Meteorological Drought Forecasting Using Machine Learning Techniques

---

A Report Submitted

For B.Tech Project Part II (CE 4291)

By

ANIRBAN GHOSH (Enrollment No.-510419003)

SHAYAN GHOSH (Enrollment No.-510419009)

Under the supervision of

DR. SUJATA BISWAS

Associate Professor

Department of Civil Engineering

Indian Institute of Engineering Science & Technology, Shibpur

Howrah -711103

MAY-2023

Department of Civil Engineering

Indian Institute of Engineering Science & Technology, Shibpur

Howrah -711103

FOREWARD

*I hereby forward the Preliminary Report entitled **"Meteorological Drought Forecasting Using Machine Learning Techniques"** submitted by **ANIRBAN GHOSH (Enrollment No.-510419003)** and **SHAYAN GHOSH(Enrollment No.-510419009)** under my guidance and supervision in partial fulfillment of the requirements for the award of the degree of '**BACHELOR OF TECHNOLOGY'** in **Civil Engineering** in the department of Civil Engineering, Indian Institute of Engineering Science and Technology, Shibpur.*

Dr. Sujata Biswas

Associate Professor

Department of Civil Engineering

Indian Institute of Engineering Science & Technology, Shibpur

ACKNOWLEDGEMENT

I hereby convey my sincere respect, thanks and heartfelt gratitude to my supervisor **Dr. Sujata Biswas** for her immense support, monitoring and guidance. I would also like to extend my gratitude to **Mr. Suvro Aon and Mr. Subimal Nandi**, research scholar, who helped me throughout this journey. This 4th Year Major Project would never have been completed without their constant help and encouragement. The blessing, help and guidance given by them time to time shall carry us a long way in the journey of our life.

Anirban Ghosh

Enrollment No.- 510419003

Shayan Ghosh

Enrollment No.- 510419009

Date: 18.05.2023

Department of Civil Engineering
Indian Institute of Engineering Science & Technology, Shibpur
Howrah -711103, West Bengal.

CONTENTS

CHAPTER 1
1. Introduction..2
   1.1 Drought...2
   1.2 Classification of drought..3
      1.2.1 On the basis of source of water availability...3
      1.2.2 On the basis of source of time of occurrence...3
      1.2.3 On the basis of source of medium...4
   1.3 Impacts of Drought..4
      1.3.1 Impact on Public Health...4
      1.3.2 Impact on Agriculture..4
      1.3.3 Impact on Transportation...5
      1.3.4 Impact on Economy...5
      1.3.5 Impact on Environment..5
   1.4 Drought Map of India...6
   1.5 Drought Forecasting In Modern Times..7
      1.5.1 Drought Indices...7
      1.5.2 Discussion on SPI..8
      1.5.3 Discussion on Machine Learning..10
      1.5.4 Deep Learning...10
      1.5.5 Neural Network..13
      1.5.6 Time Series Forecasting...13
      1.5.7 Gradient Descent..14
      1.5.8 Stochastic Gradient Descent..15
      1.5.9 Adam optimizer...15
      1.5.10 Stationarity & Augmented Dickey- fuller Test......................................15
      1.5.11 Activation function...15
      1.5.12 Types of Activation Functions...16

CHAPTER 2
2. Literature Review & A Detailed view on SPI...19
   2.1 Literature Review on SPI...19
   2.2 Literature Review on Machine Learning...19

CHAPTER 3
3. Study area & Data..20
   3.1 Study area..20
   3.2 Data collection..23

CHAPTER 4
4. Methodology...24
   4.1 Standardized Precipitation Index (SPI)..24
      4.1.1 Normalization of Data...24
      4.1.2 Computation of SPI...25
   4.2 Mathematical aspects of Adam optimizer...25
      4.2.1 Momentum..25
      4.2.2 Root Mean Square Propagation (RMSP)..26
   4.3 Mathematical aspects of activation function in neural network.....................27

`

4.4 Methodology of LSTM....28
4.5 Mean Squared Error (MSE)....30
4.6 Root Mean Squared Error (RMSE)....31
4.7 Normalized root mean squared error (NRMSE)....31
4.8 R Squared....31
CHAPTER 5
5. Results and Discussions....32
5.1 Drought identification using SPI....32
5.2 Forecasting of Meteorological Drought using LSTM model....35
5.2.1 Regression graphs....36
5.2.2 Error Evaluation Data....37
5.2.3 Error Histograms....38
5.2.4 Comparison of results between observed and predicted data....39
CHAPTER 6
6. Conclusions....41
CHAPTER 7
7. References....42

`

ABSTRACT

Prolonged drought affects the natural environment of any region. Therefore, drought forecasting plays an important role in the planning and management of river basin natural resources and water resource systems. Neural networks and stochastic models have performed well in modelling and forecasting SPI time series. Considering the severity of droughts, different random models will be developed using LSTM technique.Using the Standardized Precipitation Index (SPI) series, LSTM models will be applied to forecast droughts in the Bankura district of the state of West Bengal, India. Computation of those indices requires precipitation data which will be downloaded from the India Meteorological Department (IMD) website. The computed SPI time series will help to characterize meteorological drought viz. duration and severity of the event.

Key Words: LSTM, SPI, Drought Forecasting, Bankura.

CHAPTER 1

1. INTRODUCTION

Droughts produce a complex set of impacts that span many sectors of the economy and reach well beyond the area experiencing the drought. Since 1967, about half of nearly 2.8 billion people in the world who have suffered from all natural disasters have been affected by droughts. Out of 3.5 million people killed by disasters, 1.3 million human lives have directly or indirectly been claimed by droughts (Obasi et al.,1994). Like other countries, droughts are also common in India. Although drought-affected areas are mainly confined to the peninsular and western parts of the country, there are drought-prone pockets in other parts of India as well. Out of 3.28 million km^2 of the geographical area of India, about 1.07 million km2 of land is subjected to different degrees of water stress and drought conditions (Subramanya et al. 2005). This condition is being further aggravated by the rising demands for municipal and industrial water due to the growing population coupled with the rising standard of living. Droughts differ from other natural hazards such as floods, tropical cyclones, and earthquakes in several ways. Since the effects of a drought often accumulate slowly over a considerable period of time, they may linger for several years after the drought period ends. As a result, the onset and end of a drought are difficult to determine precisely, and that is why a drought is often referred to as a creeping phenomenon.

1.1 Drought

Drought is a normal feature of climate and its recurrence is inevitable. However, there remains much confusion within the scientific and policy-making community about its characteristics. Research has shown that the lack of a precise and objective definition in specific situations has been an obstacle to understanding drought which has led to indecision and inaction on the part of managers, policy-makers, and others. A drought may mean different things to different people. For example, for a meteorologist it is a deviation from normal precipitation; for a hydrologist a fall in stream flow, lake level, or groundwater level; for an agricultural scientist lack of soil moisture to sustain crop growth; for an economist a famine condition; and for an urbanite shortage of tap water supply.

All points of view seem to agree that drought is a condition of insufficient moisture caused by a deficit in precipitation over some time period. (Dracup et al. 1980). The IPCC Sixth Assessment Report defines a drought simply as drier than normal conditions. This means that a drought is a moisture deficit relative to the average water availability at a given location and season. According to the National Integrated Drought Information System, a multi-agency partnership, drought is generally defined as a deficiency of precipitation over an extended period of time (usually a season or more), resulting in a water shortage. The National Weather Service office of the NOAA (National Oceanic and Atmospheric Administration, USA) defines drought as a deficiency of moisture that results in adverse impacts on people, animals, or vegetation over a sizable area.

1.2 Classification of Drought

Wilhite and Glantz (1985) completed a thorough review of dozens of drought definitions and identified six overall categories: meteorological, climatological, atmospheric, agricultural, hydro-logic and water management. There are three major categories of drought based on where in the water cycle the moisture deficit occurs: meteorological drought, hydro logical drought, and agricultural or ecological drought. Meteorological drought occurs when there is a prolonged time with less than average precipitation. Hydro logical drought is brought about when the water reserves available in sources such as aquifers, lakes and reservoirs fall below a locally significant threshold. Agricultural or ecological droughts affect crop production or ecosystems in general. On the other hand Drought differs in time and period of their occurrence and on this basis Thormathwite delineated following three areas:Permanent drought, seasonal drought and contingent drought. Also on the basis of medium on which drought occurs Mexico(1929) classified drought into two types: soil drought and atmospheric drought.

1.2.1 On the basis of source of water availability

Drought is classified into three types on the basis of water availability.

A. Meteorological drought: Meteorological droughts are primarily indicative of varying degrees of lack of rainfall.The IMD classified the drought as follows, based on the decrease in rainfall.

a) Slight drought : When rainfall is 11 to 25% less from the normal rainfall.

b) Moderate drought: When rainfall is 26 to 50% less than the normal rainfall.

c) Severe drought: When rainfall is more than 50% less than the normal rainfall.

B. Hydro-logical drought: It is defined as a situation of lack of precipitation when water sources such as streams, rivers, lakes and wells are depleted and the groundwater table is low. This affects industry and power generation.

C. Agricultural drought: This is a situation that results from insufficient precipitation when the soil moisture is too low to meet the water needs during plant growth. This can lead to crop wilting and reduced yields due to soil moisture stress.

1.2.2 On the basis of source of time of occurrence: Drought is classified into three types on the basis of occurrence.

A. Permanent drought: This is generally the territory of permanent arid, arid and desert areas. Without irrigation, there is insufficient rainfall for agricultural production. In these areas, vegetation such as cacti. Commonly observed are thorn bushes, xerophytes, etc.

B. Seasonal drought: It occurs in areas that are well defined as wet (wet) and dry climates. Large seasonal cycles can lead to seasonal droughts. This occurs in monsoon regions.

C. Contingent drought: This result due to irregular and variability in rainfall, especially in humid and sub humid regions. The occurrence of such droughts may coincide with grand growth periods of the crops when the water needs are critical and greatest resulting into severity of the effects i.e. yield reduction.

1.2.3 On the basis of source of medium:

1.2.3 On the basis of source of medium: Drought is classified into three types on the basis of medium.

A. **Soil drought:** This is a condition in which soil moisture has decreased and is not meeting the potential evapotranspiration of crops.
B. **Atmospheric drought:** This is a result of low humidity, dry hot air, which leads to dehydration of the plant. This can occur even with adequate rainfall and water supply.

1.3 Impacts of Drought

1.3.1 Impact on Public Health

When drought leads to water and food shortages, it can have many health implications for those affected, increasing the risk of illness and death. Drought may have acute and chronic health effects, including, malnutrition due to the decreased availability of food, including micro nutrient deficiency, such as iron-deficiency anaemia. Increased risk of infectious diseases, such as cholera, diarrhoea, and pneumonia, due to acute malnutrition, lack of water and sanitation, and displacement. Psycho-social stress and mental health disorders. Disruption of local health services due to a lack of water supplies, loss of buying power, migration and/or health workers being forced to leave local areas etc. Severe drought also affects air quality, making wildfires and dust storms more likely, posing health risks for people who already have lung diseases such as asthma, chronic obstructive pulmonary disease (COPD), and heart disease. increase.

1.3.2 Impact on Agriculture

Agriculture bears most of the impacts and is the most affected sector in developing countries, absorbing up to 80% of direct impacts and improving water availability, agricultural production, food security and rural livelihoods. It has multiple impacts on your life. With about 1.3 billion people (40% of the world) relying on agriculture as their main source of income, drought threatens the livelihoods of many, halting and often ineffective progress in food security and poverty reduction. and hinders efforts to achieve SDG1 and SDG2. Experience shows that proactive and risk-based management approaches are effective in improving community resilience and ability to cope with drought, but drought management and planning can be overlooked until the crisis becomes apparent. It often happens. This passive, crisis-driven response results in a fragmented policy space where interventions are isolated by sector and drought mitigation strategies do not work well. FAO also supports the SDGs, the Paris Agreement on Climate Change, and other international actors in developing national drought management strategies and their recent efforts to support the transition to a proactive drought management approach at the heart of strategic strategies. We are doing our best to further increase the momentum that has been brought to us. Aim for FAO to make life more resilient to threats and crises. FAO’s support materializes through the following actions:

Awareness creation to improve understanding among policy-makers and decision-makers of the importance of drought risk management (e.g. through the organization of high-level events, such as the International Seminar on Drought and Agriculture, MHNDP, the African Drought Conference). The development of guidelines tailored for specific drought-prone regions and of other technical tools to facilitate the adoption of proactive drought management policies at the country level. Hosting the Drought & Agriculture Forum, a common learning, sharing and planning platform on best practices in drought management to enable the design of multi-sectoral initiatives that benefit all stakeholders.

1.3.3 Impact on Transportation

When drought occurs, water levels in rivers and waterways begin to drop. As the drought worsens, water transportation becomes increasingly restricted, leading to higher transportation costs. For example, the 2012 Great Plains drought closed the Mississippi River several times, causing him $300 million in trade losses for each day the river was closed. Increased shipping costs are usually passed on to consumers and reflected in higher market prices. Drought can affect more than waterways. Rising temperatures, usually associated with drought, often cause cracks and damage to pavement, requiring government agencies to allocate emergency funds for repairs. Airport runways can soften and deteriorate in extreme heat. Additionally, some small aircraft used for regional transport cannot fly when temperatures exceed 118 degrees. This problem is likely to become more prevalent as average temperatures continue to rise. Drought can also increase wildfire activity and cause visibility problems that affect both planes and cars, leading to derailments and train schedules. Sun buckles can also occur on commuter trains and can cause injuries if commuter service is interrupted and major derailments occur.When water supply is low and demand is high, more groundwater is withdrawn, which can lead to land subsidence (subsidence). This destabilizes the ground, affects infrastructure (buildings, roads, water supplies, etc.) and even leads to the formation of sinkholes.

1.3.4 Impact on Economy

The economic impact is the costly impact of drought on people (or businesses). Below are just a few examples of the economic impact. Farmers may lose money if a drought destroys their crops. If a farmer's water supply is too low, the farmer may have to spend more money on irrigation or to drill new wells. Ranchers may have to spend more money on feed and water for their animals. Businesses that depend on farming, like companies that make tractors and food, may lose business when drought damages crops or livestock. People who work in the timber industry may be affected when wildfires destroy stands of timber. Businesses that sell boats and fishing equipment may not be able to sell some of their goods because drought has dried up lakes and other water sources. Power companies that normally rely on hydroelectric power (electricity that's created from the energy of running water) may have to spend more money on other fuel sources if drought dries up too much of the water supply. The power companies' customers would also have to pay more. Water companies may have to spend money on new or additional water supplies. Barges and ships may have difficulty navigating streams, rivers, and canals because of low water levels, which would also affect businesses that depend on water transportation for receiving or sending goods and materials.

1.3.5 Impact on Environment

Drought also affects the environment in many ways. Just like humans, plants and animals depend on water.During drought, their food supply is reduced and habitat can be compromised. Damage may be temporary, and habitat and food supplies may return to normal once the drought is over. However, the environmental effects of drought can last for long periods of time, even forever. Examples of environmental impacts include: Losses or destruction of fish and wildlife habitat. Lack of food and drinking water for wild animals. Increase in disease in wild animals, because of reduced food and water supplies. Migration of wildlife. Increased stress on endangered species or even extinction. Lower water levels in reservoirs, lakes, and ponds. Loss of wetlands. More wildfires etc.

