRADESH	356.10	1127.90	720.01	721.05	177.56
EAST UTTAR PRADESH	434.30	1426.00	830.15	837.60	185.89
BIHAR	536.20	1515.10	989.44	988.55	181.60
JHARKHAND	578.40	1539.00	1044.88	1033.70	185.92
ORISSA	790.40	1542.10	1131.28	1136.05	171.04
GANGETIC WEST BENGAL	770.70	1624.20	1181.47	1186.10	200.09
SUB HIMALAYAN WEST BENGAL & SIKKIM	1443.60	2846.80	2119.32	2058.20	319.33
NAGA MANI MIZO TRIPURA	898.70	3050.20	1517.40	1418.35	371.88
ASSAM & MEGHALAYA	1208.10	2356.30	1705.10	1675.35	257.51

### 3.5 Chapter Summary

In this chapter, a monthly statistical analysis is performed, based on summer monsoon rainfall of 36 meteorological subdivisions, 7 homogeneous regions, and the country as a whole. ISMR is the area-weighted observed rainfall of the country as a whole. The various graph depicts the witnessed fluctuations of drought, excess, and normal-precipitation years of monsoon rainfall.

# CHAPTER 4

## DATA EXTRACTION & PREDICTION METHODOLOGY

### 4.1 Overview

Data is a critical component of any machine learning model and, in many ways, the sole basis for machine learning’s current popularity. Scalable machine learning algorithms have become viable as actual products that may add value to a business, rather than being a raw material for the production of its main activities, thanks to the availability of data. Machine learning is a branch of artificial intelligence that is at the core of the subject. Thus, without explicit programming, computers can enter a self-learning state. These kind of computers adapt, learn, grow, and evolve on their own when given fresh data. Following the input of these training and validation sets into the system, you can employ successive datasets to shape your machine learning model. The more relevant data you give the machine learning system, the faster and more accurate the model will learn and develop. We’ll talk about how to get data from climate predictors in this chapter. Also included is a brief description of the machine learning algorithms that were used to train and test the climate predictor models.

### 4.2 Data Extraction Sources

The southwest monsoon season (June–September) rainfall across India is calculated using the area-weighted mean of the seasonal precipitation rainfall data from all 36 meteorological subdivisions in India. The seasonal rainfall long period average (LPA) (1961–2010) is 87CM, with a coeff. variation of roughly 10%. For the period 1958 to 2018, we used the ISMR series. In India, two historical rainfall data sets are useful and widely utilized. The Indian Institute of Tropical Meteorology in Pune, performs the collection of 306 rain gauges the area-weighted average monsoon season precipitation for all of India is widely known, a trustworthy summer monsoon precipitation index in the Indian region ([www.tropmet.res.in](http://www.tropmet.res.in)). Since 1871, a long time series of this index has revealed a variety of fascinating features of the monsoon’s inter-annual and decadal fluctuations, as well as regional and worldwide connections. Parthasarathy et al. (1996)

utilised an ISMR series that differs from another commonly used series, which relied on a fixed 306 rain gauge stations spread over the plains of India.

Since **SST** has been one of the most researched factors in the ocean, scientists have paid close attention to it. We described the worldwide and seasonal trends of SST using a mix of ship SSTs, moored and drifting buoy SSTs, and satellite infrared SSTs. SSTs of the ocean's 10 m-thick skin layer can only be determined by satellite infrared SSTs, thus future research should focus on separating satellite skin SSTs from 1–5 m-deep, bulk-level measurements collected by buoys and ships. In simplest terms, the ocean's skin layer is the surface layer that connects a restless Ocean to an unstable atmosphere. There is a direct relationship between **wind speed and net air-sea heat transfer**. As a result, a better knowledge of this connection can aid in the resolution of the ocean-atmosphere net heat and momentum fluxes. Extended Reconstructed Sea Surface Temperature (ERSST) v5 (<https://www.ncdc.noaa.gov>) is used to derive North Atlantic Ocean SST (NAO SST), NINO 3.4, and Southeast Indian Ocean SST (SEIO SST). The Extended Reconstructed Sea Surface Temperature (ERSST) dataset is produced from the International Comprehensive Ocean–Atmosphere Dataset and is a worldwide monthly sea surface temperature dataset (ICOADS). The ERSST is produced on a  $2^\circ \times 2^\circ$  grid, with spatial completeness improved using statistical techniques. This monthly study, which runs from January 1854 to the present, contains anomalies calculated using a 1981–2010 monthly climatology.

**Sea level pressure** is abbreviated as **SLP**. The Physical Science Lab (PSL) keeps track of reanalysis datasets that might be used in climate diagnostics and attribution. To limit the effects of modeling changes on climate statistics, reanalysis datasets are generated by assimilating ("inputting") climatic observations using the same climate model throughout the reanalysis period. Ships, satellites, ground stations, RAOBS, and radar are among the various sources of observation. PSL currently provides reanalysis datasets available to the public in our standard netCDF format. This was NOAA's first reanalysis of this sort. NCEP utilized the same climate model, which was initialized with a range of weather observations, including ships, aircraft, RAOBS, station data, satellite observations, and more. Scientists can investigate climate/weather data and dynamic processes without the complications that model modifications might bring by utilizing the same model. Near-real-time observations are used to keep the dataset updated. We used the NCEP/NCAR

Reanalysis 1 Dataset (<https://psl.noaa.gov>) to calculate East Asia Sea Level Pressure and North Atlantic Surface Pressure for this study. For the period 1948/01 to the present, the NCEP/NCAR Reanalysis 1 Dataset provides monthly SLP means and other derived variables in daily and monthly values. Long-term monthly means (LTM) are calculated using data from 1981 to 2010. It is spread as a worldwide grid of 2.5 degrees latitude x 2.5 degrees longitude (144x73) [90N - 90S, 0E - 357.5E]. It has 17 distinct levels for measuring the data that has been observed.

NCEP/NCAR Reanalysis 1 Dataset additionally used to extracts Equator Winds and North Central Pacific Zonal Winds data. The Southern Oscillation Indicator (SOI) is a standardised index that measures variations in sea level pressure between Tahiti and Darwin, Australia. During El Nino and La Nina periods, the SOI is one indicator of large-scale air pressure changes between the western and eastern tropical Pacific (i.e., the state of the Southern Oscillation). In the eastern tropical Pacific, smoothed time series of the SOI is often well correlated with variations in ocean temperatures. When the SOI is in the negative phase, it means that the air pressure is below normal in Tahiti and above normal in Darwin. For extended periods, the SOI levels can be negative or positive, depending on whether El Nino or La Nina is occurring in the eastern tropical Pacific. The methods for calculating SOI is listed below. The Climate Prediction Center's SOI website has further information. The Arctic Oscillation (AO) is a climatic indicator that measures the status of the Arctic's atmospheric circulation. Two phases exist: a positive phase with geopotential heights below average (also known as negative geopotential height anomalies) and a negative phase with geopotential heights above average the reverse.

The Pacific Decadal Oscillation (PDO) is a long-term pattern of Pacific climatic fluctuation similar to El Nio. Extremes in the PDO pattern, like the more well-known El Nino/Southern Oscillation (ENSO), are defined by broad changes in the Pacific Basin and North American climate. Along with the ENSO event, ocean mean temperature (OMT) anomalies in the the tropical Pacific Ocean and northeast is used to describe the extreme stages of the PDO's temperature cycle as either warm or cold. The Pacific decadal oscillation has a positive tariff when SSTs are abnormally warm near the Pacific Coast, and chilly in the interior North Pacific, and when sea level pressures are below average across the North Pacific. When the climatic anomaly patterns are inverted, the PDO gets negative value, with warm SST anomalies in the interior and cold SST

anomalies near the North American coast, or when sea level pressures are above normal across the North Pacific.

