International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

RAINFALL PREDICTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES : A REVIEW

Manjunath N [1], Muralidhar B R [2], Sachin Kumar S[3], Vamshi K[4], Mrs. Savitha P[5]

[1-4]B.E. Student, Department of CSE, Sir M Visvesvaraya Institute of Technology, Bengaluru, India
[5]Assistant Professor, Department of CSE, Sir M Visvesvaraya Institute of Technology, Bengaluru, India

---***---

Abstract - *In India , Agriculture is the key point for survival. For agriculture, rainfall is most important. These days rainfall prediction has become a major problem. Prediction of rainfall gives awareness to people and know in advance about rainfall to take certain precautions to protect their crop from rainfall. Many techniques came into existence to predict rainfall. Machine Learning algorithms are mostly useful in predicting rainfall. Some of the major Machine Learning algorithms are ARIMA Model (Auto-Regressive Integrated Moving Average), Artificial Neural Network, Logistic Regression, Support Vector Machine and Self Organizing Map. Two commonly used models predict seasonal rainfall such as Linear and Non-Linear models. The first models are ARIMA Model. While using Artificial Neural Network (ANN) predicting rainfall can be done using Back Propagation NN, Cascade NN or Layer Recurrent Network. Artificial NN is same as Biological Neural Networks.*

1.INTRODUCTION

In today's situation, rainfall is considered to be one of the responsible factors for most of the significant things across the world. In India, agriculture is considered to be one of the important factors for deciding the economy of the country and agriculture is solely dependent on rainfall. Apart From that in the coastal areas across the world, getting to know the amount of rainfall is very much necessary. In some of the areas which have water scarcity, to establish rain water harvester, prior prediction of the rainfall should be done. This project deals with the prediction of rainfall using machine learning & neural networks. The project performs the comparative study of machine learning approaches and neural network approaches then accordingly portrays the efficient approach for rainfall prediction. First of all, pre-process is performed. Pre-process is the process of representing the data set in the form of several graphs such as bar graph, histogram etc. The prediction has been done using the data set which contains rainfall data from year 1901 to 2015 for different regions across the country. It contains month wise data as well as annual rainfall data for the same.

1.1 MOTIVATION

Rainfall information in the past helps farmers better manage their crops, leading to economic growth in the country. Prediction of precipitation is beneficial to prevent flooding that saves people's lives and property.

2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

Machine learning approach deals with predicting rainfall using machine learning approach. It finds the accuracy of the machine learning approach using two types of errors i.e., RE and RMSE. In these four major trends of machine learning are being used. The first one is called hybridization, which means multiple machine learning approaches are being used together and accordingly prediction is being done. The second one deals with improving the quality of dataset which is being used. Data mining approach helps to find the hidden pattern, which will help to predict the rainfall correctly. This approach takes all the parameters, which affect the rainfall such as climate, wind speed etc. and predict the future rainfall. Customized, integrated and modified data mining technique is used to predict rainfall. Many climate variables are being taken to predict rainfall.

2.2 PROPOSED SYSTEM

We have proposed ANN (ARTIFICAL NEEURAL NETWORK) based rainfall prediction and forecasting system to efficiently predict the rainfall and to do forecasting for upcoming years. It provides the better accuracy comparing to the existing approach. It consumes less time for huge amount of data.