1.4 Drought Map of India

In the past, droughts have periodically led to major Indian famines, including the Bengal famine of 1770, in which up to one third of the population in affected areas died; the 1876–1877 famine, in which over five million people died; and the 1899 famine, in which over 4.5 million died. The 2013 Maharashtra drought affected 25 million people.In simple words, drought has destroyed India on a large scale. Eighteen meteorological and 16 hydrological droughts occurred in India between 1870 and 2018. The most severe meteorological droughts were in the years 1876, 1899, 1918, 1965, and 2000, while the five worst hydrological droughts occurred in the years 1876, 1899, 1918, 1965, and 2000. The drought of 1899 can be classified as meteorological as well as hydrological and was the most severe documented drought India has ever experienced to date. A typical picture of drought prone area of India is shown below.

Figure1: Drought prone areas in India.

1.5 Drought Forecasting In Modern Times

Drought indices are quantitative measures that characterize drought levels by data assimilation from one or several variables (indicators) such as precipitation and evapotranspiration into a single numerical value.The nature of drought indices reflects different events and conditions: they may reflect climatic dryness anomalies (based on precipitation) or correspond to delayed agricultural and hydrological impacts such as loss of soil moisture or lowering of underground water levels.Several forecasting models have been introduced for forecasting droughts such as: LSTM, ANN, Auto-regression Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR) and Markov Chain. Most of them are probabilistic, highly stochastic, and non-linear. Comprehensive drought indices use a variety of meteorological, agricultural and hydrological variables to draw a comprehensive picture of drought.

1.5.1 Drought Indices: Commonly, drought indices are categorized based on the type of impacts they relate to. The taxonomy can also be based on the variables they relate to (Steinemann et al. 2005) or use of disciplinary data (Niemeyer 2008). Three popular categories are meteorological, agricultural and hydrological drought indices. Niemeyer (2008) adds three categories to this list: comprehensive, combined and remote-sensing-based drought indices. Comprehensive drought indices use a variety of meteorological, agricultural and hydrological variables to draw a comprehensive picture of drought. Major operational drought indices are Standardized Precipitation Index (SPI), Standardized Precipitation-Evapotranspiration Index (SPEI), Palmer Drought Severity Index (PDSI) etc.

A. Standardized Precipitation Index (SPI): It is an index based on the probability of cumulative precipitation for any time scale. The understanding that a deficit of precipitation has different impacts on the ground water, reservoir storage, soil moisture, snow pack, and stream flow led McKee (1993) to develop the Standardized Precipitation Index (SPI). The SPI was designed to quantify the precipitation deficit for multiple time scales. These time scales reflect the impact of drought on the availability of the different water resources. Soil moisture conditions respond to precipitation anomalies on a relatively short scale, while groundwater, stream flow, and reservoir storage reflect the longer-term precipitation anomalies. For these reasons, McKee (1993) originally calculated the SPI for 3, 6,12, 24, and 48-month time scales.

B. Standardized Precipitation-Evapotranspiration Index (SPEI): The SPEI is a multi scale drought index based on climatic data. It can be used for determining the onset, duration and magnitude of drought conditions with respect to normal conditions in a variety of natural and managed systems such as crops, ecosystems, rivers, water resources, etc.

C. Palmer Drought Severity Index (PDSI): In 1965, Palmer developed an index to measure the departure of the moisture supply. Palmer based his index on the supply-and-demand concept of the water balance equation, taking into account more than just the precipitation deficit at specific locations. The objective of the Palmer Drought Severity Index (PDSI), as this index is now called, was to provide measurements of moisture conditions that were standardized so that comparisons using the index could be made between locations and between months. In this project forecasting of drought using machine learning we are using the concept of SPI, so further we can elaborate our literature survey in SPI or Standardized Precipitation Index.

1.5.2 Discussion on SPI

In this study SPI is described elaborately as it is drought index, being used for forecasting of drought in the context of this project. Over the years, many drought indices were developed and used by meteorologists and climatologists around the world. Those ranged from simple indices such as percentage of normal precipitation and precipitation percentiles to more complicated indices such as the Palmer Drought Severity Index. However, scientists in the United States realized that an index needed to be simple, easy to calculate and statistically relevant and meaningful. Moreover, the understanding that a deficit of precipitation has different impacts on groundwater, reservoir storage, soil moisture, snow pack and stream flow led American scientists McKee, Doesken and Kleist to develop the Standardized Precipitation Index (SPI) in 1993.

A. Types

SPI-1 to SPI-3: When SPI is calculated over a shorter accumulation period (e.g., 1-3 months), it can be used as an indicator of immediate impacts such as reduced soil moisture, snow cover, and smaller river flows.

SPI-3 to SPI-12: When the SPI is calculated over a medium retention period (e.g., 3 to 12 months), it can be used as an indicator of stream flow decline and reservoir retention.

SPI-12 to SPI-48: When SPI is calculated for longer retention periods (e.g., 12 to 48 months), it can be used as an indicator of reservoir depletion and groundwater recharge.

B. Values

McKee and others (1993) used the classification system shown in the SPI value table below (Table 1) to define drought intensities resulting from the SPI. They also defined the criteria for a drought event for any of the timescales. A drought event occurs any time the SPI is continuously negative and reaches an intensity of -1.0 or less. The event ends when the SPI becomes positive. Each drought event, therefore, has a duration defined by its beginning and end, and an intensity for each month that the event continues. The positive sum of the SPI for all the months within a drought event can be termed the drought's "magnitude". SPI values are shown below.

SPI Values	Condition
>2	extremely wet
1.5 to 1.99	very wet
1.0 to 1.49	moderately wet
-0.99 to 0.99	near normal
-1.0 to –1.49	moderately dry
-1.5 to –1.99	severely dry
<-2	extremely dry

Table 1: SPI Values w.r.t drought severity condition

The definition of drought thus far has included a beginning date, ending date, and a current drought intensity. Duration of drought can be either a current duration since the beginning or the duration of a historic drought event from beginning to ending. Peak intensity can easily be determined from the SPI. A measure of the accumulated magnitude of the drought can be included.

C. Use

Many drought planners appreciate the SPI's versatility. It is also used by a variety of research institutions, universities, and National Meteorological and Hydrological Services across the world as part of drought monitoring and early warning efforts.

D. Advantages & Disadvantages

Advantages

Due to the fact that SPI values are given in units of standard deviations from the long-term average, SPI can be calculated and compared at any geographic location and any number of time scales. In addition, the SPI is normalized, so it works just as well for wet season and cycle analysis as it does for dry season and cycle analysis. The SPI index is less complex to calculate than other drought indices, such as evaporation, because it is based on only one input parameter (precipitation accumulation).

Disadvantages

As mentioned earlier, the parameters of the gamma distribution are calculated (fitted) using historical non-zero rainfall accumulations. The cumulative probability of observed rainfall is then adjusted using the probability (frequency) of zero rainfall accumulations. However, if there are many historical zero rainfall accumulations, the estimated gamma distribution may not adequately fit the frequency distribution of the historical rainfall. Therefore, in regions with a high probability of zero rainfall, the SPI indicator should be interpreted with care. In such cases (typical of arid climates), the concept of a drought needs to be adapted, and it may be best to restrict SPI calculation and analysis to the normal rainy season, or to use alternative drought indicators. In Europe this applies to very limited areas.Because SPI is based only on precipitation, it does not address the effects of high temperatures on drought conditions, such as by damaging cultivated and natural ecosystems, and increasing evapotranspiration and water stress. A new variation of SPI - the Standardized Precipitation and Evapotranspiration Index (SPEI) - has been developed, which includes precipitation and temperature, in order to identify increases in drought severity linked with higher water demand by evapotranspiration.]

1.5.3 Discussion on Machine Learning

Machine Learning is a subset of Artificial intelligence (AI) which provides machine the ability to learn automatically and improve from experience without being explicitly programmed. It uses algorithms to parse data, learn from the data, and make informed decisions on what it has learned.

A. Types

Supervised Learning: Supervised Learning is a type of Machine Learning in which we teach machine using labeled data. Popular algorithm used in the case of supervised learning are linear regression, logistic regression, Support Vector Machine, K nearest Neighbour, Random forest method. Its basically used for Risk Evaluation and Forecast Sales.

Unsupervised Learning: In unsupervised learning the machine is trained on unlabeled data without any guidance.In this case popular algorithm used are K-Means, Apriori, C-Means.Its basically used in Recommendation Systems and anomaly detection systems.

Reinforcement Learning: In Reinforcement Learning an agent interacts with its environment by producing actions and discovers errors or rewards. In this case popular algorithm used are Q-Learning, SARSA etc.Its basically used in Self driving cars and gaming.

B. Applications

Self driving cars, Facebook Face Recognition, VR Games-Kinect, Voice recognition, Boston Dynamics, Alpha Go Games etc.

1.5.4 Deep Learning

Deep Learning is a particular kind of machine learning that is inspired by the functionality of our brain cells called neurons which led to the concept of artificial neural network. It structures algorithms in layers to create an artificial neural network that can learn and make intelligent decisions on it’s own. Deep Learning and Machine learning both falls under the category of artificial intelligence, deep learning is usually what’s behind the most human like artificial intelligence.While deep learning algorithms feature self-learning representations, they depend upon ANNs that mirror the way the brain computes information. During the training process, algorithms use unknown elements in the input distribution to extract features, group objects, and discover useful data patterns. Much like training machines for self-learning, this occurs at multiple levels, using the algorithms to build the models. Deep learning models make use of several algorithms. While no one network is considered perfect, some algorithms are better suited to perform specific tasks. To choose the right ones, it’s good to gain a solid understanding of all primary algorithms.

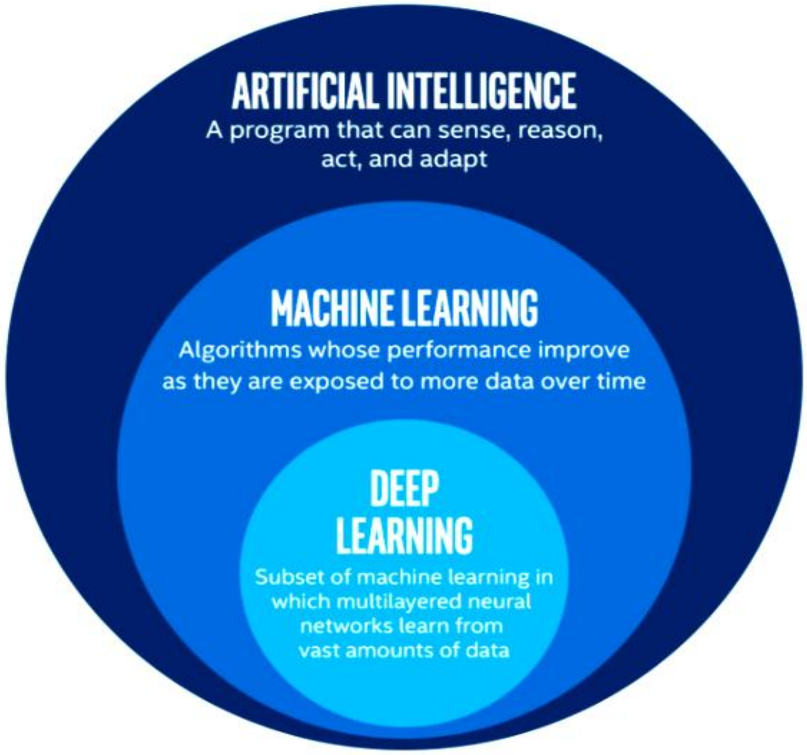


Figure 2: Schematic view of Machine Learning, AI, Deep Learning

A. Types of Algorithms used in Deep Learning

Here is the list of top 10 most popular deep learning algorithms:Conventional Neural Networks (CNN), Artificial Neural Networks (ANN), Long Short Term Memory Networks (LSTM), Recurrent Neural Networks (RNN), Generative Adversarial Networks (GAN), Radial Basis Function Networks (RBFN), Self Organizing Maps (SOM), Deep Belief Networks (DBN), Restricted Boltzmann Machines(RBM). In this study LSTM is described elaborately as it is being used for forecasting of drought in the context of this project.

B. Long Short Term Memory (LSTM)

Long Short Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does not give an efficient performance. LSTM can by default retain the information for a long period of time. It is used for processing, predicting, and classifying on the basis of time-series data. It is specifically designed to handle sequential data, such as time series, speech, and text. LSTM networks are capable of learning long-term dependencies in sequential data, which makes them well suited for tasks such as language translation, speech recognition, and time series forecasting.A traditional RNN has a single hidden state that is passed through time, which can make it difficult for the network to learn long-term dependencies. LSTMs address this problem by introducing a memory cell, which is a container that can hold information for an extended period of time. The memory cell is controlled by three gates: the input gate, the forget gate, and the output gate.

These gates decide what information to add to, remove from, and output from the memory cell. The input gate controls what information is added to the memory cell. The forget gate controls what information is removed from the memory cell. And the output gate controls what information is output from the memory cell. This allows LSTM networks to selectively retain or discard information as it flows through the network, which allows them to learn long-term dependencies. LSTMs can be stacked to create deep LSTM networks, which can learn even more complex patterns in sequential data. LSTMs can also be used in combination with other neural network architectures, such as Convolutional Neural Networks (CNNs) for image and video analysis.

C. Applications of LSTM:

Language Modeling: LSTMs have been used for natural language processing tasks such as language modeling, machine translation, and text summarization. They can be trained to generate coherent and grammatically correct sentences by learning the dependencies between words in a sentence.

Speech Recognition: LSTMs have been used for speech recognition tasks such as transcribing speech to text and recognizing spoken commands. They can be trained to recognize patterns in speech and match them to the corresponding text.

Time Series Forecasting: LSTMs have been used for time series forecasting tasks such as predicting stock prices, weather, and energy consumption. They can learn patterns in time series data and use them to make predictions about future events.

Anomaly Detection: LSTMs have been used for anomaly detection tasks such as detecting fraud and network intrusion. They can be trained to identify patterns in data that deviate from the norm and flag them as potential anomalies.

Recommender Systems: LSTMs have been used for recommendation tasks such as recommending movies, music, and books. They can learn patterns in user behavior and use them to make personalized recommendations.

Video Analysis: LSTMs have been used for video analysis tasks such as object detection, activity recognition, and action classification. They can be used in combination with other neural network architectures, such as Convolutional Neural Networks (CNNs), to analyze video data and extract useful information.