#### 4.2.1 Anomaly Calculation

The climatological normal value is considered as a standard value and the relative deviation from the same is known as the anomaly. The long-term mean of an individual predictor is used as the base value. This standard (normal) climatology is usually computed by calculating a climatology over a time span of at least 30 years (the climate normal period). Following steps are involved in calculating data anomaly;

- Calculate averaged value of all grid points that comes under geographical location of particular predictor as observed data
- Calculate monthly climatology over a long period of [1981-2010]
- Calculate anomaly by taking the difference of observed data and monthly climatology.  $processDataym = Xym - \text{mean}(Xm)$ ,

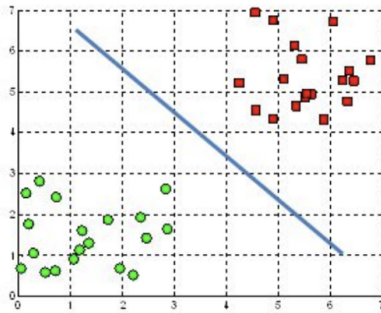
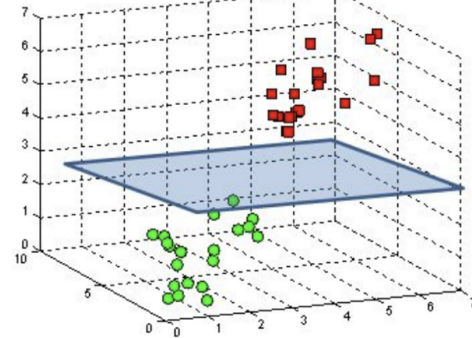
$$Anomaly \begin{pmatrix} month \\ year \end{pmatrix} = observed \begin{pmatrix} month \\ year \end{pmatrix} - LPA(month) \quad (4.1)$$

### 4.3 Machine Learning

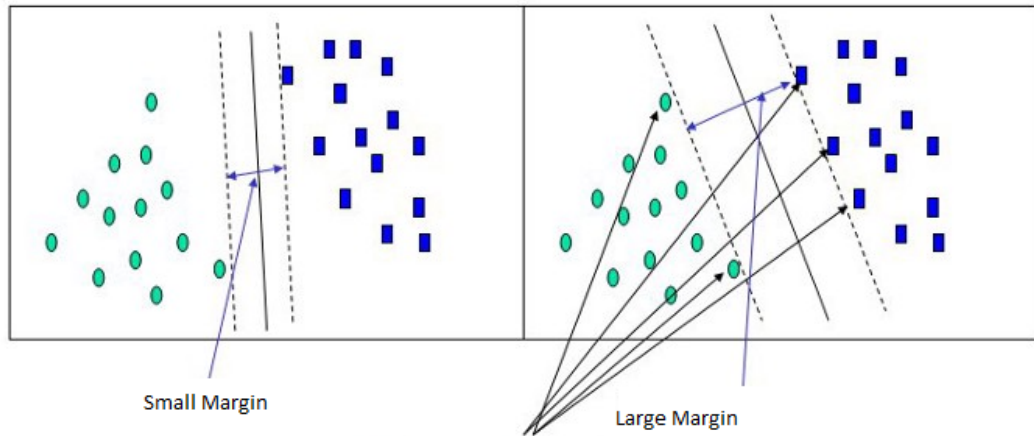
Machine learning is a data analysis technique that automates the implementation of analytical models. It is an artificial intelligence discipline predicated on the idea that computers can absorb from data, detect patterns, and make the fruitful decisions with hardly any human intervention. Supervised learning, reinforcement learning and unsupervised learning are the categories of machine learning methodology. The following are some supervised machine learning models used in this study:

#### 4.3.1 Support Vector Regressor

For classification and regression analysis in machine learning, Support Vector Machines (SVMs, also known as support vector networks) are supervised learning models with associated learning algorithms. It's a discriminative classifier with a separating hyper-plane as its formal definition, the Support Vector Machine (SVM). Also known as support vector machines, their objective is to find a point of intersection in N-dimensional space

A hyperplane in  $\mathbb{R}^2$  is a lineA hyperplane in  $\mathbb{R}^3$  is a plane

(a) SVM Plane



(b) SVM Margin

Figure 4.1: Showing SVM plane and SVM Margin

( $N$  = the number of features) that may be used to differentiate data points. To separate the two types of data points, you can pick from a number of hyper-planes. Our aim is to find the plane with the largest margin, or distance between data points from both classes. Maximizing the margin distance provides some reinforcement, making future data points simpler to categories. Hyper-planes are decision boundaries that help categories data. On each side of the hyper-plane, several categories can be applied to data points.

#### 4.3.2 Multiple Linear Regression

MLR is the most trivial and often utilized predictive analysis type. It's traditional a statistical way for developing a model on dependent variable's connection with a group of independent variables. The observed data is to be fitted in linear equation, Multiple Linear Regression aims to describe the connection between different predictors and a



target variable. Multiple linear regression is essentially identical to basic linear regression in terms of stages. The distinction is in the assessment. It may be used to determine which factor has the most influence on the projected output, and it can also be used to determine how different variables relate to one another.

$$Y = w_0 + w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \dots + w_n * x_n \quad (4.2)$$

$Y$  = Dependent variable and  $x_1, x_2, x_3, \dots, x_n$  = multiple independent variables

### 4.3.3 Ridge Regression

Ridge regression is a vastly used ML model developing method that can be utilized to observe data-set with multicollinearity. L2 regularisation is achieved using This approach is used to achieve L2 regularisation. When there is a problem with multicollinearity, least-squares are unbiased, and variances are significant, the projected values are distant from the actual values. The value function for ridge regression:

$$Min(||Y - X(\theta)||^2 + ||\theta||^2) \quad (4.3)$$

Lambda is known as punishment term. An alpha parameter in the ridge function denonte given here. So, by modifying the values of alpha, we are commanding  $\lambda$ . Lower the significance of alpha, lower is the penalty and therefore the magnitude of coefficients is increased. Whereas higher alpha value corresponds to higher penalty, as a result the value of coefficient reduce, it shrinks the parameters. As a result, it's used to avoid multicollinearity. It pulls down the model complexity by coefficient shrinkage.

### 4.3.4 Stacked Auto-encoder

An autoencoder is a type of artificial neural network that is used to develop unsupervised machine learning model. It has three layers: input, internal, and output. This architecture was used to get the data's complex characteristics as well as to reduce the data's dimensionality. The autoencoder tends to reproduce the feeded input information from the learnt representation in the internal layer by using the output data-set to the same as the input data-set. The architecture of auro-encoder is enough skillful of recognise a non-linear function using an iterative model training procedure. The main goal of the model is to reduce the re-construction faults in rebuilding the output from the data representation witnessed in the internal layer. A collection of

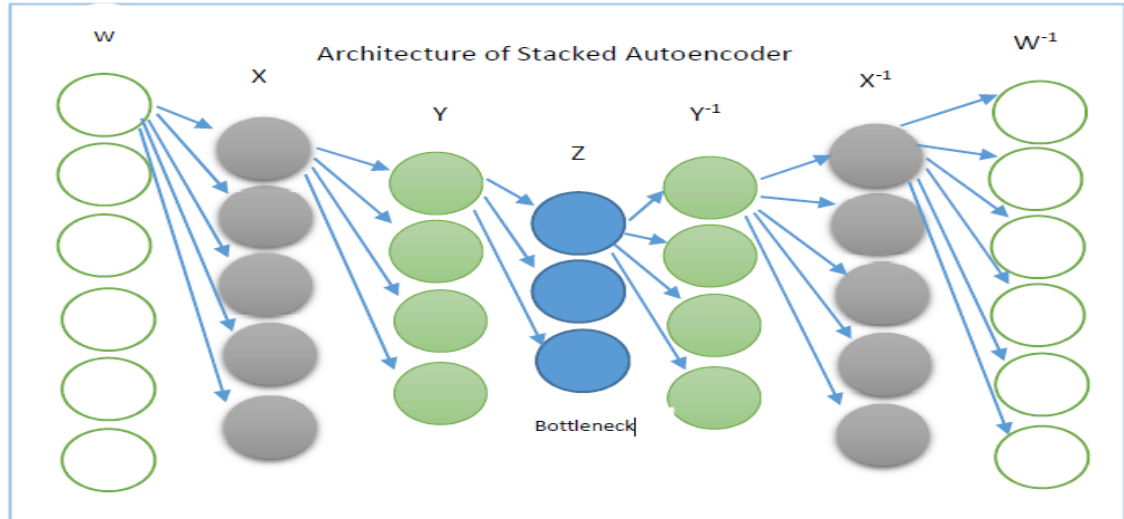


Figure 4.2: Architecture of Auto-encoder used to recreate the input vector

single-layered auto-encoders are stacked to form a deep neural network model. The results of the first auto-encoder forwarded as an input to the second network model, and this continues to the required depth of the network architecture. The training of the stacked auto-encoder is kind of unsupervised learning, which no longer expect for any deep information about the target variable. Each layer training is performed as a single layer known as pre-training. The individual layers of the heaped auto-encoder are edified with the main motivation of reduction of errors in input data-set reformation individually for the layers. Figure 4.2 shows the layered architecture of auto-encoder used to recreate the input matrix.