3. SYSTEM DESIGN

3.1 System Architecture

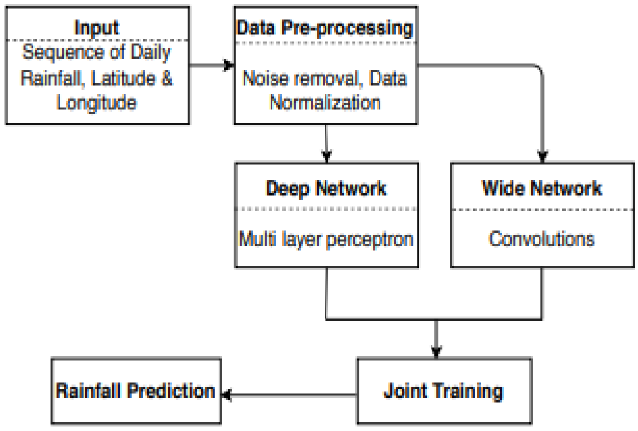

Fig 3.1 Architecture outline of the classification of rain fall prediction

The model takes sequence of dail rainfall intensities and geographical parameters, namely latitude and longitude as input. After initial pre-processing, input goes to a deep network, which is a ANN (ARTIFICAL NEURAL NETWORK) and a wide network consists of convolutions. The model is trained using joint training approach, considering outputs from deep and wide networks simultaneously.