1.5.5 Neural Network

A neural network is structured like the human brain and consists of artificial neurons, also known as nodes. These nodes are stacked next to each other in three layers: The input layer, The hidden layer(s), The output layer

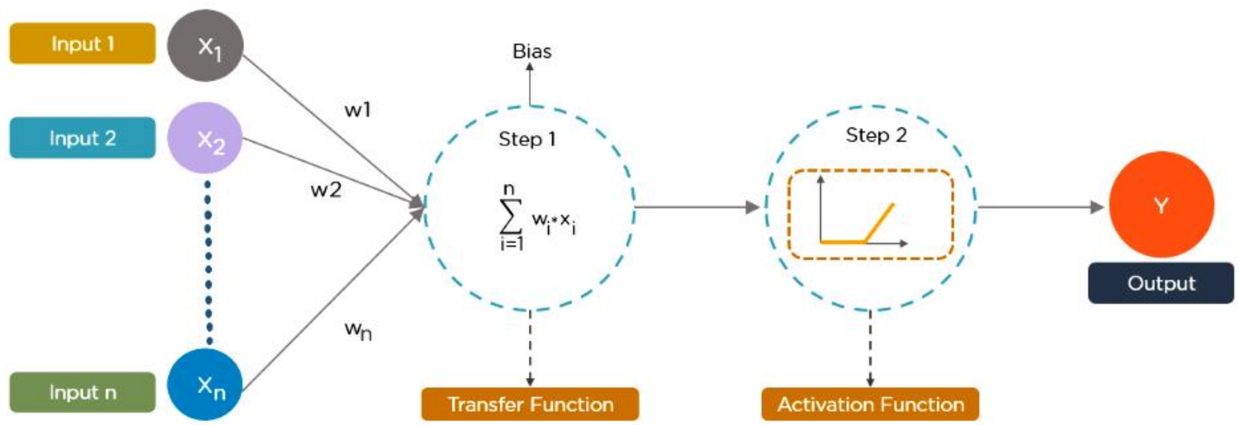


Figure 3: A typical neural network

Data provides each node with information in the form of inputs. The node multiplies the inputs with random weights, calculates them, and adds a bias. Finally, nonlinear functions, also known as activation functions, are applied to determine which neuron to fire.

1.5.6 Time Series Forecasting

A Time Series is defined as a series of data points indexed in time order. The time order can be daily, monthly, or even yearly.Time Series forecasting is the process of using a statistical model to predict future values of a time series based on past results.Example-Time Series that illustrates the number of passengers of an airline per month from the year 1949 to 1960.

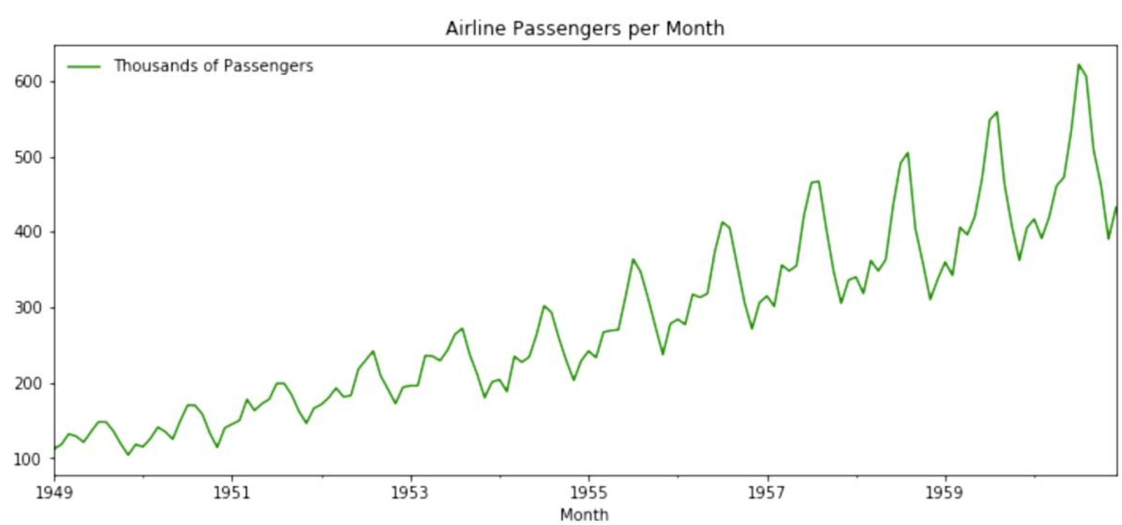


Figure 4: Time Series that illustrates the number of passengers of an airline per month from the year 1949 to 1960.

A. Components of Time series:

Trend

The trend shows a general direction of the time series data over a long period of time. A trend can be increasing(upward), decreasing(downward), or horizontal(stationary).

Seasonality

The seasonality component exhibits a trend that repeats with respect to timing, direction, and magnitude. Some examples include an increase in water consumption in summer due to hot weather conditions, or an increase in the number of airline passengers during holidays each year.

Cyclical Component

These are the trends with no set repetition over a particular period of time. A cycle refers to the period of ups and downs, booms and slums of a time series, mostly observed in business cycles. These cycles do not exhibit a seasonal variation but generally occur over a time period of 3 to 12 years depending on the nature of the time series.

Irregular Variation

These are the fluctuations in the time series data which become evident when trend and cyclical variations are removed. These variations are unpredictable, erratic, and may or may not be random.

B. Application

To explaining seasonal patterns in sales, To detect unusual events and estimate the magnitude of their effect, To Estimate the effect of a newly launched product on number of sold units, To predict the number of incoming or churning customers etc.

1.5.7 Gradient Descent

Gradient Descent is a generic optimization algorithm capable of finding optimal solutions to a wide range of problems. The general idea is to tweak parameters iteratively in order to minimize the cost function.An important parameter of Gradient Descent (GD) is the size of the steps, determined by the learning rate hyper-parameters. If the learning rate is too small, then the algorithm will have to go through many iterations to converge, which will take a long time, and if it is too high we may jump the optimal value. It is applicable in the scenarios where the function is easily differentiable with respect to the parameters used in the network. It is easy to minimize continuous functions than minimizing discrete functions. The weight update is performed after one epoch, where one epoch represents running through an entire data set. This technique produces satisfactory results but it deteriorates if the training data set size becomes large and does not converge well. It also may not lead to a global minimum in case of the existence of multiple local minima.Typically, there are three types of Gradient Descent: Batch Gradient Descent, Stochastic Gradient Descent, Mini-batch Gradient Descent. In this study Stochastic Gradient Descent is described as it is being used for forecasting of drought in the context of this project.

1.5.8 Stochastic Gradient Descent

Stochastic Gradient Descent (SGD) is a variant of the Gradient Descent algorithm used for optimizing machine learning models. In this variant, only one random training example is used to calculate the gradient and update the parameters at each iteration.It overcomes this drawback by randomly selecting data samples and updating the parameters based on the cost function. Additionally, it converges faster than regular gradient descent and saves memory by not accumulating the intermediate weights. It has the ability to escape from local minima and converge to a global minimum.

1.5.9 Adam optimizer

Adaptive Moment Estimation is an algorithm for optimization technique for gradient descent. The method is really efficient when working with large problem involving a lot of data or parameters. It requires less memory and is efficient. Intuitively, it is a combination of the 'gradient descent with momentum' algorithm and the 'RMSP' algorithm.

1.5.10 Stationarity & Augmented Dickey- fuller Test:

Stationarity

The Time series data model works on stationary data. The Stationarity of data is described by the following three criteria: It should have a constant mean(average value of all the data), It should have a constant variance(difference of each point value from the mean), Auto covariance (it is a relationship between any two values at a certain amount of time) does not depend on the time.

Augmented Dickey- fuller Test

In this method, we take a null hypothesis that the data is non-stationary. After executing this test, it will give some results comprised of test statistics and some other critical values that help to define the Stationarity. If the test statistic is less than the critical value then we can reject the null hypothesis and say that the series is stationary.

1.5.11 Activation function

The activation function decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. The purpose of the activation function is to introduce non-linearity into the output of a neuron. The neural network has neurons that work in correspondence with weight, bias, and their respective activation function. In a neural network, we would update the weights and biases of the neurons on the basis of the error at the output. This process is known as back-propagation. Activation functions make the back-propagation possible since the gradients are supplied along with the error to update the weights and biases. A neural network without an activation function is essentially just a linear regression model. The activation function does the non-linear transformation to the input making it capable to learn and perform more complex tasks.

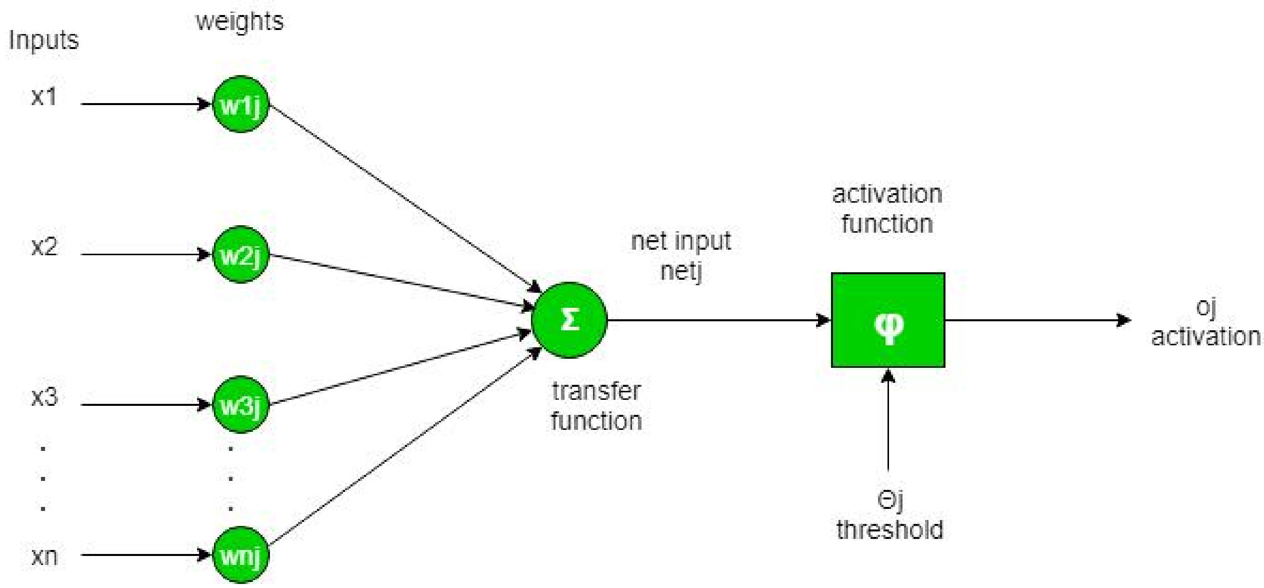


Figure 4: Activation function in the neural network

$$\text{net input} = \sum(\text{weight} * \text{input}) + \text{bias}$$

1.5.12 Types of Activation Functions

A. Step Function

Step Function is one of the simplest kind of activation functions. In this, we consider a threshold value and if the value of net input say y is greater than the threshold then the neuron is activated.

Mathematically,

$$f(x) = 1, \text{if } x >= 0, f(x) = 0, \text{if } x < 0$$

Graphically,

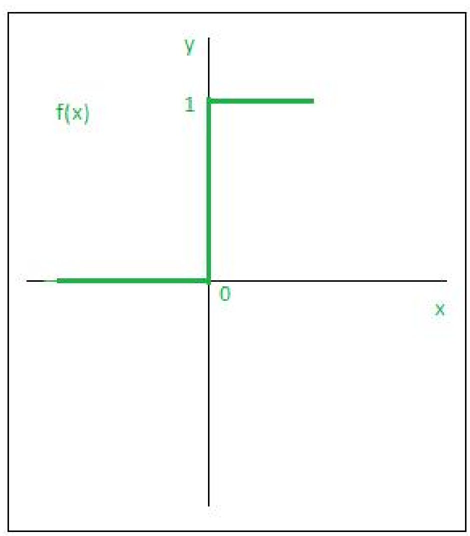


Figure 5: Step Function

B. Sigmoid Function

Sigmoid function is a widely used activation function. It is defined as:

$$f(x) = \frac{1}{1 + e^{-x}}$$

This is a smooth function and is continuously differentiable. The biggest advantage that it has over step and linear function is that it is non-linear.This essentially means that when I have multiple neurons having sigmoid function as their activation function – the output is non linear as well. The function ranges from 0-1 having an S shape.

Graphically,

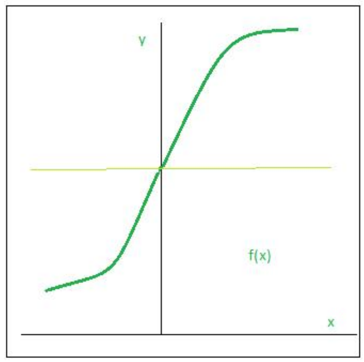


Figure 6: Sigmoid Function

C. tanh Function

The activation that works almost always better than sigmoid function is tanh function also known as Tangent Hyperbolic function. It's actually mathematically shifted version of the sigmoid function.

Mathematically, $f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$

Graphically,

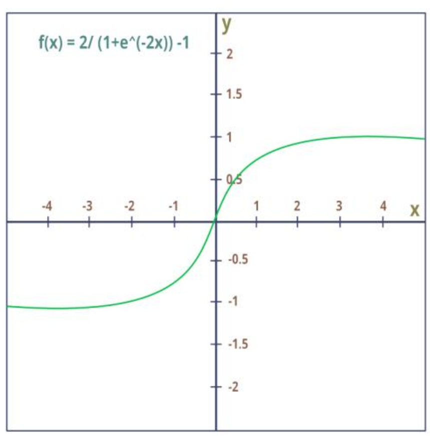


Figure 7: Tanh Function

D. ReLU Function: The ReLU function is the Rectified linear unit. It is the most widely used activation function. It is defined as:

$$f(x) = \max(0, x)$$

Graphically,

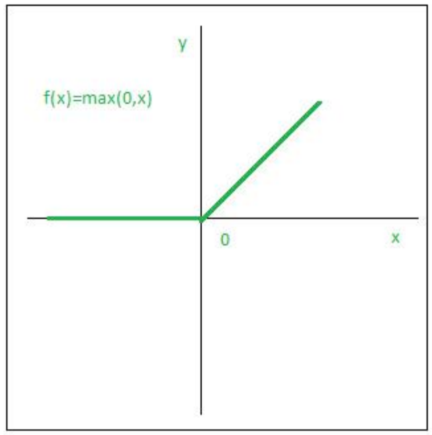


Figure 8: ReLU Function

The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time.

E. Leaky ReLU Function: Leaky ReLU function is nothing but an improved version of the ReLU function.Instead of defining the Relu function as 0 for x less than 0, we define it as a small linear component of x. It can be defined as:

$$f(x) = ax \text{ if } x < 0, \quad f(x) = 0, \text{otherwise}$$

Graphically,

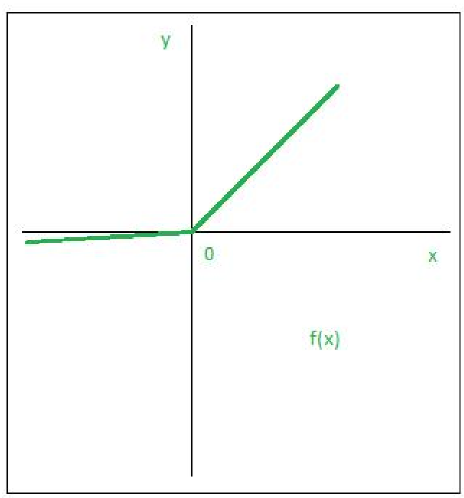


Figure 9: Leaky ReLU Function

CHAPTER 2

2. Literature Review

2.1 Literature Review on SPI

McKee et al. (1993) stated that SPI is based on the long-term precipitation record for a desired period, is used as drought assessment tool. It is an index to categorize the severity of a drought due to rainfall deficit locally or globally. The SPI was used in this study because it is based on rainfall data alone, so the drought assessment is possible where even other hydro meteorological measurements are not available. It is calculated in the following sequence. A monthly precipitation data set is prepared for a period of m months, ideally a continuous period of at least 30 years. A set of averaging periods are selected to determine a set of time scales of period 3, 6, 12, 24, or 48 months.