## 4.4 Rainfall Forecasting Approach

Indian summer monsoon rainfall forecasting has always being a tough task, IMD always try to use most significant climate predictors, best suited training time and efficient ML model. In this section, we will discuss two forecasting Approaches implemented.

### 4.4.1 Approach - I

These are the important steps followed to develop machine learning model and forecasting rainfall:

- The prediction performance of each of the potential models during a set common period 1965–2000 was investigated to determine the appropriate duration of training period. We utilised the **sliding model training period** approach for all available training period lengths. For example, in the instance of a 25-year training period,

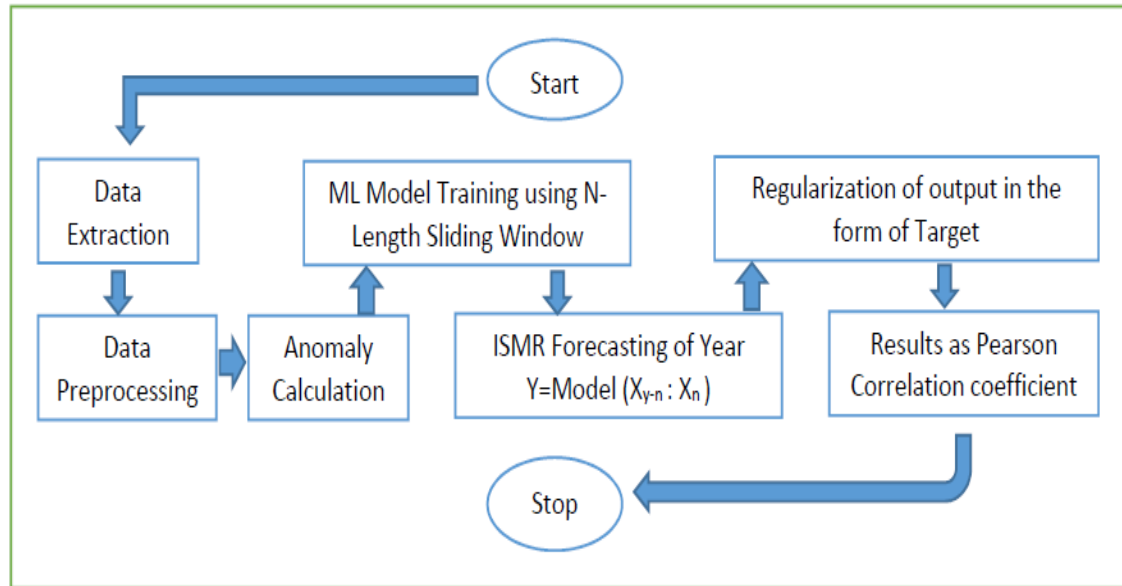


Figure 4.3: Steps involved in training Simple regression model using 10 predictors

data from 1976–2000 were used to forecast rainfall in 2001, and data from 1977–2001 were used to predict rainfall in 2002, and so on.

- We train SVR, MLR and ridge regression ML models using sliding model training period to predict the rainfall of the next year.
- As indicators of predictive skill, we utilised the root mean square error (RMSE) and PCC of forecasts from 2001 to 2018. For each of the ML models, RMSE and PCC values were calculated for various training period durations (Window size). The ideal length of training period for each model is defined as the length of training period with the lowest RMSE of forecasts for the period 2001-2018.

#### 4.4.2 Approach - II

The suggested predictor identification method begins with pre-processing of input climate data. These are the important steps for followed to develop machine learning model using second approach:

- We consider **sea-level pressure** (SLP) for our research work. In pre-elusive, the globe is spilt-up into spatial-rectangular boxes of measurement 20 longitude 10 latitude, in total which counts up-to **324 rectangular** grid boxes  $[(360/20) \times (180/10)]$ .
- All of the time-series values of climatic variables within the grid's spatial cover are averaged to produce a single time series of climatic variable for that specific grid.

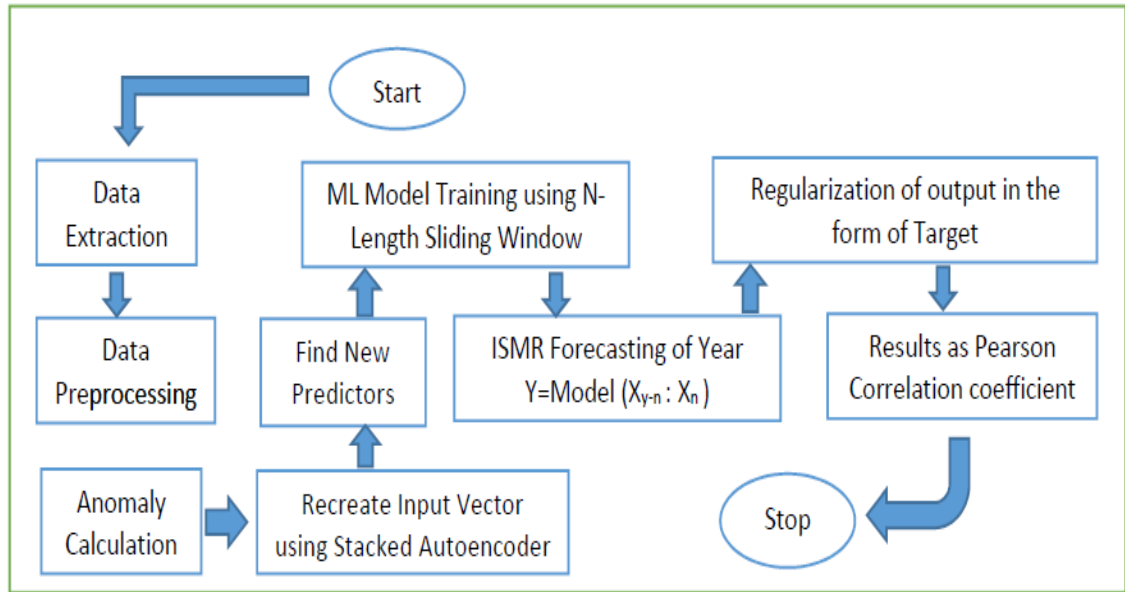


Figure 4.4: Steps involved in training Simple regression model using 10 predictors

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 324)]	0
dense_5 (Dense)	(None, 324)	105300
re_lu_3 (ReLU)	(None, 324)	0
dense_6 (Dense)	(None, 97)	31525
re_lu_4 (ReLU)	(None, 97)	0
dense_7 (Dense)	(None, 29)	2842
re_lu_5 (ReLU)	(None, 29)	0
dense_8 (Dense)	(None, 9)	270
dense_9 (Dense)	(None, 29)	290
re_lu_6 (ReLU)	(None, 29)	0
dense_10 (Dense)	(None, 97)	2910
re_lu_7 (ReLU)	(None, 97)	0
dense_11 (Dense)	(None, 324)	31752
Total params: 174,889		

Figure 4.5: Layered architecture of auto-encoder model

- Each such smoothed sequence of the chosen grids is regarded as an input neuron of the auto-encoder.
- Figure 4.5 shows the layered architecture [324,97,29,9,29,97,324] of auto-encoder model which learns from the monthly SLP dataset [1948-2000]. And using the trained model, we recreate monthly [Jan, Feb, March, Apr, and May] SLP dataset from first layer [X] of auto-encoder for the entire data period [1948-2020].
- We extract top most 8 predictors for each months [Jan, Feb, March, Apr, and May] from the recreated data-set for the period [1948-2020]. The significance of the new predictors measured by PCC between the ISMR and recreated dataset of the period 1981-2010.
- We train SVR regression ML models using sliding model training period to predict the rainfall of the next year using all combinations of monthly data (i.e  $2^5$  ways).
- As indicators of predictive skill, we utilised the root mean square error (RMSE) and PCC of projections from 2001 to 2020. For the ML model, RMSE & PCC values were calculated for different size of learning period (Window size). The ideal length of training period for each model is defined as the length of training period with the lowest RMSE of forecasts for the period 2001-2020.