3.2 Flow Chart

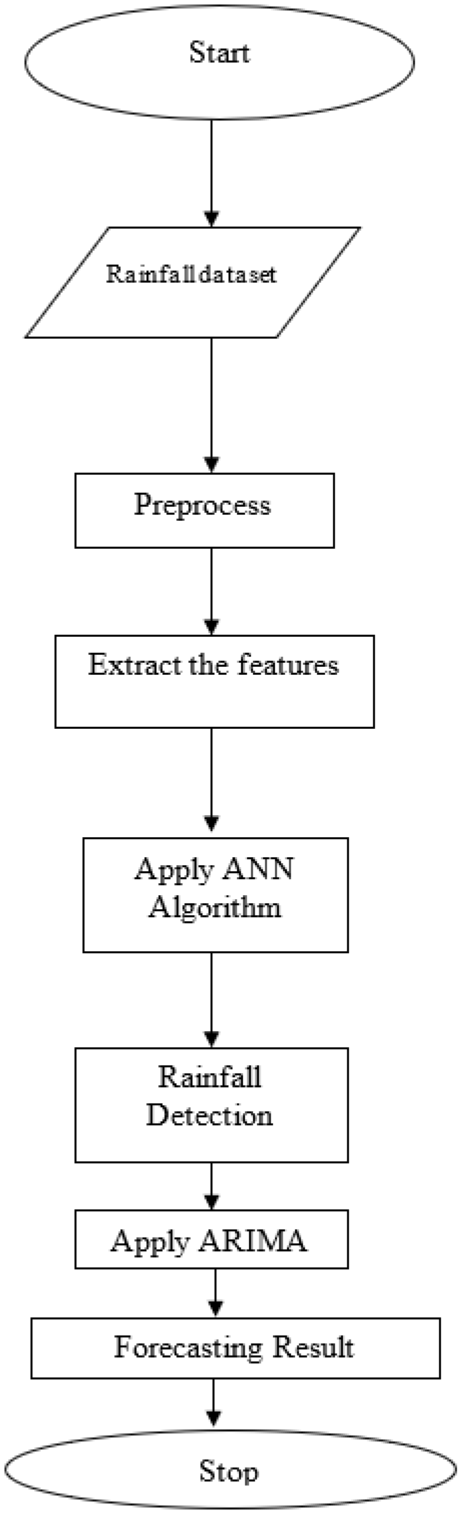

Fig 3.2 Flow Chart

3.3 Use Case Diagram

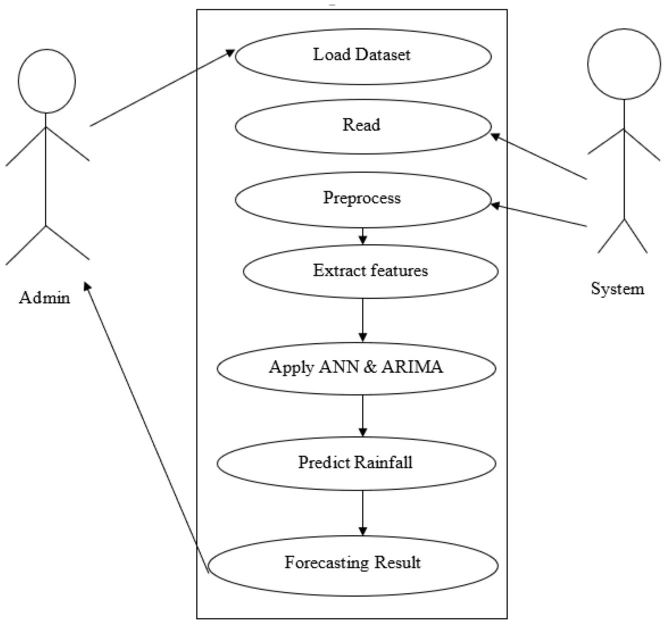

Fig 3.3 Use Case Diagram

3.4 Sequence Diagram

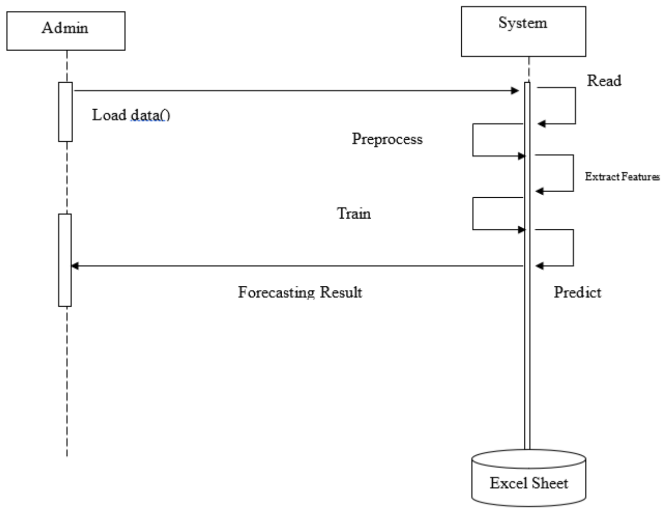

Fig 3.4 Sequence Diagram

3.5 Data Flow Diagram

Level 0:

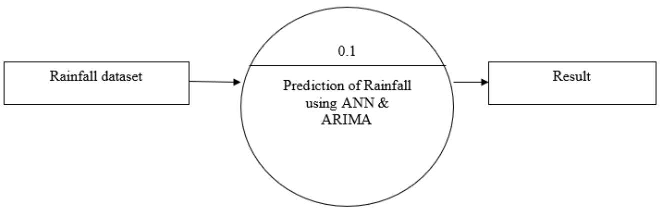

Fig 3.5.1 Level 0 Data Flow Diagram

Level: 0 describes the overall process of the project. We are using rainfall data set as input. System will use the ANN algorithm to predict the rainfall ARIMA mode is used to forecast the future result.

Level 1:

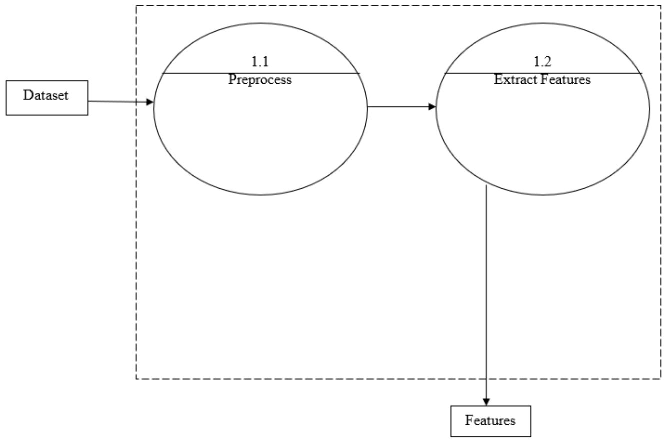

Fig 3.5.2 Level 1 Data Flow Diagram

Level2:

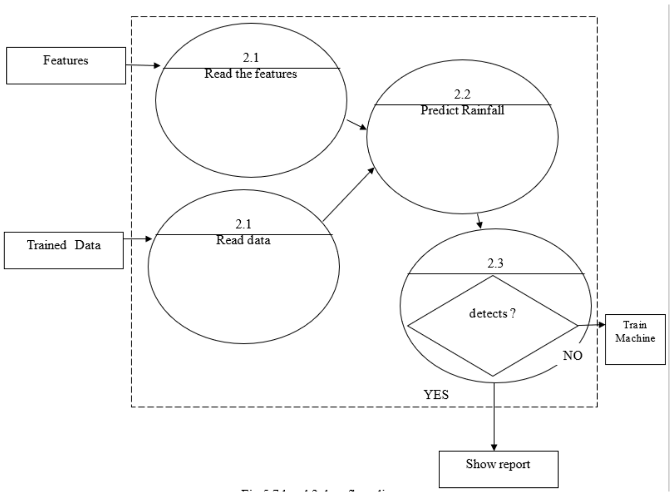

Fig 3.5.3 Level 2 Data Flow Diagram

3.6 Class diagram

Class diagram is a static diagram. It represents the static view of an application. Class diagram isnot only used for visualizing , describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modelling of object-oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram.

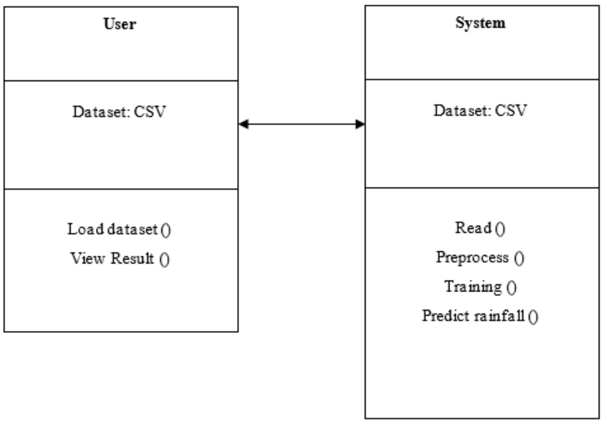

Fig 3.6 Class Diagram

3.7 Activity diagram

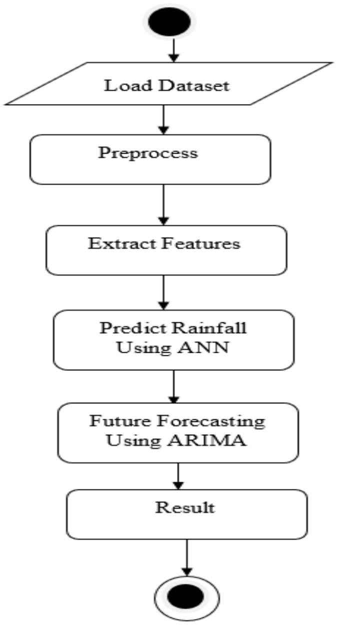

Fig 3.7 Activity Diagram

4. ALGORITHM APPLIED

4.1 ARTIFICAL NEURAL NETWORK

Artificial neural network model ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during its learning phase. The neural network is neurons connected together with the output from one neuron becoming input to others until the final output is reached. The network learns when an example of a set of input data with known results/output are presented to it, the weighting factors are adjusted (either through human intervention or by a programmed algorithm) and these connection weights store the knowledge necessary to bring the final output closer to the known result (Haykin 1999). In this present study, ANN models with three training algorithms were developed to forecast the daily rainfall. Using the available data of the study area, trial and error approach has been employed in finalizing the present ANN structure. The Neuro solution version 5 (http://www.nd.com) has been used in the model development.

The first ANN model (A) was trained using MLP back-propagation algorithm network with simple structure, four nodes in the input layer, single hidden layer with seven nodes and one node in the output layer. Input to the model is the present-day rainfall data (t) and the 3-day lagged rainfall [(t 1) (t 2) (t 3)], while the output is rainfall of the next day (t þ 1). The transfer function used is the sigmoid function with 400 numbers of epochs.