Guttman et al. (1999) stated that the Palmer Drought Severity Index (PDSI) has been calculated for about 30 years as a means of providing a single measure of meteorological drought severity. It was intended to retrospectively look at wet and dry conditions using water balance techniques. The Standardized Precipitation Index (SPI) is a probability index that was developed to give a better representation of abnormal wetness and dryness than the Palmer indices.

Mishra et al. (2006) stated that SPI has a variable time scale and is thus conducive to monitor short-term water supplies, such as soil moisture which is important for agricultural production, and long-term water resources, such as groundwater supplies, stream flow, and lake and reservoir levels.

Hayes et al. (2012) stated that the SPI is a powerful, flexible index that is simple to calculate. In fact, precipitation is the only required input parameter. In addition, it is just as effective in analyzing wet periods/cycles as it is in analyzing dry periods/cycles. It is based on the probability of precipitation for any time scale. The probability of observed precipitation is then transformed into an index. It is being used in research or operational mode in more than 70 countries.

2.2 Literature Review on Machine Learning

Belayneh et al.(2013) stated that the data-driven modelling for forecasting the meteorological time series prediction is becoming more powerful and flexible with computational intelligence techniques. Machine learning (ML) techniques have demonstrated success in the drought prediction process and are becoming popular to predict the weather, especially the minimum temperature using back propagation algorithms. The favourite ML techniques for weather forecasting include singular vector machines (SVM), support vector regression, random forest, decision tree, logistic regression, Naive Bayes, linear regression, gradient boosting tree, k-nearest neighbour (KNN), the adaptive neuron-fuzzy inference system, the feed-forward neural networks, Markov chain, Bayesian network, hidden Markov models, and auto-regressive moving averages, evolutionary algorithms, deep learning etc.

Dikshit et al. (2022) stated that since drought is nonlinear and multivariate in nature, the ability of neural networks to easily and efficiently capture dynamic relationships is increasingly being utilized.

CHAPTER 3

3. Study Area & Data

3.1 Study area

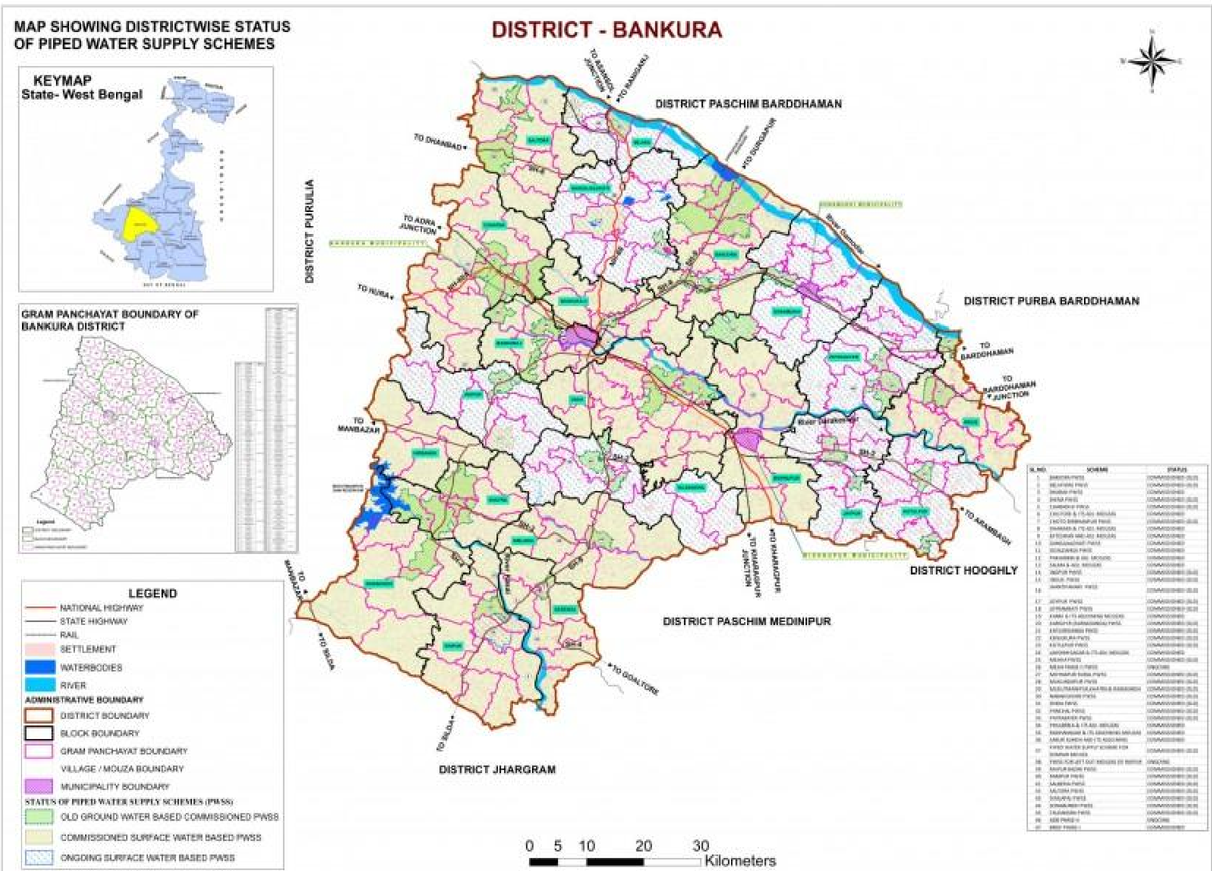


Figure 10: Bankura District, West Bengal, India

Bankura district is an administrative unit in the Indian state of West Bengal. It is part of Medinipur division—one of the five administrative divisions of West Bengal. Bankura district is surrounded by Purba Bardhaman district and Paschim Bardhaman district in the north, Purulia district in the west, Jhargram district and Paschim Medinipur district in the south, and some part of Hooghly district in the east. Damodar River flows in the northern part of Bankura district and separates it with the major part of Burdwan district.The district head quarter is located in Bankura town. The district has been described as the connecting link between the plains of Bengal on the east and Chotanagpur plateau on the west. The areas to the east and north-east are low-lying alluvial plains while to the west the surface gradually rises, giving way to undulating country, interspersed with rocky hillocks.

A. Geography

The areas to the east and north-east are low-lying alluvial plains, similar to predominating rice lands of Bengal. To the west the surface gradually rises, giving way to undulating country, interspersed with rocky hillocks. Much of the country is covered with jungles. The western part of the district has poor, ferruginous soil and hard beds of laterite with scrub jungles and Sal woods. Long broken ridges with irregular patches of more recent alluvium have marks of seasonal cultivation. During the long dry season large extents of red soil with hardly any trees lend the country a scorched and dreary appearance. In the eastern part the eye constantly rests on wide expanses of rice fields, green in the rains but parched and dry in summer.

B. Hills

The hills of the district consist of outliers of the Chotanagpur plateau and only two are of any great height – Biharinath and Susunia**.**

C. Rivers

The rivers of the area flow from the north-east to the south-west in courses roughly parallel to one another. They are mostly hill streams, originating in the hills in the west. The rivers come down in floods after heavy rains and subside as rapidly as they rise. In summer, their sand beds are almost always dry. The principal rivers are: Damodar, Dwarakeswar, Shilabati, Kangsabati, Sali, Gandheswari, Kukhra, Birai, Jaypanda and Bhairabbanki. There are some small waterfalls along the course of the Shilabati near Harmasra, and along the course of the Kangsabati in the Raipur area.

D. Geology

The greater portion of the district consists of a rolling country covered by laterite and alluvium. While metamorphic or gneissose rocks are found to the extreme west, to the east there is a wide plain of recent alluvium. Strong massive runs of hornblendic varieties stretch across the region in tolerably continuous lines, the general strike being nearly east and west. The most characteristic geological feature of the district is the area of laterite and associated rocks of sand and gravel. At some places one finds hard beds of laterite. At other places it is decomposed and reorganized. Locally, the ferruginous rock is called kankar. The calcareous concretions, commonly used as the sources of lime, are known as ghutin.

E. Climate

The climate, especially in the upland tracts to the west, is much drier than in eastern or southern Bengal. From the beginning of March to early June, hot westerly winds prevail, the thermometer in the shade rising to around 45 °C (113 °F). The monsoon months, June to September, are comparatively pleasant. The total average rainfall is 1,400 millimeters (55 in), the bulk of the rain coming in the months of June to September. Winters are pleasant with temperatures dropping down to below 27 °C (81 °F) in December.

F. Droughts in Bankura

Bankura though being a rain fed district, it is widely known as the drought prone district of the State. Drought is a regular feature in the North-West part of the district covering Chhatna, Saltora, Gangajalghati, Barjora, Bankura-I, Bankura-II, Mejia, Indpur, Hirbandh & Ranibandh Blocks. Though this district receives good amount of rainfall, around 1400 mm. annually, is received per year yet cultivation and production of crop primarily depends on constricted period of erratic rainfall. About 80% to 90% rainfall is generally received by the district from June to September depending on the onset of monsoon.A conspicuous feature of this district is the absence of significant rainfall in the month of September and October.

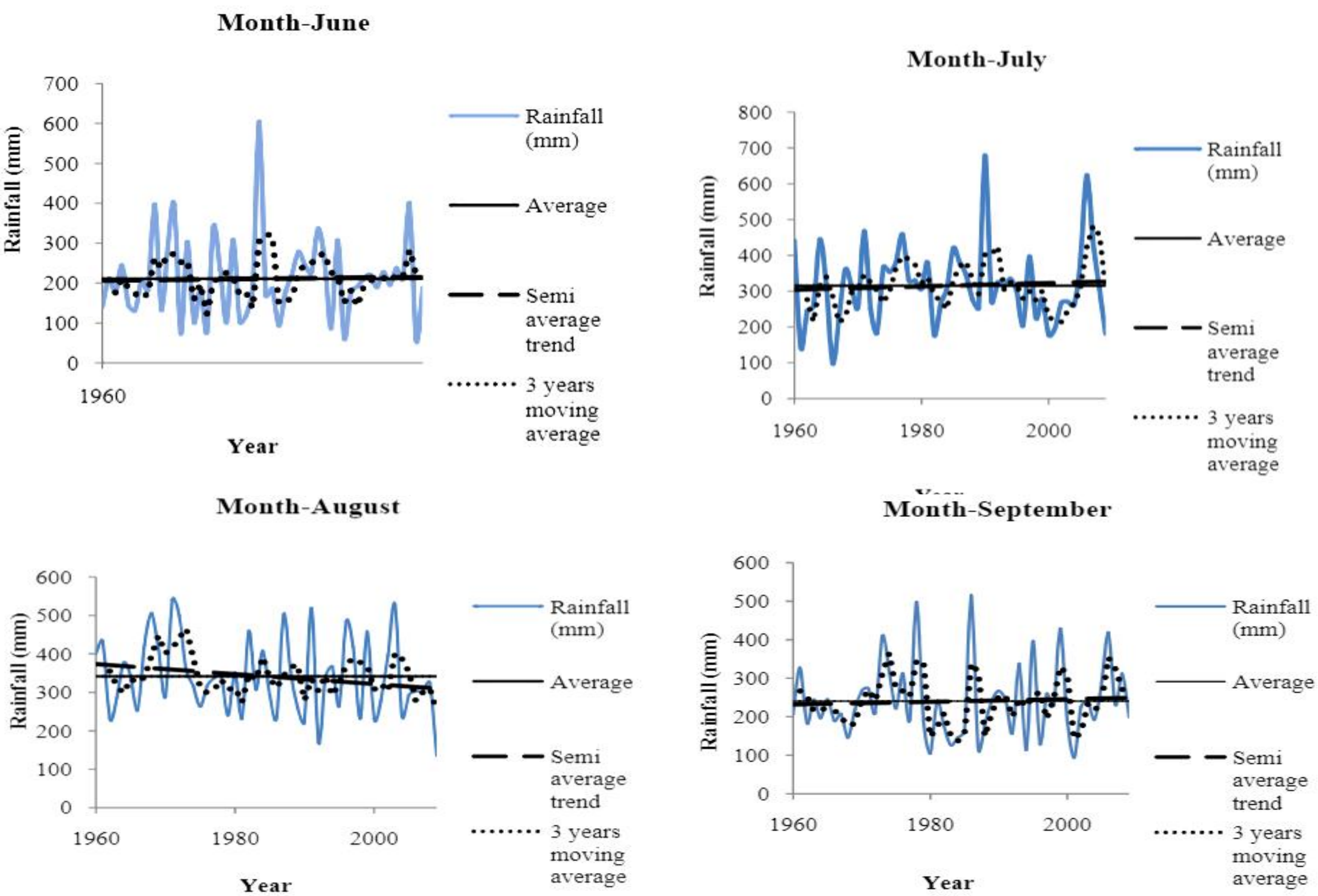


Figure 11: Rainfall data of Bankura from 1960 t0 2009 in the month of June, July, August, September from IMD Website.