## 4.5 Chapter Summary

In this chapter, we discuss how data extraction is carried out, how pre-processing of the extracted data is implemented, how data anomaly is calculated, different machine learning models (like auto-encoder, support vector regressor, ridge regression, and multiple linear regression), and model training approaches using a sliding period window model.

# CHAPTER 5

## FORECASTING RESULTS OF DIFFERENT MODELS AND THEIR COMPARATIVE ANALYSIS

### 5.1 Overview

Because India is an agricultural country, the success or failure of agricultural products and water scarcity in each year is always a big worry. ISMR is the monthly area-weighted average of Indian summer monsoon rainfall, which may be used to forecast the whole monsoon season across the nation. Forecasting ISMR is always being a grate challenge for the researcher and also for the IMD ministry of earth sciences (Government of India). Here, we tried a few new ways that can enhance the accuracy of forecasting ISMR. In this chapter, we will depicts the findings and machine learning model results of our study in the form of tables and charts.

### 5.2 Linkage Between Predictors and the ISMR Monthly Data

We examine the relationship of ISMR with all the potential climate predictors monthly data, some shows a good correlation with rainfall and some found to be neutral. We extract these 10 climate parameters which are significantly linked with the ISMR data. Figure 5.1 shows the currents in the measured 30 year running PCC of all the 10 climate predictors for the period of 1980-2018.

Indian is divided into 7 homogeneous regions (like Peninsula, North East, North-West, West-coastal, North-central, East-Coast, and Western Himalaya,) based upon their climatology of the respective region. Here, we try to determine the influencing factor of these 10 climate parameters on Indian homogeneous region. Table 5.1 is the correlation matrix of Indian regions rainfall (like R1:Peninsula, R2:North East, R3:North-West, R4:West-coastal, R5: North-central, R6: East-Coast, R7:Western Himalaya, and Area-weighted rainfall over the country as a whole) with all the 10 climate predictors for the period 1981-2010.

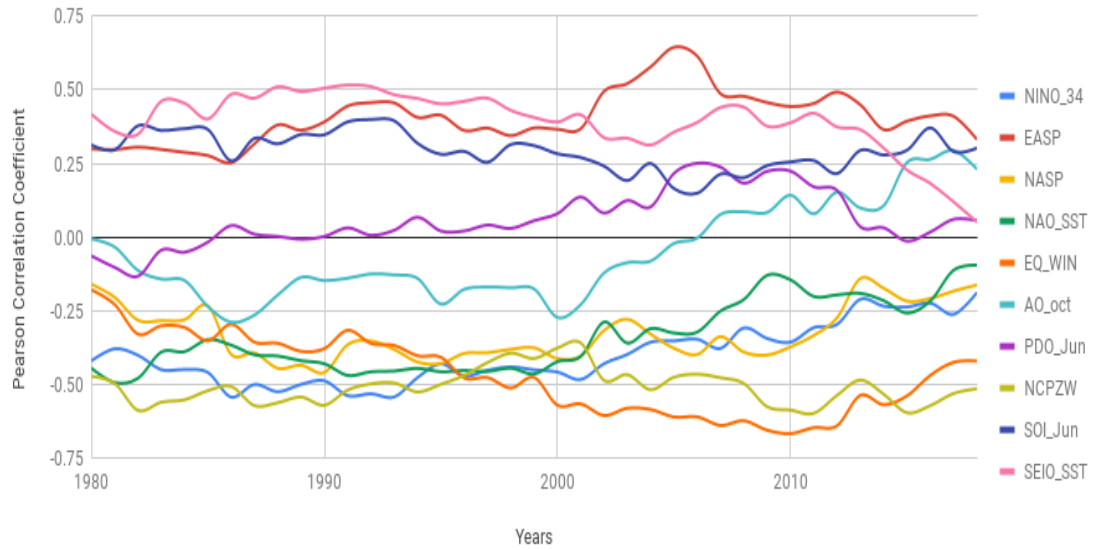


Figure 5.1: 30 Years running PCC between ISMR and climate predictors for the period [1980-2018]

Table 5.1: Correlation Matrix of Indian Regions Rainfall and Climate predictors. R1:Peninsula, R2:North East, R3:North-West, R4:West-coastal, R5: North-central, R6: East-Coast, R7:Western Himalaya

Correlation Table (1981-2010)								
Region	ISMR	R1	R2	R3	R4	R5	R6	R7
NINO_34	-0.36	-0.4	-0.08	-0.32	0.18	-0.34	-0.29	-0.35
EASP	0.44	0.21	-0.02	0.51	0.23	0.36	0.46	0.48
NASP	-0.37	-0.03	-0.2	-0.36	-0.33	-0.43	-0.43	-0.42
NAO SST	-0.14	-0.16	0.42	-0.13	0.11	-0.35	-0.21	-0.24
EQ Wind	-0.67	-0.28	-0.13	-0.44	-0.74	-0.52	-0.45	-0.51
AO	0.14	0.16	0.33	0.04	0.01	-0.04	0.01	0.01
PDO	0.22	0.27	0.17	0.18	-0.12	0.37	0.31	0.28
NCPZ Wind	-0.59	-0.33	-0.04	-0.56	-0.39	-0.42	-0.48	-0.53
SOI	0.25	0.44	-0.25	0.25	0.09	0.18	0.14	0.24
SEIO SST	0.39	0.33	-0.07	0.44	0.04	0.34	0.36	0.43

Table 5.2: List of predictors used for Approach-1

No.	Predictors	Period	Geographical Location
P1	NINO_34	May	[5N-5S, 170W-120W]
P2	North Atlantic Surface Pressure	May	[35N-45N, 30W-10W]
P3	North Atlantic Ocean SST	Dec(-1) + Jan	[20N-30N, 100W-80W]
P4	EQ Wind	May	[5N-15N, 165W-110W]
P5	East Asia Surface Pressure	Feb + Mar	[35N-45N, 120E-130E]
P6	Arctic Oscillation	Oct	[5N-90N]
P7	Pacific Decadal Oscillation	Jun	[15N-20N]
P8	North Central Pacific Zonal Wind	May	5N-15N, 180E-150W
P9	Southern Oscillation Index (SOI)	Dec (-1)	mean(Thaiti and Darwin)
P10	SE Indian Ocean SST	Feb + Mar	5N-15N, 180E-150W

### 5.3 Approach-I Results

We use 10 climate predictor to forecast the rainfall. The list of parameters along with their geographical is discussed in the table 2.1. There are some findings of this work.