In the second ANN model (B), the radial basis function (RBF) was used for training the network. The input and the output of training data set were kept same as MLP network. However, the transfer function, TanhAxon, was used .

The third ANN model (C) was trained using time-lagged recurrent networks (TLRNs). Data used to train the model was the same as the previous two models (A and B). In the TLRN algorithm, the increased number of nodes in the hidden layer reduced the performance and hence the number of nodes in the hidden layers was reduced to two. For all the networks, the number of hidden layers and number of neurons in each layer were found by trial and error. Out of 47 years of rainfall data, 35 years of data are used for training and remaining 12 years are used for testing; this length is achieved through a trial-and-error modelling approach.

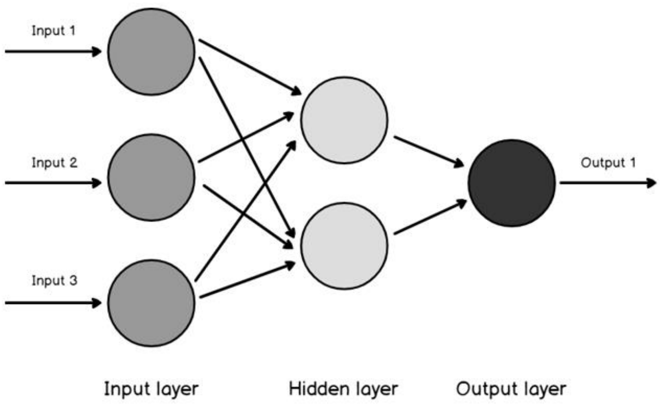

Fig 4.1 Artificial Neural Network

4.2 ARIMA forecast model:

ARIMA is an acronym that stands for Auto Regressive Integrated Moving Average. It is a class of model that captures a suite of different standard temporal structures in time series data.

Components of ARIMA:

- ARIMA has three components – AR (autoregressive term), I (differencing term) and MA (moving average term). Let us understand each of these components
- AR term refers to the past values used for forecasting the next value. The AR term is defined by the parameter 'p' in arima.
- MA term is used to defines number of past forecast errors used to predict the future values. The parameter 'q' in arima represents the MA term. ACF plot is used to identify the correct 'q' value.
- Order of differencing specifies the number of times the differencing operation is performed on series to make it stationary. Test like ADF and KPSS can be used to determine whether the series is stationary and help in identifying the d value.

Steps of ARIMA

- Load the data: This step will be the same. Load the data into your notebook
- Preprocessing data: The input should be univariate, hence drop the other columns
- Fit Auto ARIMA: Fit the model on the univariate series
- Predict values on validation set: Make predictions on the validation set
- Calculate RMSE: Check the performance of the model using the predicted values against the actual values.

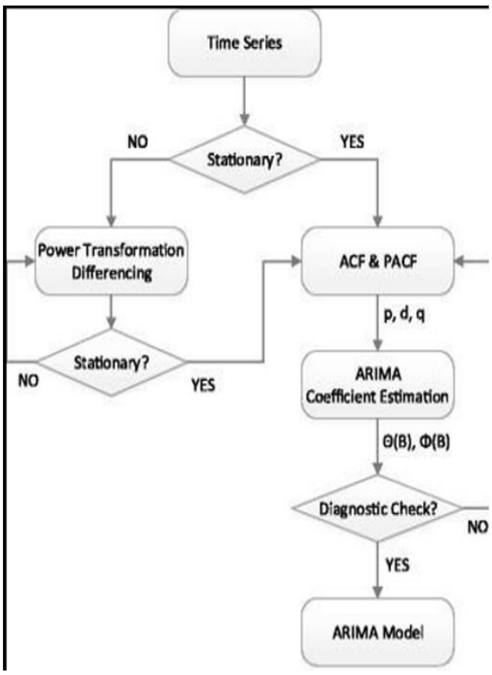

Fig 4.2 ARIMA Flow Chart

5. IMPLEMENTATION

5.1 Dataset description and pre-processing

A data set is a collection of data. Dataset contains only the matrix of numerical data. We use the data set which contains the rainfall data from 1901 to 2015 for different regions across the country. For Data pre-processing we use the most commonly used min-max normalization method to convert all rainfall intensity values to a number between 0 and 100.