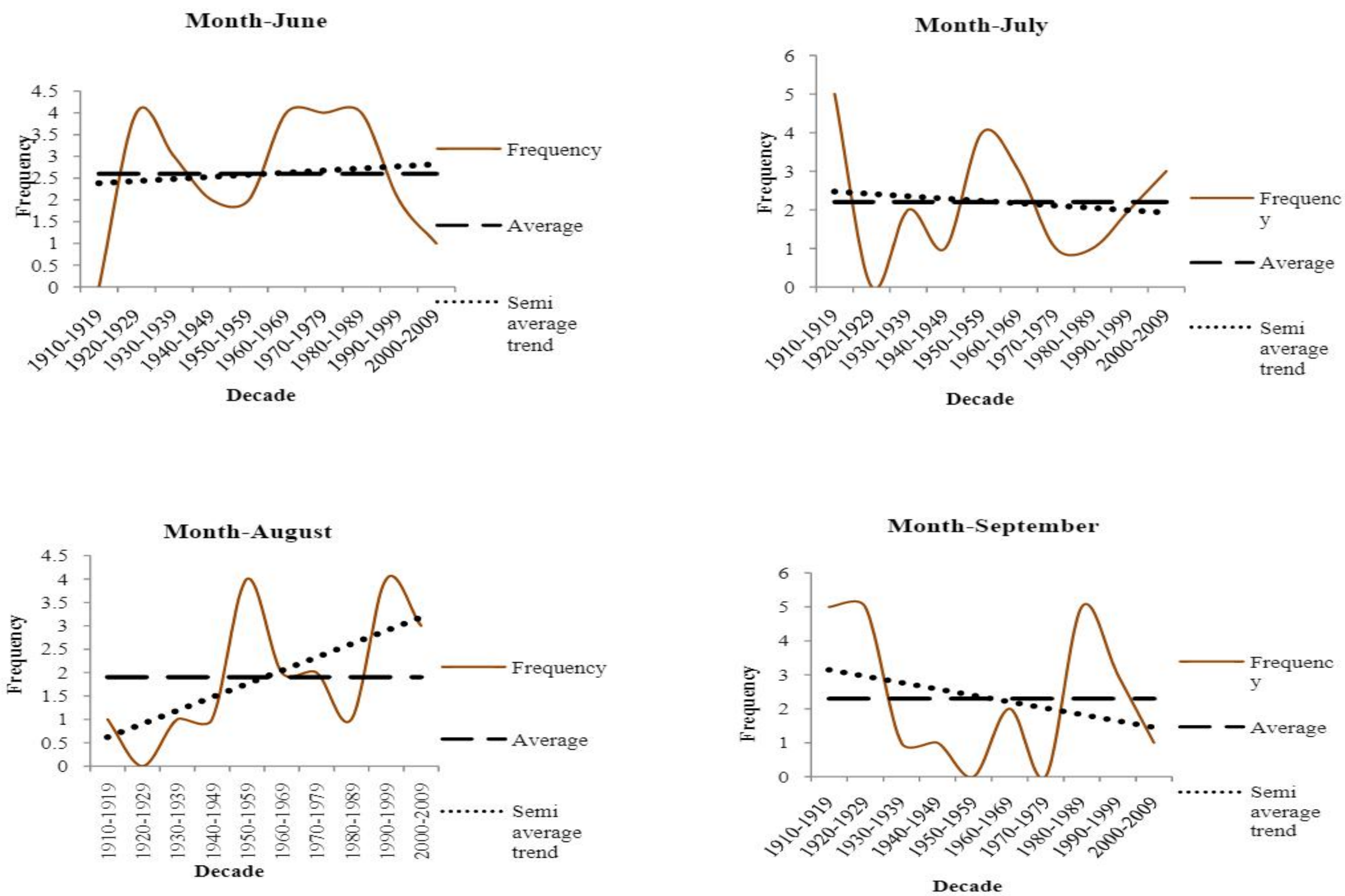


Figure 12: Decade trend of meteorological drought in Bankura from 1910-2009 in the month of June, July, August, September.

3.2 Data collection

Rainfall and temperature data for analysis and forecasting of drought is collected from the Bankura district in the from the year 1981 to 2020 at 23.75 degree North and 87.75 degree East for every month and each year from Indian Meteorological Department(IMD) Website.

CHAPTER 4

4. Methodology

4.1 Standardized Precipitation Index (SPI)

The statistical distribution of rainfall accumulations for defined intervals (e.g., daily; 1 month; 3 months) over a long time-series (at least 30 years), can be effectively represented by the two-parameter continuous probability distribution known as the "gamma distribution". In order to compute the SPI for an observed rainfall accumulation for a period of interest (e.g., 1,3,6,12 or 48 months), the two parameters (i.e., shape and scale) of the gamma distribution are first fitted on the frequency distribution of the historical non-zero rainfall accumulations for all years in the available time-series, using one of two alternative approximations of the "maximum likelihood estimators" for the gamma distribution that were developed by Thom (1958) and Greenwood and Durand (1960). For any observed rainfall accumulation, the cumulative probability is then derived, based on the parameters of the gamma distribution and using algorithms provided by Press et al. (1992). Following adjustment for the probability of zero rainfall accumulation, the cumulative probability of the observed rainfall is then transformed (converted) to the standard normal random variable Z with mean zero and variance of one, using an approximation described by Abramowitz and Stegun (1965). This transformed value is the SPI.

4.1.1 Normalization of Data

The rainfall record of a long-term time (any base time may be chosen, depending on the time scale of interest) is fitted with a probability distribution (generally by Gamma distribution as it fits the rainfall time series well) after fitting the probability distribution, the cumulative probability of an observed rainfall is computed. Gamma probability distribution function (PDF) is calculated by,

$$g(x) = \frac{1}{(\beta^{\alpha} \cdot \Gamma(\alpha))} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

Where α and β are shape and scale parameters, x is the rainfall and $\Gamma(\alpha)$ is the gamma function. The aim of fitting the distribution to the data is to calculate α and β . Integrating PDF and inserting the estimated values of the parameters α and β , the gamma cumulative distribution function (CDF) can be calculated for each value of x.

4.1.2 Computation of SPI

The CDF is now transformed into the standard normal distribution or we can say that the inverse normal Gaussian function, with mean zero and variance one, is then applied to the cumulative probability distribution function, which results in the SPI based on probability of precipitation for any time scale and calculated as,

$$SPI = \frac{(X - Xm)}{\sigma}$$

Where, X=Precipitation for the station, Xm= Mean Precipitation, σ = Standard Deviation

The SPI calculation for any location is based on the long-term precipitation record for a desired period. This long-term record is fitted to a probability distribution, which is then transformed into a normal distribution so that the mean SPI for the location and desired period is zero. Positive SPI values indicate greater than median precipitation, and negative values indicate less than median precipitation. Drought, according to the SPI, starts when the SPI value is equal or below -1.0 and ends when the value becomes positive. Computation of SPI requires precipitation which will be downloaded from the India Meteorological Department (IMD) website. The computed SPI time series will help to characterize meteorological drought viz. duration and severity of the event.

4.2 Mathematical aspects of Adam optimizer

Adam optimizer involves a combination of two gradient descent methodologies:Momentum, Root Mean Square Propagation (RMSP)

4.2.1 Momentum

This algorithm is used to accelerate the gradient descent algorithm by taking into consideration the 'exponentially weighted average' of the gradients. Using averages makes the algorithm converge towards the minima in a faster pace.

$$W_{t+1} = W_t - \alpha m_t$$

Where,

$$m_t = \beta m_{t-1} + (1 - \beta)\left[\frac{\delta L}{\delta \omega_t}\right]$$

m_t = aggregate of gradients at time t

m_{t-1}= aggregate of gradients at time t-1

W_t= weights at time t

W_{t+1}= weights at time t+1

α= learning rate at time t

δL= derivative of Loss Function

$\delta \omega_t$=derivative of weights at time t

β=Moving average parameter

4.2.2 Root Mean Square Propagation (RMSP)

Root mean square prop or RMSprop is an adaptive learning algorithm that tries to improve AdaGrad. Instead of taking the cumulative sum of squared gradients like in AdaGrad, it takes the 'exponential moving average'.

$$W_{t+1} = W_t - \frac{\alpha_t}{(V_t + \varepsilon)^{1/2}}\left[\frac{\delta L}{\delta \omega_t}\right]$$

where,

$$V_t = \beta V_t + (1-\beta)\left[\frac{\delta L}{\delta \omega_t}\right]^2$$

W_t= weights at time t

W_{t+1}=weights at time t+1

α_t= learning rate at time t

δL=derivative of Loss Function

$\delta\omega_t$=derivative of weights at time t

V_t=sum of square of past gradients. [i.e sum(∂L/∂Wt-1)]

β= Moving average parameter, ε= A small positive constant.

Adam Optimizer inherits the strengths or the positive attributes of the above two methods and builds upon them to give a more optimized gradient descent.

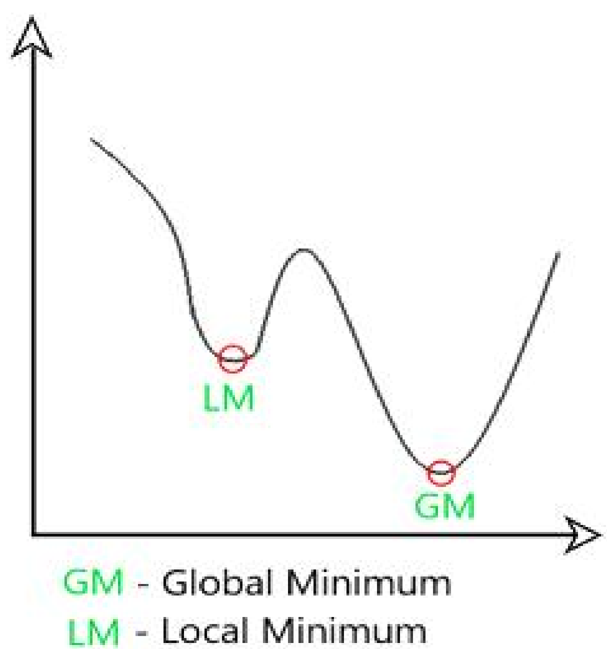


Figure 13: Local minima and Global minima in Adam optimizer

Here, we control the rate of gradient descent in such a way that there is minimum oscillation when it reaches the global minimum while taking big enough steps (step-size) so as to pass the local minima hurdles along the way. Hence, combining the features of the above methods to reach the global minimum efficiently.

Mathematical Aspect of Adam Optimizer:

Taking the formulas used in the above two methods, we get

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)[\frac{\delta L}{\delta\omega_t}] + \beta_2 V_t + (1-\beta_2)[\frac{\delta L}{\delta\omega_t}]^2$$

β_1& β_2 = decay rates of average of gradients in the above two methods.

4.3 Mathematical aspects of activation function in neural network

Suppose we have a Neural net like this.

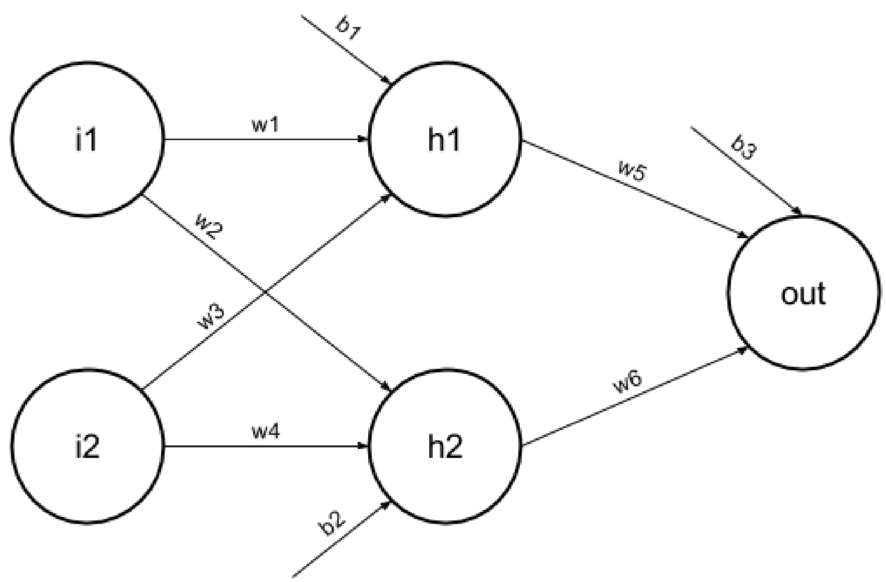


Figure 14: Calculation of activation function from neural network

Hidden layer i.e. layer 1

$z(1) = W(1)X + b(1)a(1)$

Here, z(1) is the vectorized output of layer 1

W(1) be the vectorized weights assigned to neurons of hidden layer i.e. w1, w2, w3 and w4

X be the vectorized input features i.e. i1 and i2.

b is the vectorized bias assigned to neurons in hidden layer i.e. b1 and b2.

a(1) is the vectorized form of any linear function.

We are not considering activation function here.

Layer 2 i.e. output layer

Input for layer 2 is output from layer 1.

$$z(2) = W(2)a(1) + b(2)$$
$$a(2) = z(2)$$

Calculation at Output layer:

$$z(2) = (W(2)[W(1)X + b(1)]) + b(2)$$
$$z(2) = [W(2)W(1)]X + [W(2)b(1) + b(2)]$$

Let, $[W(2)W(1)] = W, [W(2)b(1) + b(2)] = b$
Final Output:$z(2) = WX + b$

which is again a linear function

This observation results again in a linear function even after applying a hidden layer, hence we can conclude that, doesn't matter how many hidden layer we attach in neural net, all layers will behave same way because the composition of two linear function is a linear function itself. Neuron can not learn with just a linear function attached to it. A non-linear activation function will let it learn as per the difference w.r.t error. Hence we need an activation function.

4.4 Methodology of LSTM

LSTM has a chain structure that contains four neural networks and different memory blocks called cells.

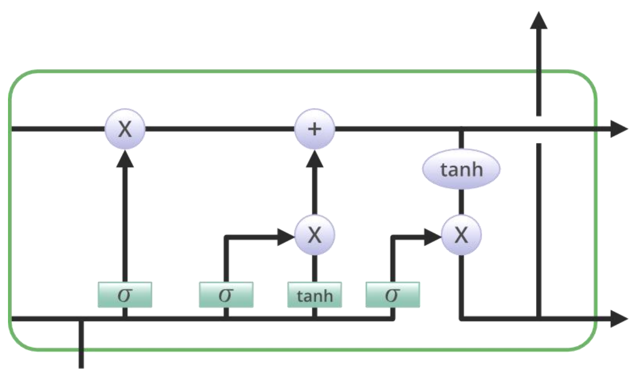


Figure 15: Structure of LSTM

Information is retained by the cells and the memory manipulations are done by the gates. There are three gates –

Forget gate

The information that is no longer useful in the cell state is removed with the forget gate. Two inputs x_t (input at the particular time) and h_t-1 (previous cell output) are fed to the gate and multiplied with weight matrices followed by the addition of bias. The resultant is passed through an activation function which gives a binary output. If for a particular cell state the output is 0, the piece of information is forgotten and for output 1, the information is retained for future use.

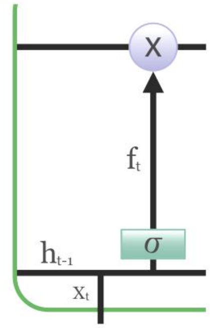


Figure 16: Forget gate

Input gate

The addition of useful information to the cell state is done by the input gate. First, the information is regulated using the sigmoid function and filter the values to be remembered similar to the forget gate using inputs h_t-1 and x_t. Then, a vector is created using tanh function that gives an output from -1 to +1, which contains all the possible values from h_t-1 and x_t. At last, the values of the vector and the regulated values are multiplied to obtain the useful information.

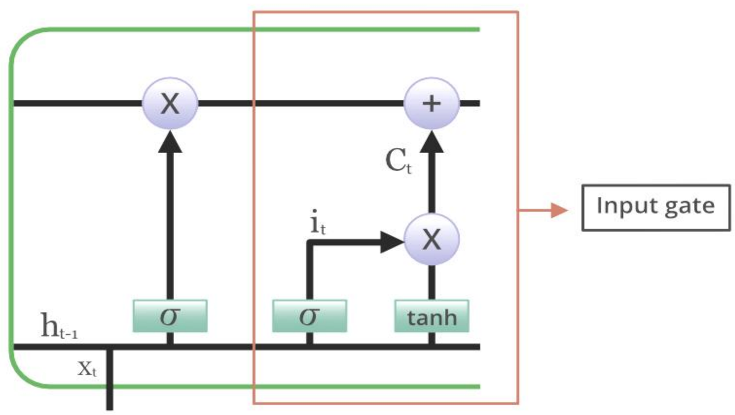


Figure 17: Input gate

Output gate

The task of extracting useful information from the current cell state to be presented as output is done by the output gate. First, a vector is generated by applying tanh function on the cell. Then, the information is regulated using the sigmoid function and filter by the values to be remembered using inputs h_t-1 and x_t. At last, the values of the vector and the regulated values are multiplied to be sent as an output and input to the next cell.