- Figures [5.2, 5.3, 5.4] show the prediction skill of SVR, MLR and Ridge regression models developed based on sliding window prototype of length 29. We forecast the ISMR (Country as a whole) for the period of 2001-2018 and got a PCC of 0.82 using SVR model.
- Figures [5.5, 5.6, 5.7] show the prediction skill of SVR, MLR and Ridge regression models, in order to forecast North central rainfall, developed based on sliding window prototype of length 21. We forecast the ISMR for the period of 2001-2012 and got a best PCC of 0.83 using MLR.
- Figures [5.8, 5.9, 5.10] show the prediction skill of SVR, MLR and Ridge regression models, in order to forecast East coastal region rainfall, developed based on sliding window prototype of length 12. We forecast the ISMR for the period of 2001-2012 and got a best PCC of 0.62 using Ridge regression.
- Figures [5.11, 5.12, 5.13] show the prediction skill of SVR, MLR and Ridge regression models, in order to forecast Western Himalayan region rainfall, developed based on sliding window prototype of length 12. We forecast the ISMR for the period of 2001-2012 and got a best PCC of 0.73 using Ridge regression.
- Figures [5.14, 5.15, 5.16] show the prediction skill of SVR, MLR and Ridge regression models, in order to forecast North-east region rainfall, developed based on sliding



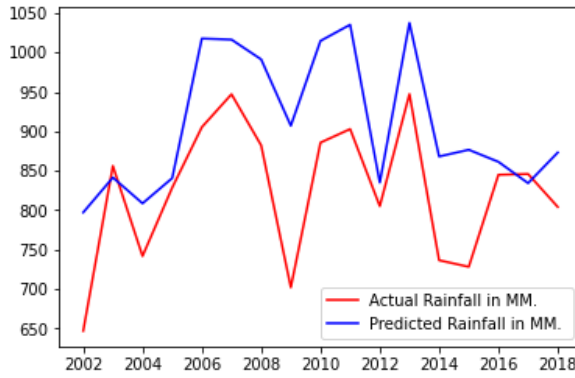


Figure 5.2: ISMR forecasting using SVR

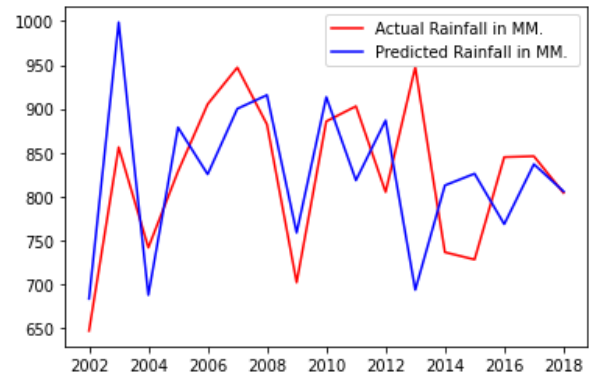


Figure 5.3: ISMR forecasting using MLR

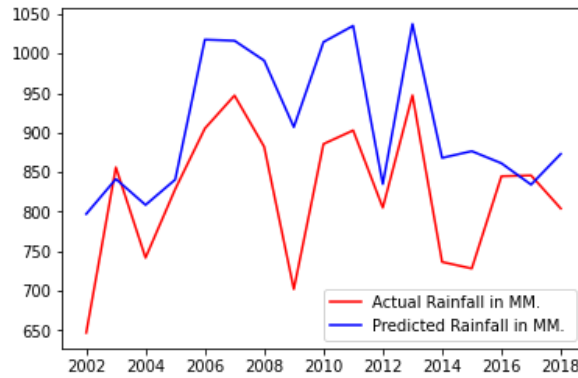


Figure 5.4: ISMR forecasting using Ridge Regression

window prototype of length 28. We forecast the ISMR for the period of 2001-2012 and got a best PCC of 0.65 using SVR .

- Figures [5.17, 5.18, 5.19] show the prediction skill of SVR, MLR and Ridge regression models, in order to forecast North-west region rainfall, developed based on sliding window prototype of length 25. We forecast the ISMR for the period of 2001-2012 and got a best PCC of 0.67 using MLR model.
- Figures [5.20, 5.21, 5.22] show the prediction skill of SVR, MLR and Ridge regression models, in order to forecast Peninsula region rainfall, developed based on sliding window prototype of length 26. We forecast the ISMR for the period of 2001-2012 and got a best PCC of 0.74 .
- Figures [5.23, 5.24, 5.25] show the prediction skill of SVR, MLR and Ridge regression models, in order to forecast East coastal region rainfall, developed based on sliding window prototype of length 34. We forecast the ISMR for the period of 2001-2012 and got a best PCC of 0.70 .

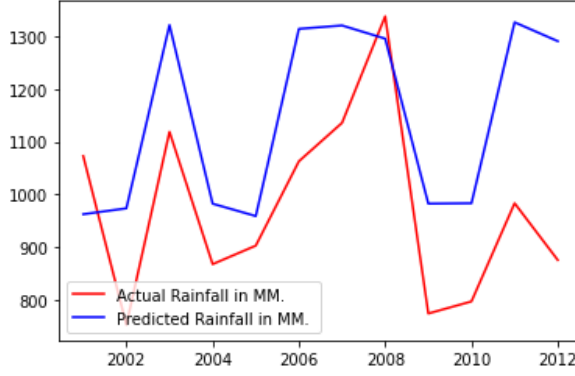


Figure 5.5: North-Central Indian rainfall forecasting using SVR

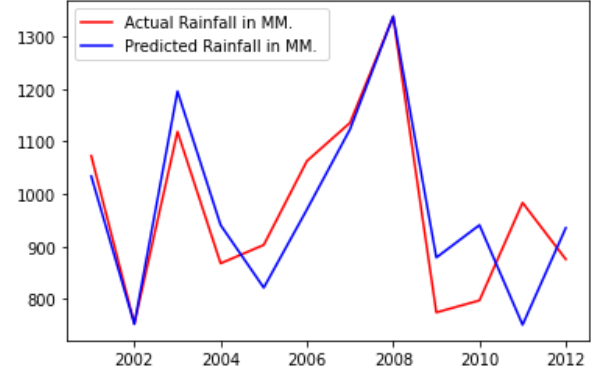


Figure 5.6: North-Central Indian rainfall forecasting using MLR

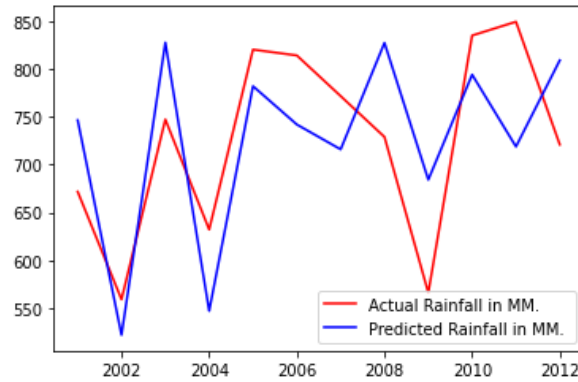


Figure 5.7: North-Central Indian rainfall forecasting using Ridge Regression

## 5.4 Approach-II Results

The suggested predictor identification method begins with pre-processing of input climate data. We utilise SLP data collected from geographic rectangular grids with dimensions of [20 longitude x 10 latitude] from all around the world. SLP monthly mean data from January to May is utilised and rebuilt using the ANN auto-encoder tool. Top 9 vectors which are most correlated with ISMR is extracted from the recreated data, as pre-processed data for each month. We try to develop the ML model using this pre-processed data of SLP. We analysis the prediction skill of  $[2^5 * 9]$  SVR models along with sliding window prototype, for the time period 2001-2018.

We experience that using top 6 vectors of April month with a sliding window of 31 length, the Pearson correlation between predicted ISMR and observed ISMR is 0.83. Figure 5.26 shows the prediction skill of this model.