Pseudo Code:

Procedure Data Collection preprocess ()

Input: Rainfall dataset

Output: Cleaned data

Begin:

Step1: Read the dataset

Step2: Detect non valid data in every row

Step3: if any non-valid data then

Step3.1: remove data from dataset.

Step4: else

Step 4.1: keep the valid data

End if

End

5.2 Building Model

We apply a convolutional layer to capture such combinations. In addition to this, to make our model more generalized with respect to different atmospheric conditions, we are using geographical parameters namely, longitude and latitude while designing and developing our model.

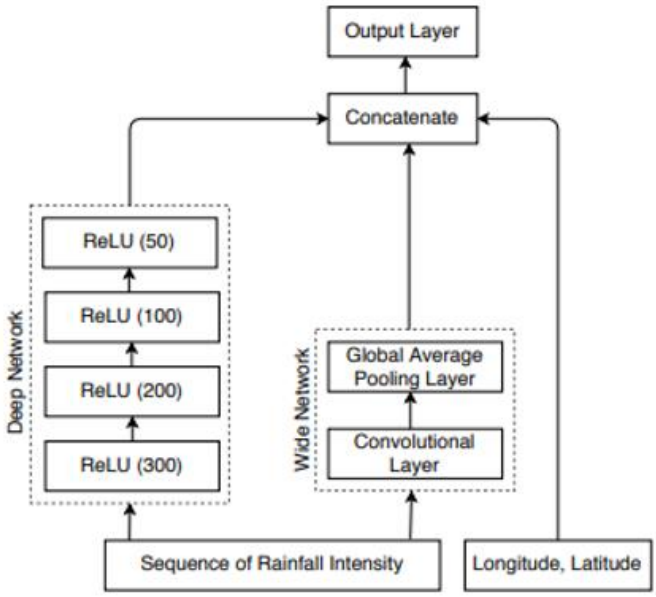

Fig:5.2 ANN model

We made a parameter exploration concerning the batch size, hidden layers, number of neurons, dropout rates and optimization algorithms using trial-and-error method. The deep part is a multi-layer perceptron with an input layer; 4 hidden layers containing 300, 200, 100 and 50 neural units with ReLU as the activation function; and finally a dense output layer. In order to prevent over-fitting of the model, dropout layers (Srivastava et al., 2014) with dropout rate 0.3 are added after each hidden layer. The wide part contains a convolutional layer with 100 filters, each of size 1x5, followed by a global average pooling layer. The output of both the wide and deep networks is concatenated, along with the latitude and longitude values, and the model is trained using the joint-training approach.

5.3 Training Phase

Pseudo Code:

Procedure Training Phase ()

Input: Dataset(csv)

Output: trained file (.h5 file)

Begin:

Step1: Read the dataset

Step2: for each row in dataset

Step3.1: Extract the features.

Step 3.2: Apply ANN.

Step 4: Train the extracted features with the Corresponding latitude and longitude using ANN.

Step 5: Save the trained file.

Step 6: return the trained message

End

5.4 Rainfall Prediction

We proposed rainfall prediction by incorporating ANN techniques. This technique does not play out any sub-sampling, but it optimizes over all dataset. This method is much accurate to predict rainfall with 99.69% accuracy.

Pseudo Code:

Procedure Rainfall ( )

Input: weather condition and geo coordinates

Output: predicted rainfall

Begin:

Step1: Read the data form the input

Step 2: Load the pre-trained model

Step3: Extract the features

Step 4: Predict rainfall.

Step 5: Save the result.

Step 6: return the result message

End

5.5 Future Forecasting:

We are using the ARIMA mode to forecast the future rainfall level.