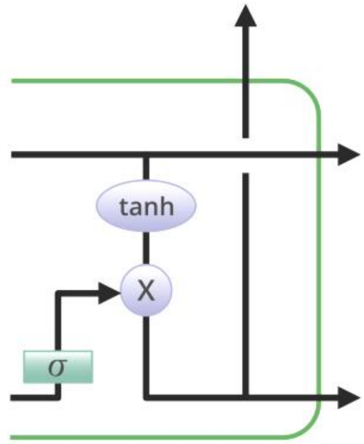


Figure 18:Output gate

First, they forget irrelevant parts of the previous state, then they selectively update the cell-state values, Finally, the output of certain parts of the cell state. Below is a diagram of how LSTMs operate:

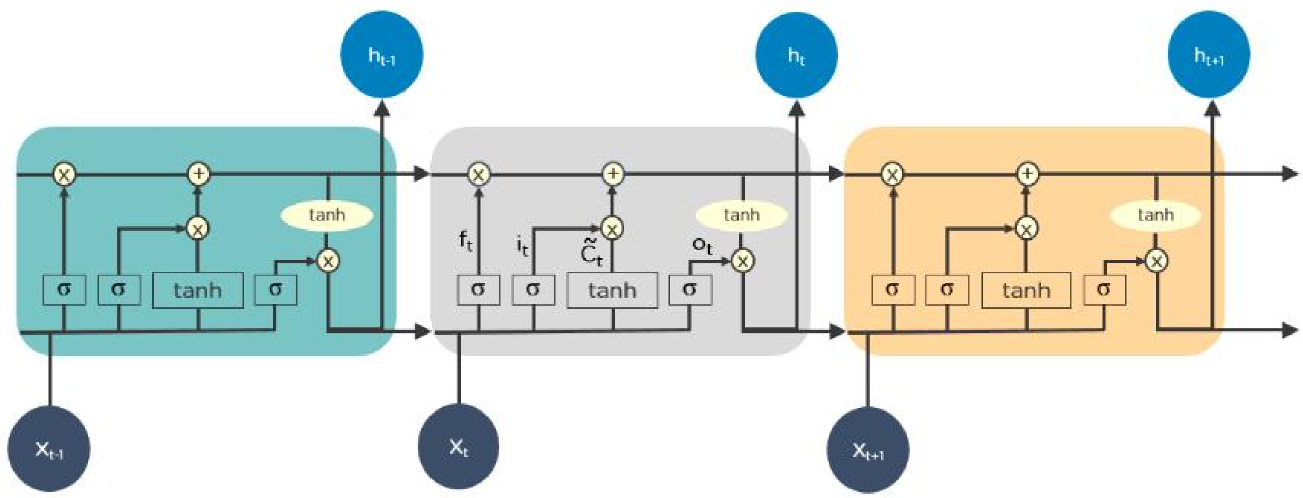


Figure 19: General framework of LSTM model

4.5 Mean Squared Error (MSE)

Mean Squared Error represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.

$$\mathrm{MSE} = \frac{\sum_{i=1}^{i=N}(\mathrm{Predicted}_i - \mathrm{actual}_i)}{N}$$

where, $\mathrm{Predicted}_i$ = The predicted value for the i_{th} observation, actual_i= The observed(actual) value for the i_{th} observation, N = Total number of observations.

4.6 Root Mean Squared Error (RMSE)

It is the square root of the mean of the square of all of the error. RMSE is considered an excellent general-purpose error metric for numerical predictions. RMSE is a good measure of accuracy, but only to compare prediction errors of different models or model configurations for a particular variable and not between variables, as it is scale-dependent. It is the measure of how well a regression line fits the data points. The formula for calculating RMSE is:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{i=N}(\mathrm{Predicted_i} - \mathrm{actual_i})}{N}}$$

where, $\mathrm{Predicted_i}$ = The predicted value for the i_{th} observation, $\mathrm{actual_i}$= The observed(actual) value for the i_{th} observation, N = Total number of observations.

4.7 Normalized root mean squared error (NRMSE)

The normalized root mean squared error (NRMSE), also called a scatter index, is a statistical error indicator defined as,

$$\mathrm{NRMSE} = \frac{\sum_{i=1}^{i=N}(\mathrm{Predicted_i} - \mathrm{actual_i})^2}{\sum_{i=1}^{i=N}(\mathrm{Predicted_i})^2}$$

where, $\mathrm{Predicted_i}$ = The predicted value for the i_{th} observation, $\mathrm{actual_i}$= The observed(actual) value for the i_{th} observation, N = Total number of observations.

4.8 R Squared

The coefficient of determination or R-squared represents the proportion of the variance in the dependent variable which is explained by the linear regression model. It is a scale-free score i.e. irrespective of the values being small or large, the value of R square will be less than one.

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{i=N}(\mathrm{Predicted_i} - \mathrm{actual_i})^2}{\sum_{i=1}^{i=N}(\mathrm{Predicted_i} - \mathrm{Mean})^2}$$

where, $\mathrm{Predicted_i}$ = The predicted value for the i_{th} observation, $\mathrm{actual_i}$= The observed(actual) value for the i_{th} observation, N = Total number of observations.

CHAPTER 5

5. Results and Discussions

5.1 Drought identification using SPI

First of all Rudimentary studies of the drought were carried out to understand the background of the problem. Extensive literature survey was carried out to define the methodology of the project. Computation of SPI for various time scales was done using R studio and SPEI package programme for data set. Plots of SPI for one months, three months, six months, nine months and twelve months accumulation period is attached here.

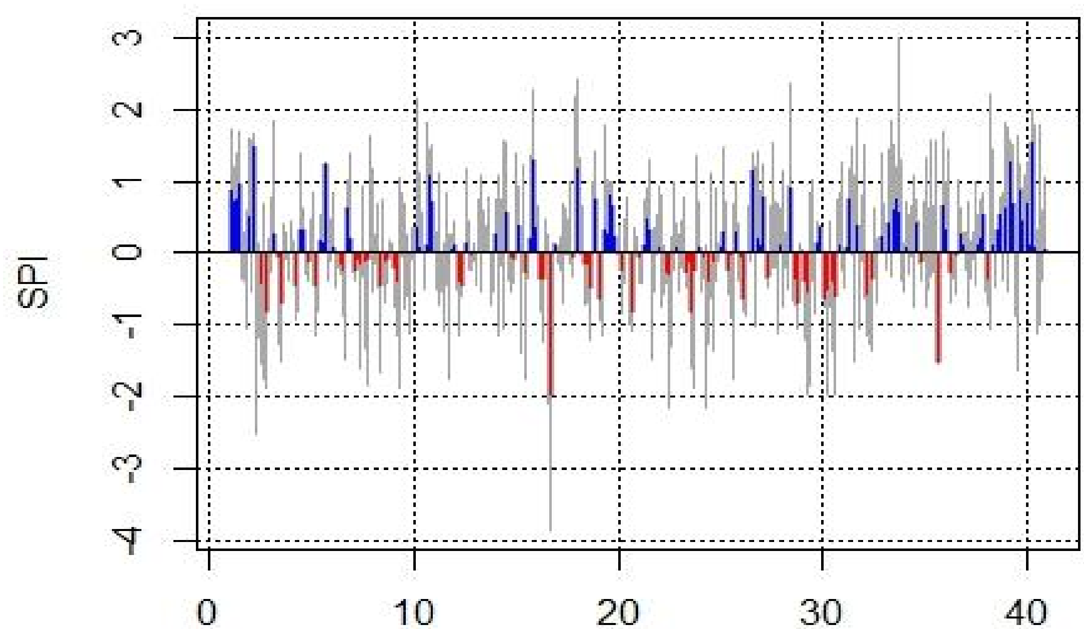


Figure 20: SPI-1(1 month SPI)

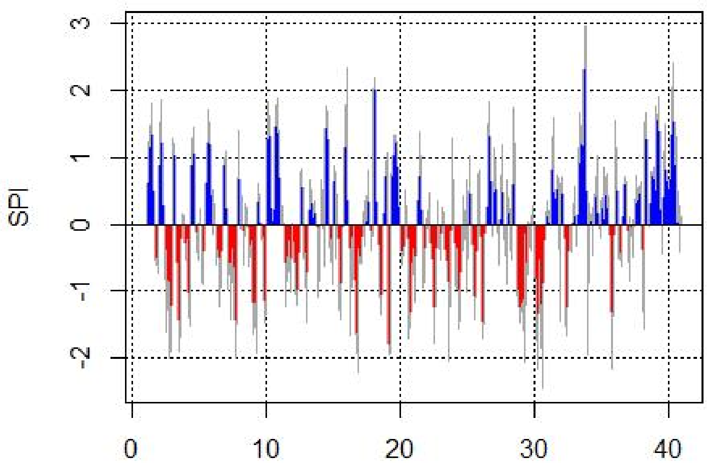


Figure 21: SPI-3(3 month SPI)

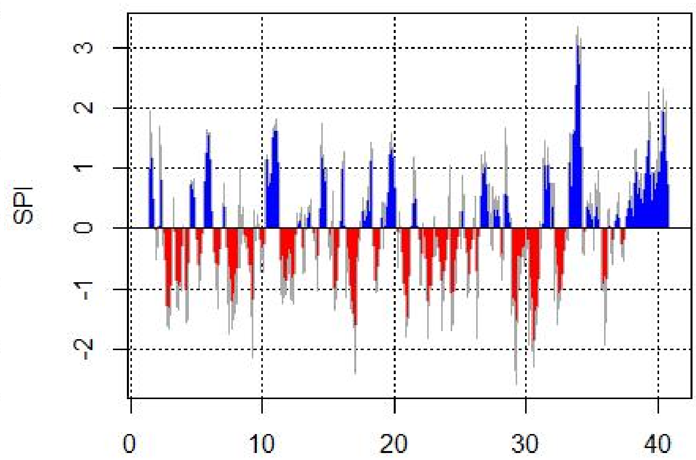


Figure 22:SPI-6(6 month SPI)

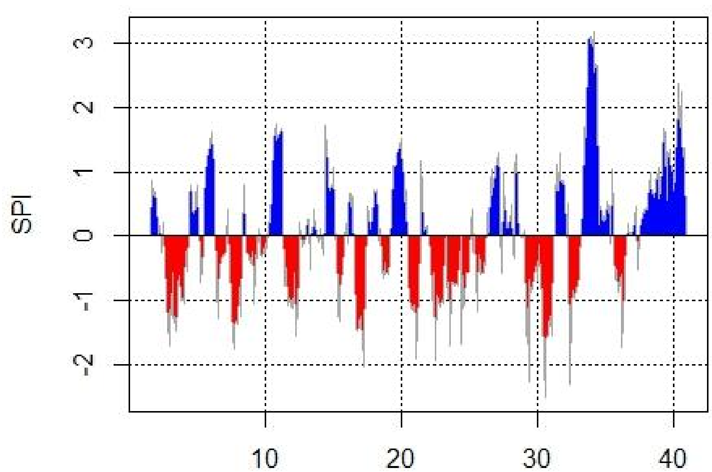


Figure 23:SPI-9(9 month SPI)

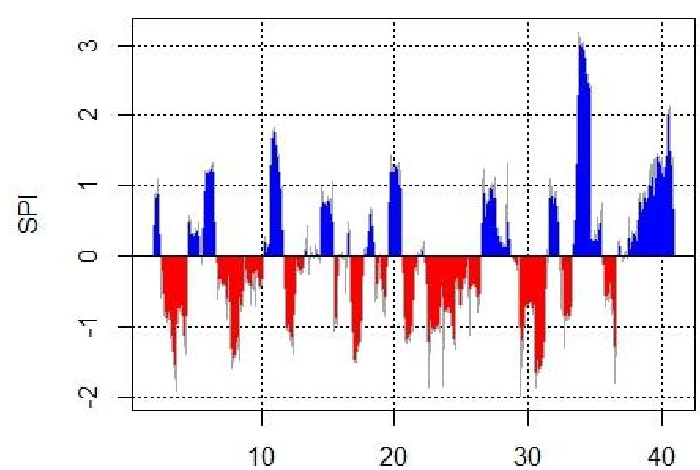


Figure 24:SPI-12(12 month SPI)

Also the SPI-3 Values of August and September from 1981 to 2020 is plotted and shown below.

For the month of August:(1981-2020)

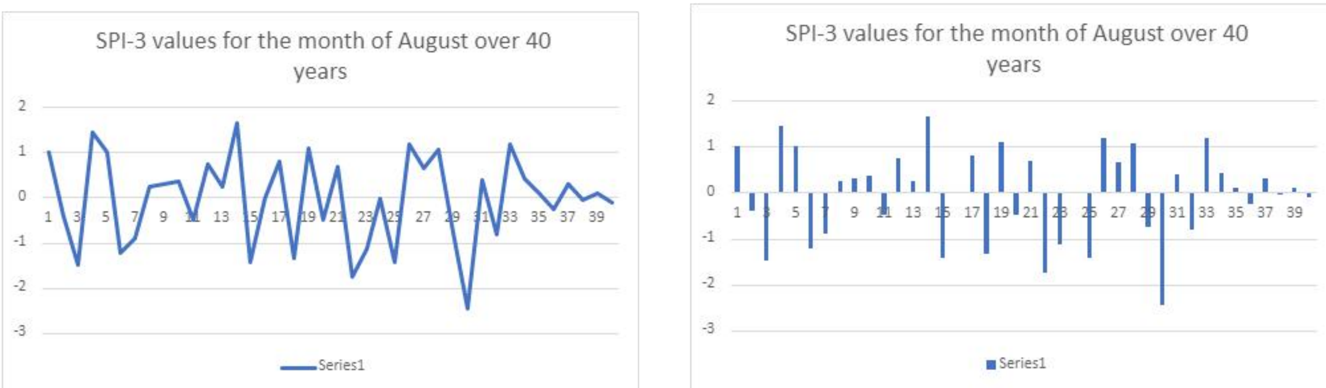


Figure 25: SPI-3 values for the month of August over 40 years

For the month of September:(1981-2020)