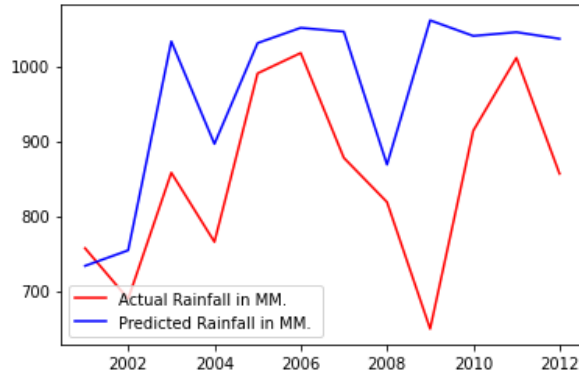


Figure 5.8: Indian East coastal region rainfall forecasting using SVR

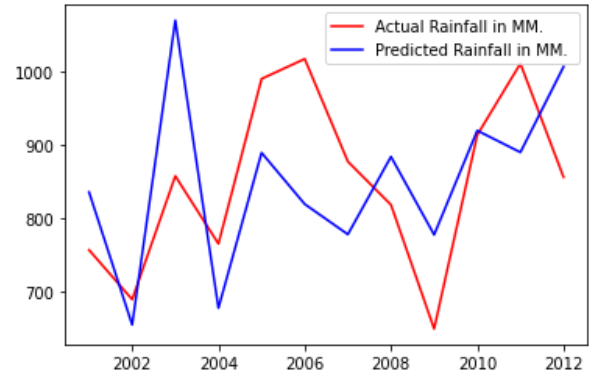


Figure 5.9: Indian East coastal region rainfall forecasting using MLR

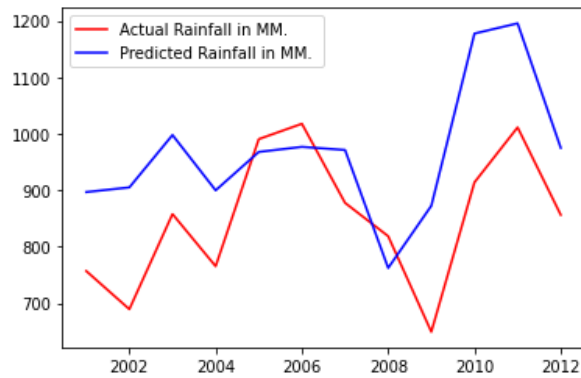


Figure 5.10: Indian East coastal region rainfall forecasting using Ridge Regression

Table 5.3: Observed rainfall VS forecated rainfall PCC analysis

Region	SVR		Multiple Linear Regression		Ridge Regression	
	WS	PCC	WS	PCC	WS	PCC
All India	29	0.82	25	0.55	29	0.79
Peninsula	26	0.74	32	0.57	25	0.67
North-east	28	0.65	18	0.61	28	0.65
West Coast	34	0.7	22	0.63	29	0.61
North-central	17	0.64	21	0.83	17	0.65
North-west	32	0.52	25	0.67	12	0.66
East coast	35	0.58	22	0.51	12	0.62
Western Himalaya	29	0.55	24	0.62	12	0.73