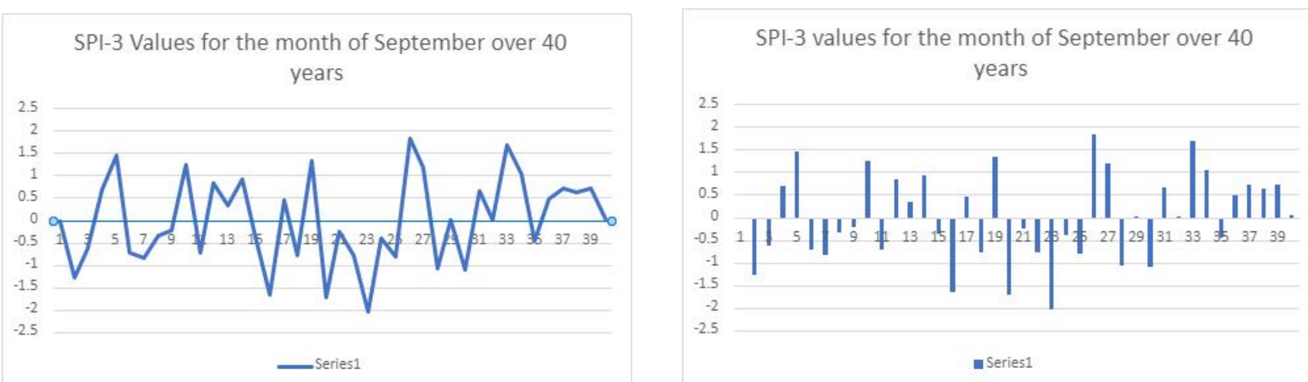


Figure 26: SPI-3 values for the month of September over 40 years

Discussion of SPI

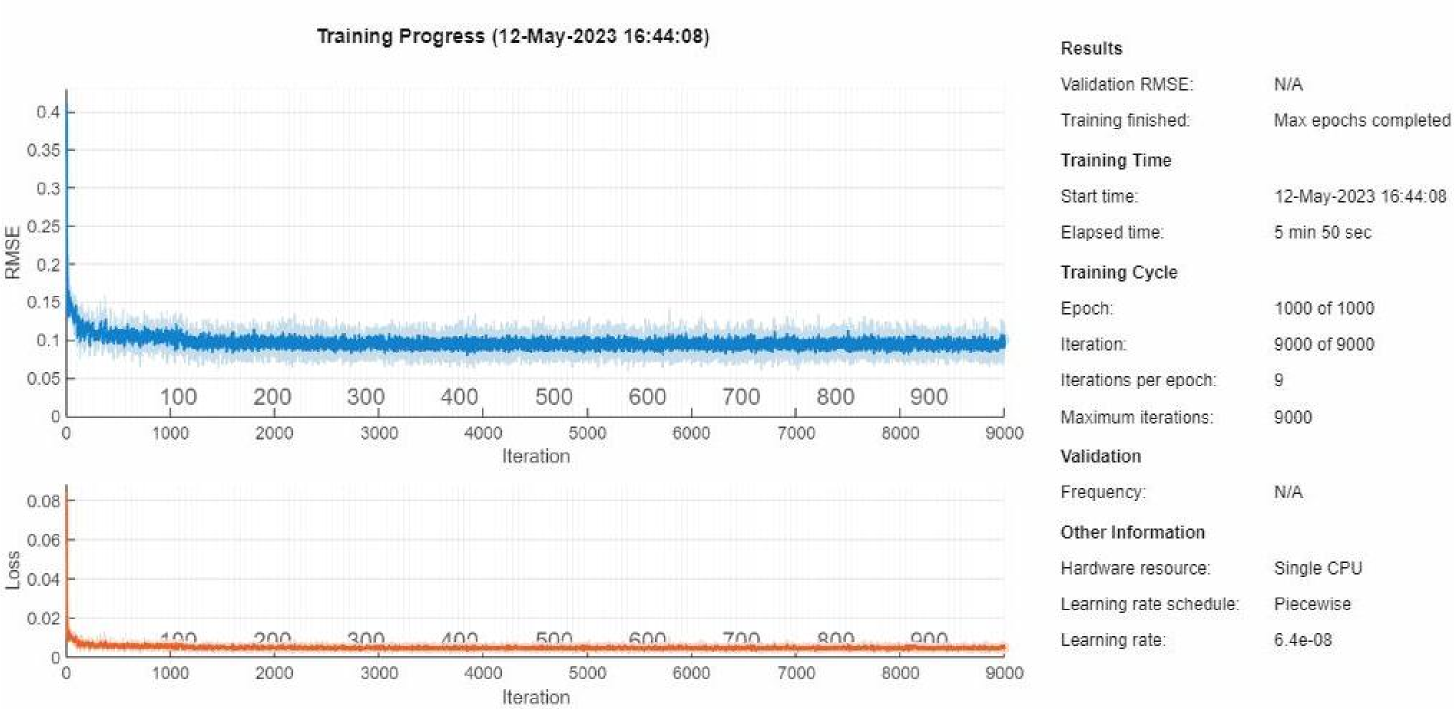
A. Smoothening effect could be observed for higher accumulation.

B. For the month of August in the year 1983,1986, 1995,1996,1998, 2002,2003,2005,2010 SPI-3 value is less than -1, so there is dry condition.

C. For the month of September in the year 1982,1996, 2000,2003, 2008, 2010 SPI-3 value is less than -1, so there is dry condition.

5.2 Forecasting of Meteorological Drought using LSTM model

5.2.1 Training progress and inputt data




```
opt.Delays = [1 2 3 4 5 6 7 8 9 10 12 16 20];
opt.dataPreprocessMode = 'Data Normalization';
opt.learningMethod = 'LSTM';
opt.trPercentage = 0.7; % divide data into Test and Train dataset

opt.NumOfHiddenLayers = 4; % number of (bi)LSTM layers

opt.NumOfUnitsInFirstlayer = 100; % number of (bi)LSTM units in the first layer
opt.NumOfUnitsInSecondlayer = 100; % number of (bi)LSTM units in the second layer
opt.NumOfUnitsInThirdlayer = 75; % number of (bi)LSTM units in the third layer
opt.NumOfUnitsInFourthlayer = 75; % number of (bi)LSTM units in the forth layer

opt.DropoutValue = 0.5;

opt.maxEpochs = 2000; % maximum number of training Epoch in bi-LSTM.
opt.miniBatchSize = 32; % minimum batch size in bi-LSTM .
opt.executionEnvironment = 'cpu'; % 'cpu' 'gpu' 'auto'
opt.LR = 'adam'; % 'sgdm' 'rmsprop' 'adam'
```

5.2.2 Regression graphs

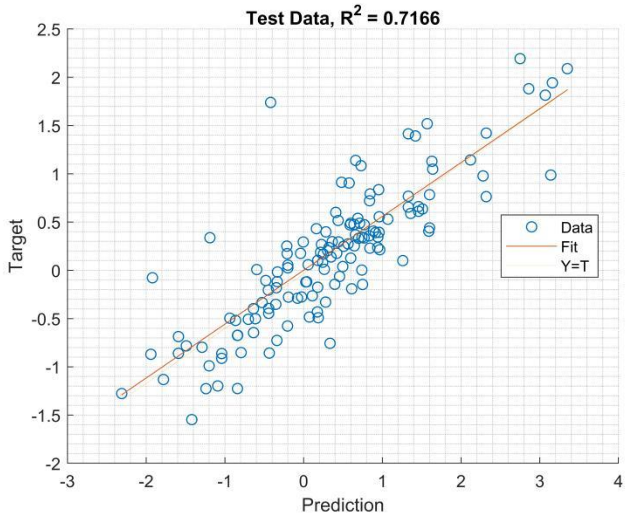


Figure 27: Regression graph evaluation Test Data

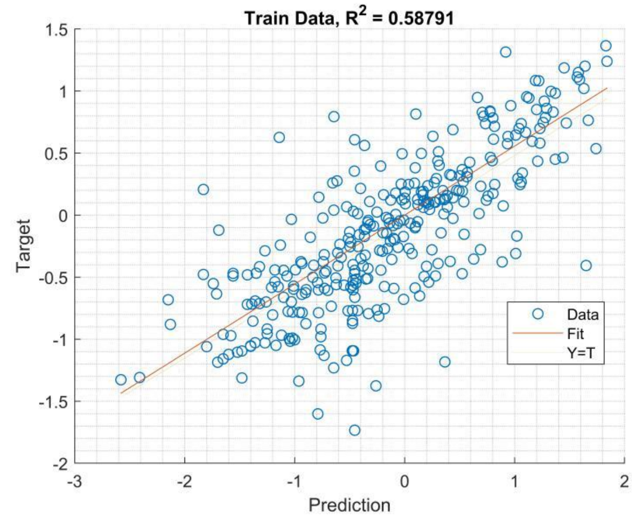


Figure 28: Regression graph evaluation Train Data

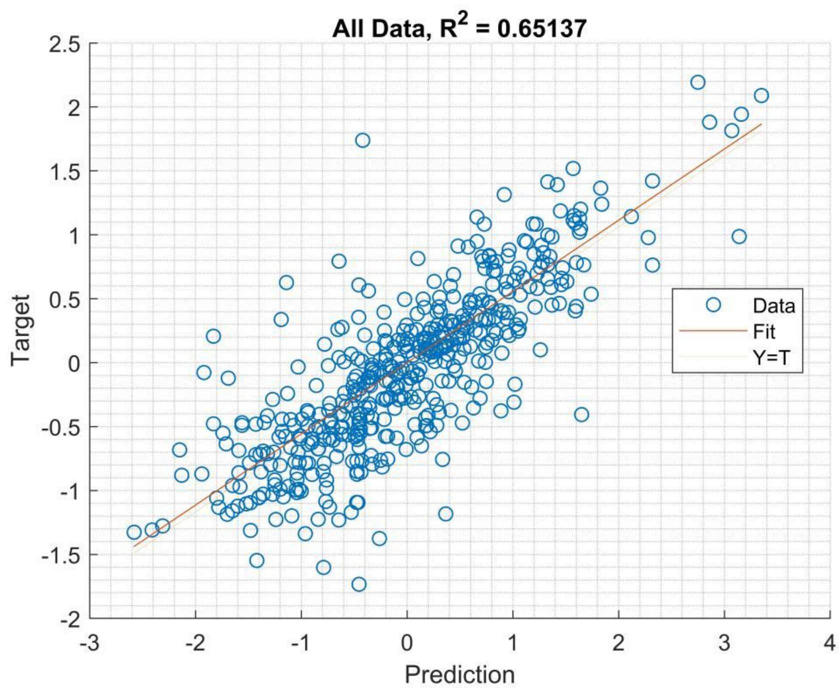


Figure 29: Regression graph evaluation All Data

Interpretation

R-squared is used for explaining the percentage of error in the data set. It is observed for both Testing and Training. It is a scale-free score i.e. irrespective of the values being small or large, the value of R square will be less than one. For our case we observe around **65.1%** data which could be explained

5.2.3 Error Evaluation Data

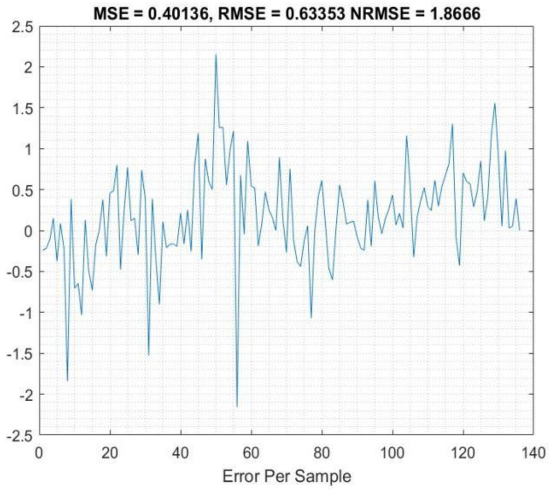


Figure 30: Error evaluation Test data

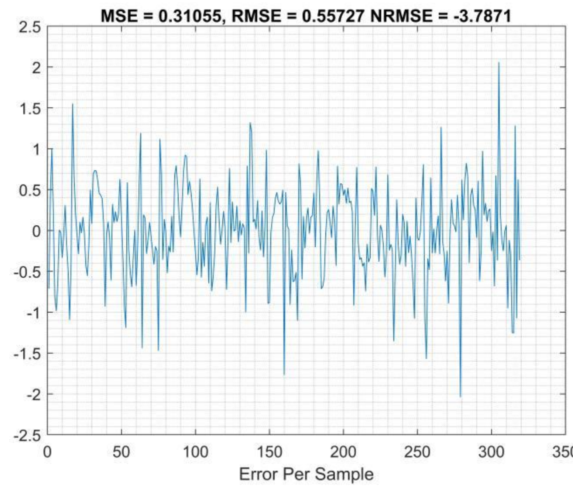


Figure 31: Error evaluation Training data

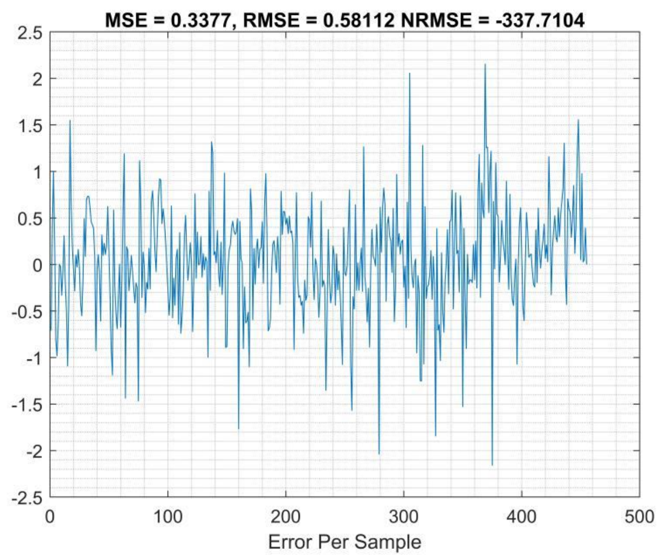


Figure 32: Error evaluation of All data

Interpretation

The error hovers around zero. For a few instances the error deviates beyond 2. The observed:

MSE value for Testing data= 0.40136, Training data= 0.31055, All data = 0.3377

RMSE value for Testing data= 0.63353, Training data= 0.55727, All data = 0.58112

NRMSE value for Testing data= 1.8666, Training data= -3.7871, All data = -337.7104

5.2.3 Error Histograms

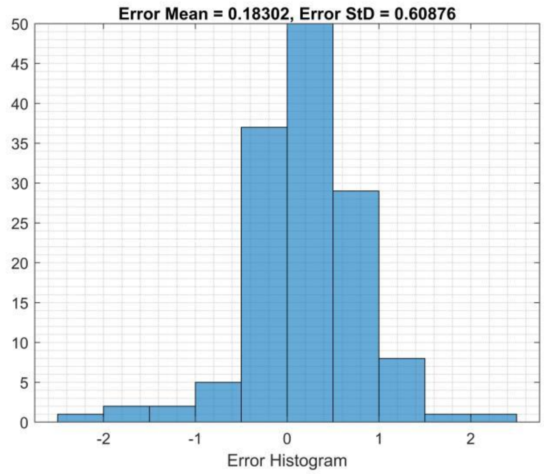


Figure 33: Error Histogram All Data

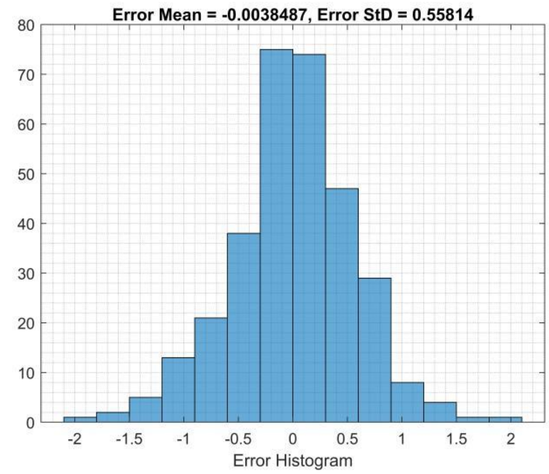


Figure 34: Error Histogram All Data

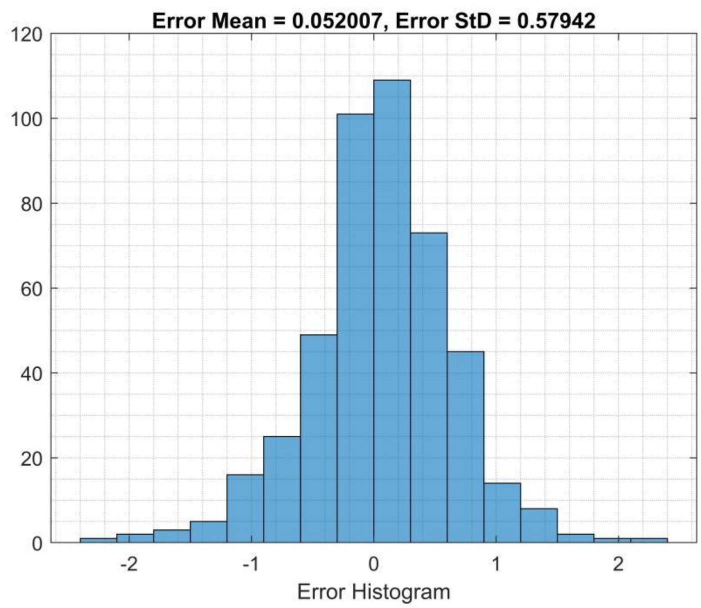


Figure 35: Error Histogram All Data

Interpretation

In case of error histogram, white noise distribution is observed about Zero mean. The error histograms shows that it is Normally distribution with mean=0.

5.2.4 Comparison of results between observed and predicted data

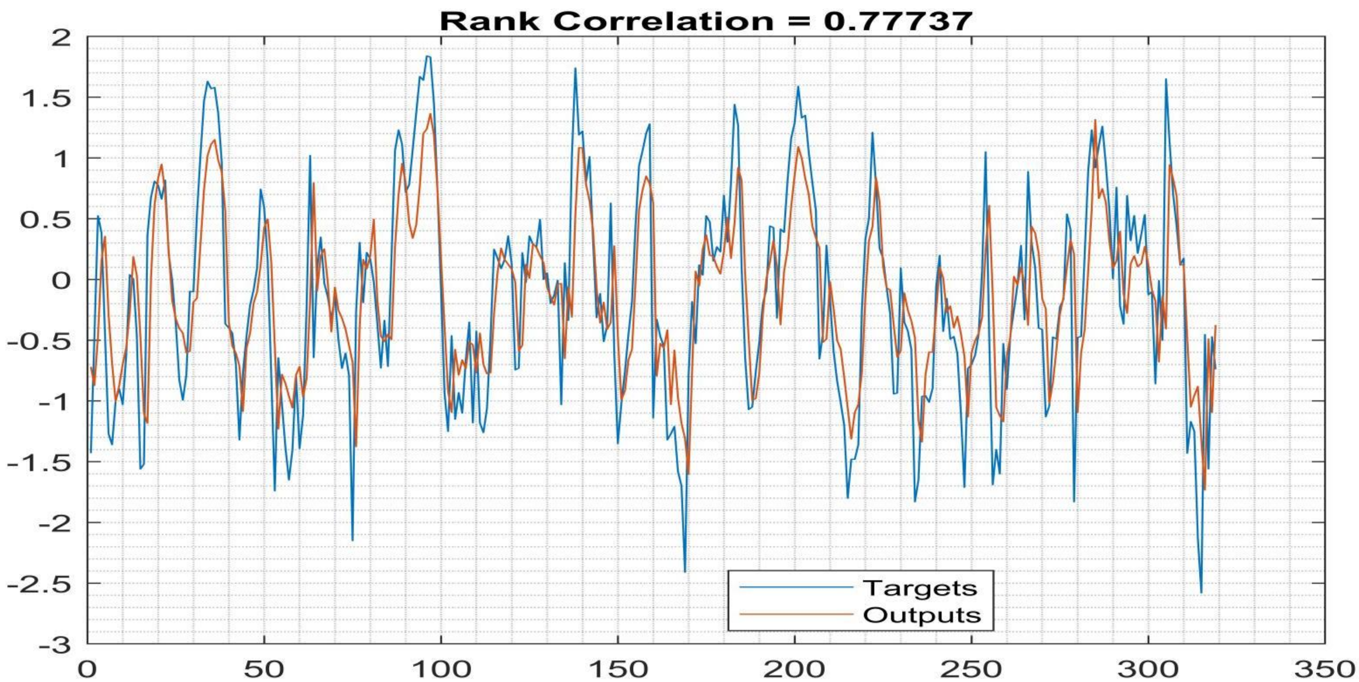


Figure 36: Comparison graph of observed(target) and predicted(outputs) for Training data

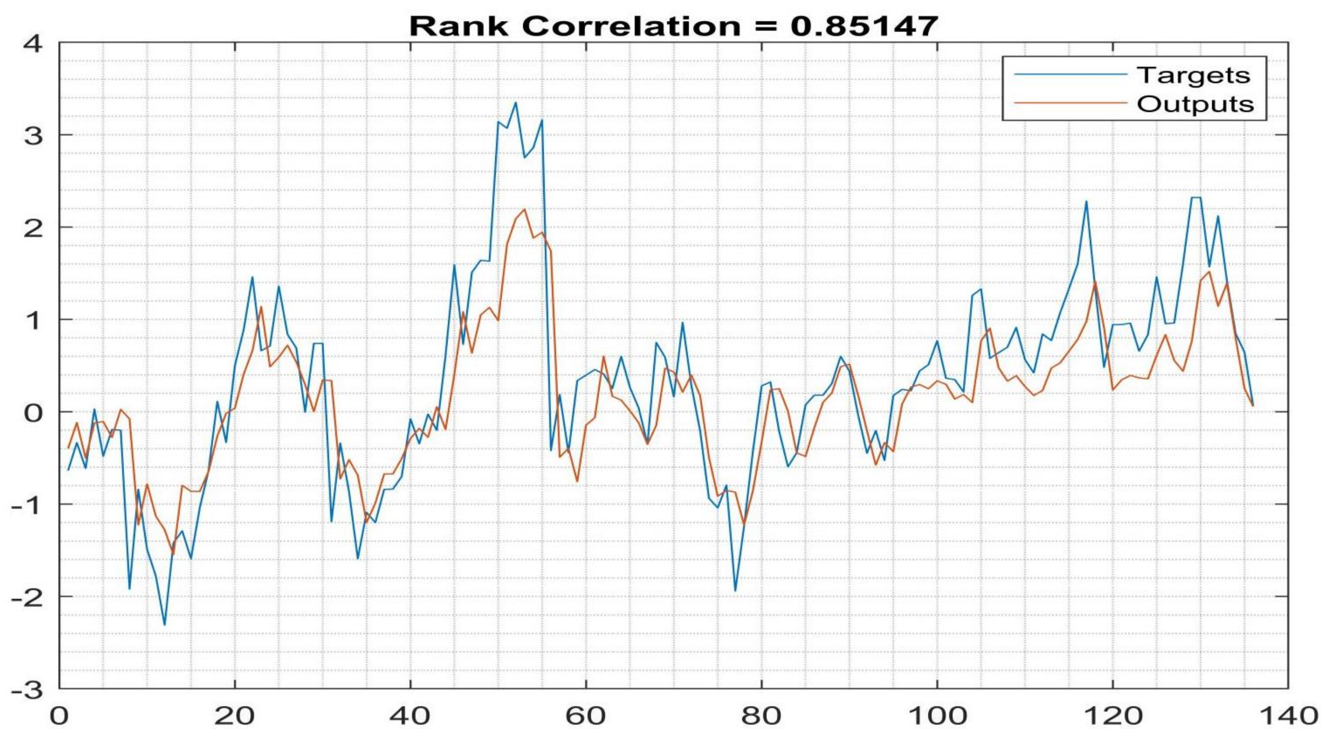


Figure 37: Comparison graph of observed(target) and predicted(outputs) for Testing data

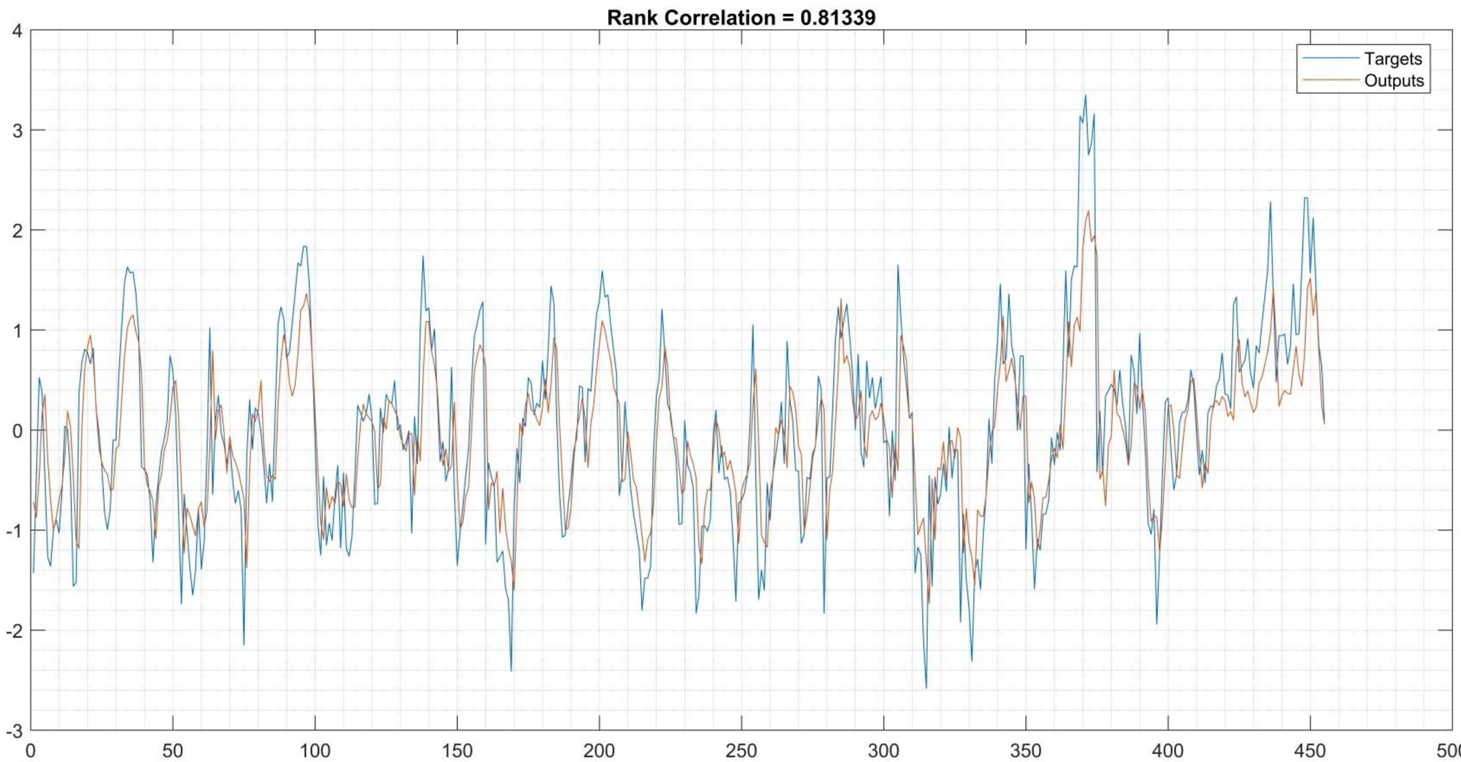


Figure 38: Comparison graph of observed(target) and predicted(outputs) for All data

Interpretation

The rank correlation used here is known as Kendall Rank Correlation. It is a scale free measure of dependence. The Rank correlation observed here is 0.81339, which is very high. It is plotted in 1 month lead time.

The lagging effect in target and output maybe due to auto-correlation. Underestimation at certain extreme ends maybe due to wet condition of the months.

CHAPTER 6

6. Conclusions

Drought being pernicious and creeping phenomena than other climatic events, it becomes difficult to predict drought in basin. Forecasting of drought is important for planning and optimal operation of irrigation system as agriculture is primary activity in Bankura district. This study was focused on drought forecasting using SPI drought index. SPI is good for understanding drought conditions as it is robust and can be used for finding drought condition for many accumulation periods like SPI1, SPI3, SPI6, SPI9, SPI 12 etc for different purposes. As it takes effect of precipitation in consideration as it would happen that if there is no rainfall then there will be surely drought for that period.

The Long run short term memory model developed for SPI6 series, SPI6 is giving good results may be because of increase in filter length as it reduces the noise effectively. The model used in this study is good as value of RMSE is acceptable which is 0.58112 in one month lead time. The predictability is more as one moves for higher lead time prediction.

CHAPTER 7

---

7. References

1. Dracup, J. A., Lee, K. S., & Paulson Jr, E. G. (1980). On the definition of droughts. *Water resources research*, *16*(2), 297-302.
2. McKee, T. B., Doesken, N. J., & Kleist, J. (1993, January). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (Vol. 17, No. 22, pp. 179-183).
3. Obasi, G. O. P. (1994). WMO's role in the international decade for natural disaster reduction. *Bulletin of the American Meteorological Society*, *75*(9), 1655-1661.
4. Guttman, N. B. (1999). Accepting the standardized precipitation index: a calculation algorithm 1. *JAWRA Journal of the American Water Resources Association*, *35*(2), 311-322.
5. Kihoro, J., Otieno, R. O., & Wafula, C. (2004). Seasonal time series forecasting: A comparative study of ARIMA and ANN models.
6. Subramanya, K. (2005). Engineering hydrology, Tata-Mcgraw Hill, New Delhi, India.
7. Steinemann, A. C., Hayes, M. J., & Cavalcanti, L. F. (2005). Drought indicators and triggers. *Drought and Water Crises: Science, Technology, and Management Issues, Wilhite DA (ed)*, 71-92.
8. Mishra, A. K., & Desai, V. R. (2006). Drought forecasting using feed-forward recursive neural network. *ecological modelling*, *198*(1-2), 127-138.
9. Mishra, A. K., Desai, V. R., & Singh, V. P. (2007). Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering*, *12*(6), 626-638.
10. Niemeyer, S. (2008). New drought indices. *Options Méditerranéennes. Série A: Séminaires Méditerranéens*, *80*, 267-274.
11. Svoboda, M., Hayes, M., & Wood, D. (2012). Standardized precipitation index: user guide.
12. Belayneh, A., & Adamowski, J. (2013). Drought forecasting using new machine learning methods. *Journal of Water and Land Development*, *18*(9), 3-12.
13. Zhang, Y., Li, W., Chen, Q., Pu, X., & Xiang, L. (2017). Multi-models for SPI drought forecasting in the north of Haihe River Basin, China. *Stochastic Environmental Research and Risk Assessment*, *31*(10), 2471-2481.
14. Vidyarthi, V. K., & Jain, A. (2020). Knowledge extraction from trained ANN drought classification model. *Journal of Hydrology*, *585*, 124804.
15. Dikshit, A., Pradhan, B., & Santosh, M. (2022). Artificial neural networks in drought prediction in the 21st century–A scientometric analysis. Applied Soft Computing, 114, 108080.