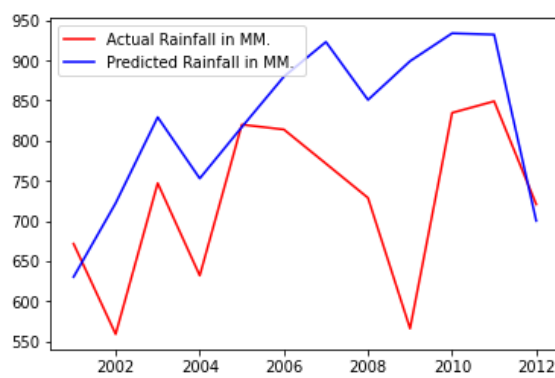


Figure 5.11: Western Himalaya region rainfall forecasting using SVR

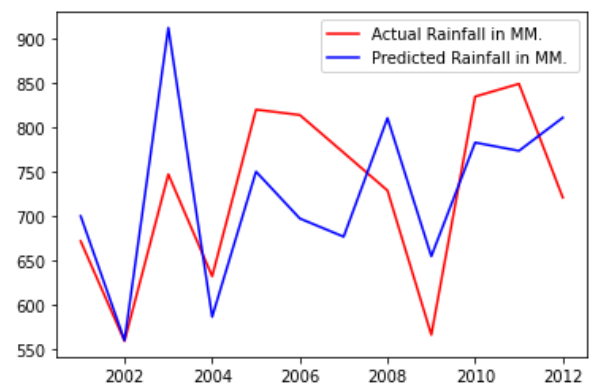


Figure 5.12: Western Himalaya region rainfall forecasting using MLR

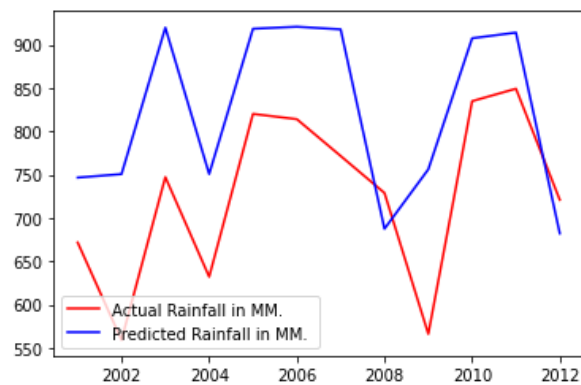


Figure 5.13: Western Himalaya region rainfall forecasting using Ridge Regression

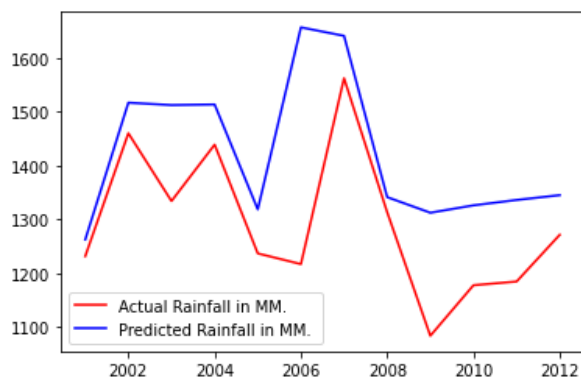


Figure 5.14: North-east region rainfall forecasting using SVR

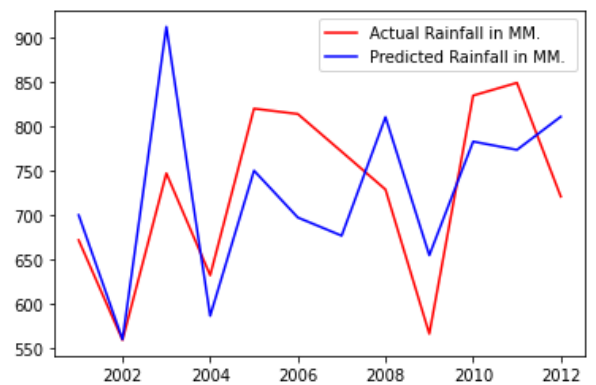


Figure 5.15: North-east region rainfall forecasting using MLR

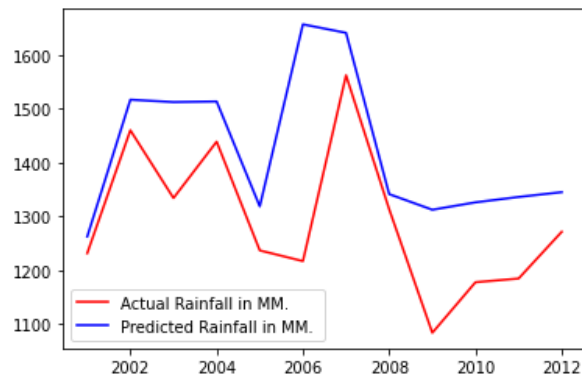


Figure 5.16: North-east region rainfall forecasting using Ridge Regression

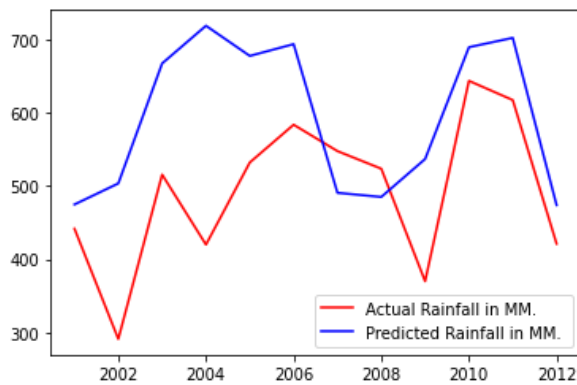


Figure 5.17: North-west region rainfall forecasting using SVR

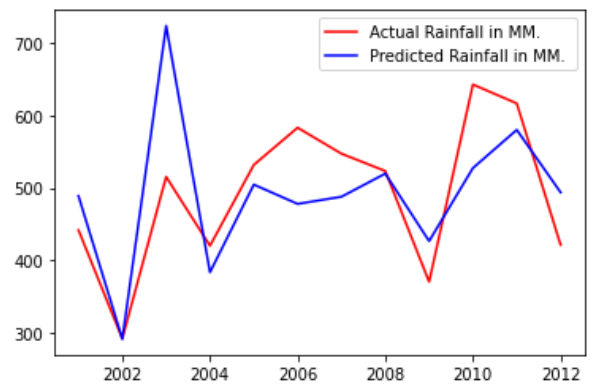


Figure 5.18: North-west region rainfall forecasting using MLR

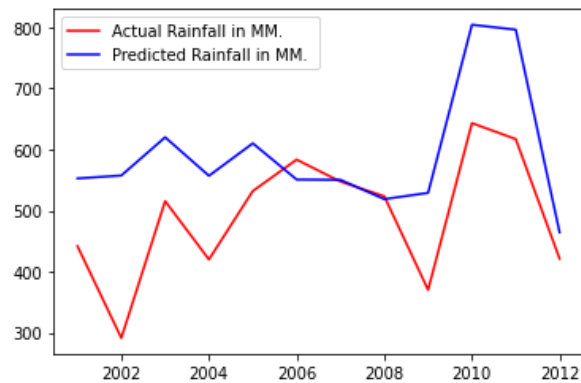


Figure 5.19: North-west region rainfall forecasting using Ridge Regression

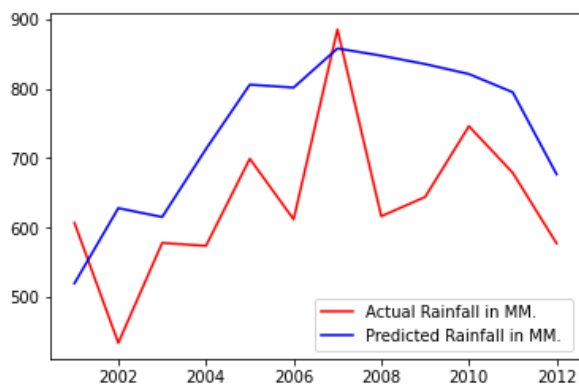


Figure 5.20: Interior Peninsula region rainfall forecasting using SVR

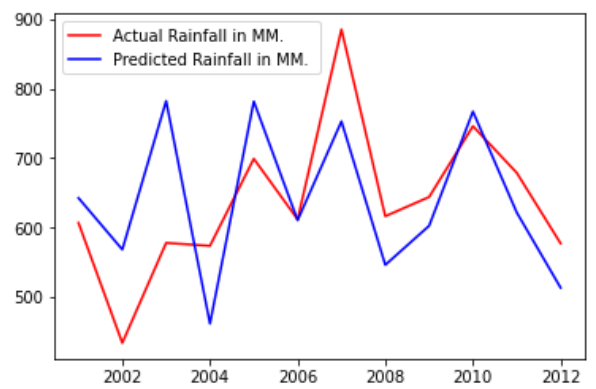


Figure 5.21: Interior Peninsula region rainfall forecasting using MLR

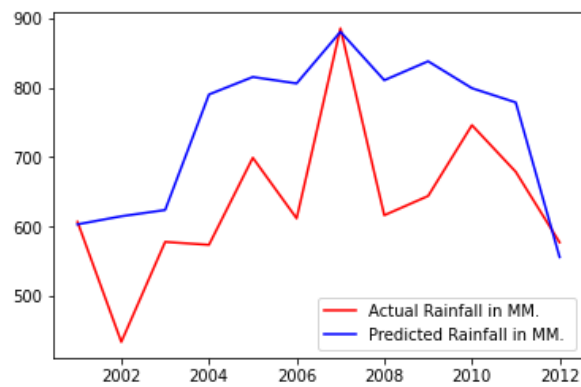


Figure 5.22: Interior Peninsula region rainfall forecasting using Ridge Regression

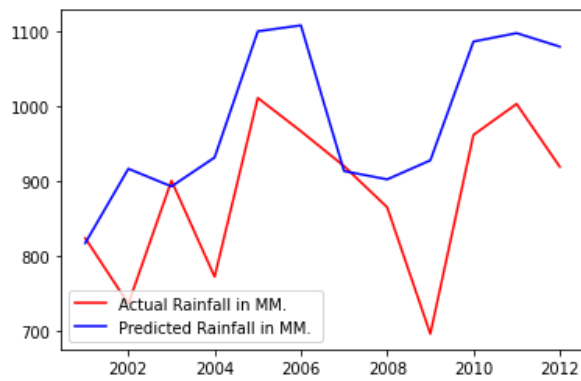


Figure 5.23: West coast region rainfall forecasting using SVR

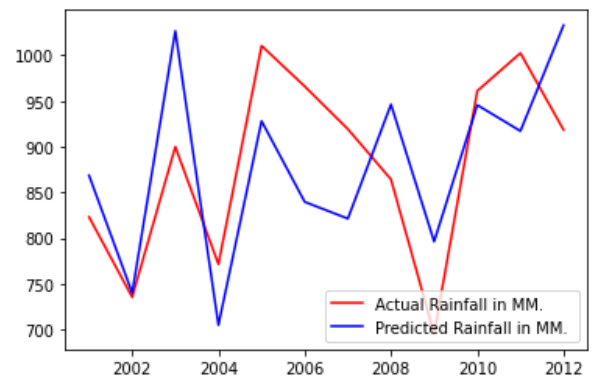


Figure 5.24: West coast region rainfall forecasting using MLR

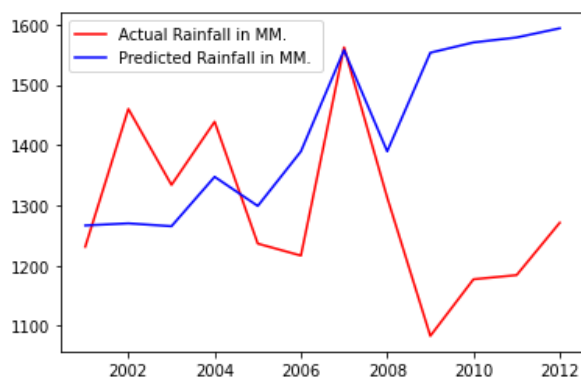


Figure 5.25: West coast region rainfall forecasting using Ridge Regression

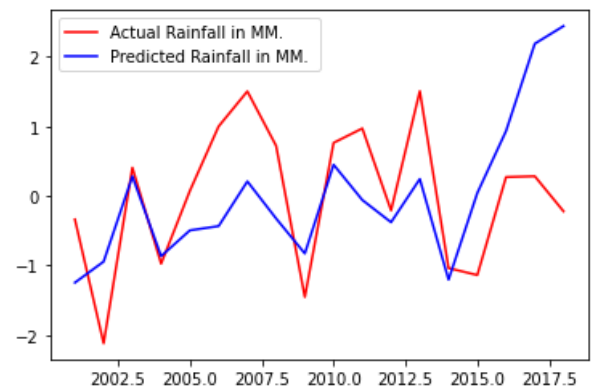


Figure 5.26: ISMR forecasting using Approach-II



## 5.5 Chapter Summary

In this chapter, we discuss about the impacts of predictors based upon their correlation with ISMR. The results of ML models depicts the findings and machine learning model results of our study in the diagrammatic ways.

## CHAPTER 6

### CONCLUSION AND FUTURE SCOPE

Agricultural and the associated products are the main pillar of the Indian economy which further depends on the monsoon precipitation. The contribution of the agricultural to national economy is declining every year due to low or excessive rainfall. The work done in the dissertation can be concluded as follows:

- The implicit challenges in the statistical predictive models such as periodical variation in the predictor-predictand linkage, predictor's inter-correlation, changing prediction skills etc., some changes should find-out in various ways, such as replacing the model size, the use of noval climate predictors, using the different combination of the climate predictors, enhancement via experimenting different the length of machine learning models training period etc.
- Two forecasting approach has been proposed with a better accuracy. First is the sliding window and SVR model with a correlation of (PCC) 83.27 for period [2001-2018].
- Second is, preparing new valuable predictors from sea level pressure over the entire globe using ANN auto-encoders and then only following 1st approach. This gives a predicted rainfall vector which is highly correlated (82.79) with the actual rainfall for the period [2001-2018].

However, there still remains a room for improvement as some issues & challenges have not been determined in the dissertation work. Some of the challenges are listed below:

- In our dissertation work, we tends to use 10 climate predictors with the best regression machine learning tools. Somehow accuracy can be increased by using more valuable and highly influencing climate parameters which are still hidden and needed to be dream up.
- In our dissertation work, we face major challenges in the data extraction as the authenticity of the data is major concern, because of very limited source of data.

# References

- Azhar, S. S. A., Chenoli, S. N., Samah, A. A., & Kim, S.-J. (2020). The linkages between antarctic sea ice extent and indian summer monsoon rainfall. *Polar Science*, 25, 100537.
- Capotondi, A., Wittenberg, A. T., Newman, M., Di Lorenzo, E., Yu, J.-Y., Braconnot, P., Cole, J., Dewitte, B., Giese, B., Guilyardi, E., et al. (2015). Understanding enso diversity. *Bulletin of the American Meteorological Society*, 96(6), 921–938.
- Charaniya, N., & Dudul, S. (2012). Focused time delay neural network model for rainfall prediction using indian ocean dipole index. *2012 Fourth International Conference on Computational Intelligence and Communication Networks*, 851–855.
- Clark, C. O., Cole, J. E., & Webster, P. J. (2000). Indian Ocean SST and Indian summer rainfall: Predictive relationships and their decadal variability. *Journal of Climate*, 13(14). [https://doi.org/10.1175/1520-0442\(2000\)013<2503:IOSAIS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2503:IOSAIS>2.0.CO;2)
- Dash, Y., Mishra, S. K., Sahany, S., & Panigrahi, B. K. (2018). Indian summer monsoon rainfall prediction: A comparison of iterative and non-iterative approaches. *Applied Soft Computing*, 70, 1122–1134.
- Dutta, U., Hazra, A., Saha, S. K., Chaudhari, H. S., Pokhrel, S., & Konwar, M. (2021). Examining the variability of cloud hydrometeors and its importance on the indian summer monsoon rainfall predictability. *arXiv preprint arXiv:2101.04521*.
- Jain, S., Scaife, A. A., & Mitra, A. K. (2019). Skill of indian summer monsoon rainfall prediction in multiple seasonal prediction systems. *Climate Dynamics*, 52(9), 5291–5301.
- Kashid, S. S., & Maity, R. (2012). Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using Genetic Programming. *Journal of Hydrology*, 454–455. <https://doi.org/10.1016/j.jhydrol.2012.05.033>
- Kumar, D. N., Reddy, M. J., & Maity, R. (2007). Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques. *Journal of Intelligent Systems*, 16(4). <https://doi.org/10.1515/jisys.2007.16.4.307>
- Kumar, V., Sunilkumar, K., & Sinha, T. (2021). Proportional trends of continuous rainfall in indian summer monsoon. *Remote Sensing*, 13(3), 398.
- Lim, E.-P., & Hendon, H. H. (2017). Causes and predictability of the negative indian ocean dipole and its impact on la niña during 2016. *Scientific reports*, 7(1), 1–11.
- Loo, Y. Y., Billa, L., & Singh, A. (2015). Effect of climate change on seasonal monsoon in asia and its impact on the variability of monsoon rainfall in southeast asia. *Geoscience Frontiers*, 6(6), 817–823.
- Lu, B., & Ren, H.-L. (2020). What caused the extreme indian ocean dipole event in 2019? *Geophysical Research Letters*, 47(11), e2020GL087768.
- Pandey, P., Dwivedi, S., Goswami, B., & Kucharski, F. (2020). A new perspective on enso-indian summer monsoon rainfall relationship in a warming environment. *Climate Dynamics*, 55(11), 3307–3326.
- Parthasarathy, B., Kumar, R., & Munot, A. (1996). Homogeneous regional summer monsoon rainfall over india: Interannual variability, teleconnections.
- Rajeevan, M., Pai, D. S., Dikshit, S. K., & Kelkar, R. R. (2004). IMD’s new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003. *Current Science*, 86(3).

- Rajeevan, M., Pai, D., & Thapliyal, V. (1998). Spatial and temporal relationships between global land surface air temperature anomalies and indian summer monsoon rainfall. *Meteorology and Atmospheric Physics*, 66(3), 157–171.
- Sabeerali, C., Ajayamohan, R., & Rao, S. A. (2019). Loss of predictive skill of indian summer monsoon rainfall in ncep cfsv2 due to misrepresentation of atlantic zonal mode. *Climate Dynamics*, 52(7), 4599–4619.
- Saha, M., & Mitra, P. (2019). Identification of Indian monsoon predictors using climate network and density-based spatial clustering. *Meteorology and Atmospheric Physics*, 131(5). <https://doi.org/10.1007/s00703-018-0637-y>
- Saha, M., Santara, A., Mitra, P., Chakraborty, A., & Nanjundiah, R. S. (2021). Prediction of the indian summer monsoon using a stacked autoencoder and ensemble regression model. *International Journal of Forecasting*, 37(1), 58–71.
- Sahai, A. K., Chattopadhyay, R., & Goswami, B. N. (2008). A SST based large multi-model ensemble forecasting system for Indian summer monsoon rainfall. *Geophysical Research Letters*, 35(19). <https://doi.org/10.1029/2008GL035461>
- Santoso, A., McPhaden, M. J., & Cai, W. (2017). The defining characteristics of enso extremes and the strong 2015/2016 el niño. *Reviews of Geophysics*, 55(4), 1079–1129.
- Shahi, N. K., Rai, S., & Mishra, N. (2019). Recent predictors of Indian summer monsoon based on Indian and Pacific Ocean SST. *Meteorology and Atmospheric Physics*, 131(3). <https://doi.org/10.1007/s00703-018-0585-6>
- Singh, P., & Borah, B. (2013). Indian summer monsoon rainfall prediction using artificial neural network. *Stochastic environmental research and risk assessment*, 27(7), 1585–1599.
- Singh, P., Rabadiya, K., & Dhiman, G. (2018). A four-way decision-making system for the indian summer monsoon rainfall. *Modern Physics Letters B*, 32(25), 1850304.
- Venugopal, T., Ali, M., Bourassa, M., Zheng, Y., Goni, G., Foltz, G., & Rajeevan, M. (2018). Statistical evidence for the role of southwestern indian ocean heat content in the indian summer monsoon rainfall. *Scientific reports*, 8(1), 1–10.
- Vittal, H., Villarini, G., & Zhang, W. (2020). Early prediction of the indian summer monsoon rainfall by the atlantic meridional mode. *Climate Dynamics*, 54(3), 2337–2346.
- Wang, C., Deser, C., Yu, J.-Y., DiNezio, P., & Clement, A. (2017). El niño and southern oscillation (enso): A review. *Coral reefs of the eastern tropical Pacific*, 85–106.
- Wang, G., Cai, W., Yang, K., Santoso, A., & Yamagata, T. (2020). A unique feature of the 2019 extreme positive indian ocean dipole event. *Geophysical Research Letters*, 47(18), e2020GL088615.
- Yeh, S.-W., Cai, W., Min, S.-K., McPhaden, M. J., Dommenges, D., Dewitte, B., Collins, M., Ashok, K., An, S.-I., Yim, B.-Y., et al. (2018). Enso atmospheric teleconnections and their response to greenhouse gas forcing. *Reviews of Geophysics*, 56(1), 185–206.
- Zahan, Y., Mahanta, R., Rajesh, P., & Goswami, B. (2021). Impact of climate change on north-east india (nei) summer monsoon rainfall. *Climatic Change*, 164(1), 1–19.
- Zhang, T., Wang, T., Krinner, G., Wang, X., Gasser, T., Peng, S., Piao, S., & Yao, T. (2019). The weakening relationship between eurasian spring snow cover and indian summer monsoon rainfall. *Science advances*, 5(3), eaau8932.