Pseudo Code:

Procedure Future Forecasting ()

Input: Rainfall Dataset

Output: forecasted result for next 10 days

Begin:

Step1: Read the dataset

Step 2: Apply ARIMA model

Step3: Predict the results for next 10 days

Step 4: Save and show the result in graph.

End

6. RESULT

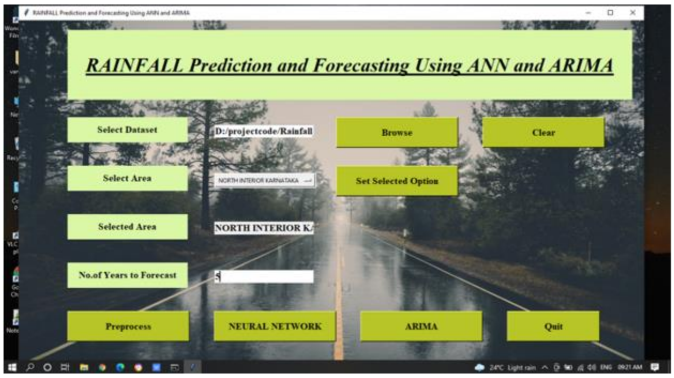

Fig 6.1(a) Main Page

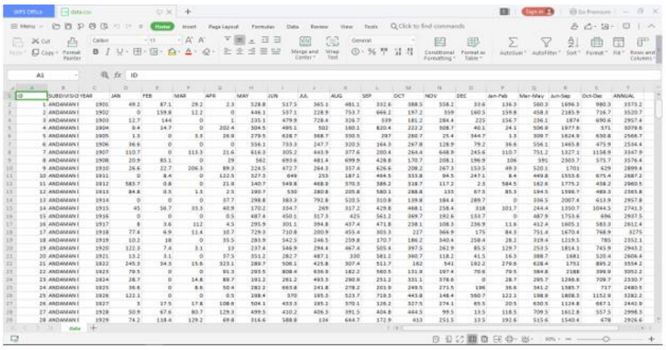

Fig 6.2(b) Data Set

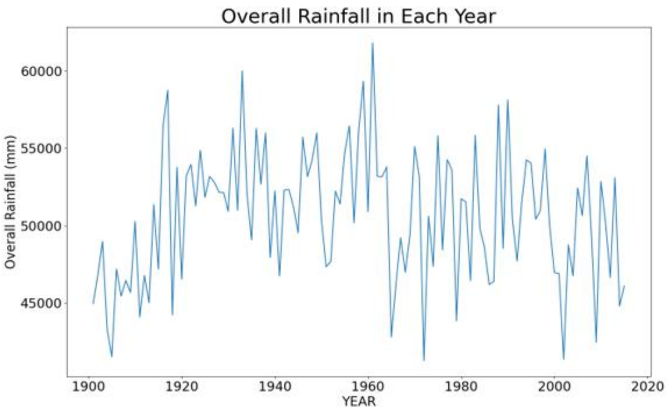

Fig 6.3(c) Overall Rainfall in Each Year

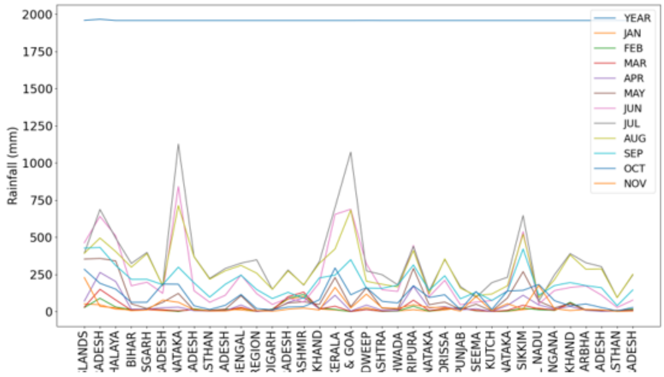

Fig 6.4(d) Overall Rainfall in Each Month of Year

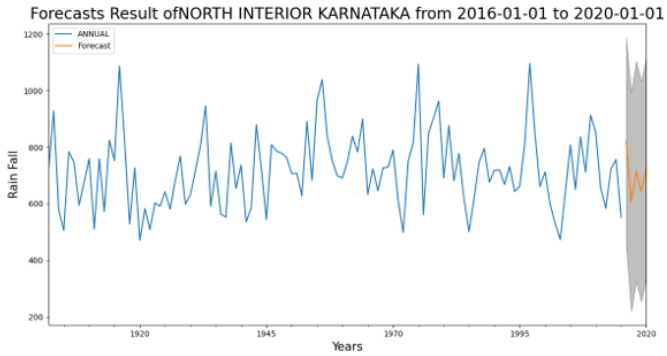

Fig 6.5(e) Forecasting using ARIMA

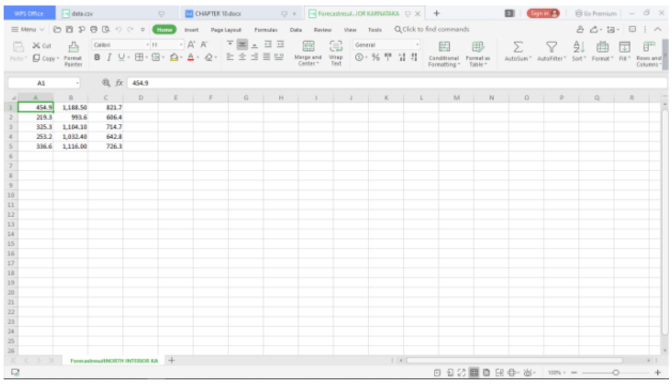

Fig 6.6(f) Forecasting values in Excel Sheet

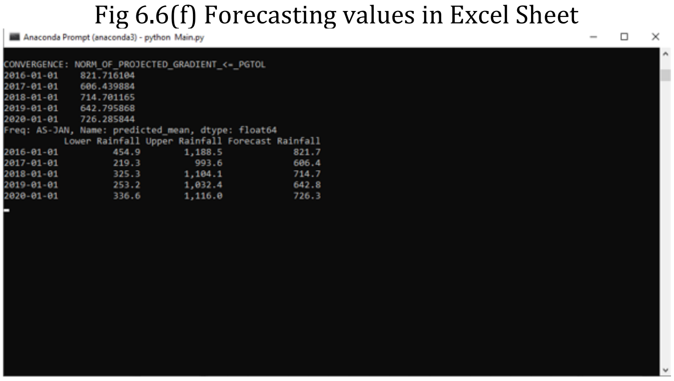

Fig 6.7(g) Forecasting Values in Console

7. CONCLUSIONS

This project represented the Deep Learning Approach for predicting the rainfall by using the ANN (ARTIFICAL NEURAL NEIWORK). Comparing the present architecture with other state approaches. This project provided a study of different types of methodologies used to forecast and predict rainfall and issues that could be found when applying different approaches to forecasting rainfall. Because of nonlinear relationships in rainfall datasets and the ability to learn from the past, Artificial Neural Network makes a superior solution to all approaches available.

The future work of the project would be the improvement of architecture for light and other weather scenarios. Also, can develop a model for small changes in climate in future. An algorithm for testing daily basis dataset instead of accumulated dataset could be of paramount Importance for further research.

REFERENCES

1. Rainfall dataset5of our country is5taken from kaggle website https://www.kaggle.com/rajanand/rainfall-in-india

2. Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using5machine learning5models: Literature review. Water (Switzerland), 10(11). https://doi.org/10.3390/w10111536

3. Janani, B; Sebastian, P. (2014). Analysis on the weather forecasting5and techniques. International Journal5of Advanced Research in Computer Engineering & Technology, 3(1), 59–61. http://ijarcet.org/wp-content/uploads/IJARCET-VOL-3-ISSUE-1-59- 61.pdf

4. Chaudhari, M. S., & Choudhari, N. K. (2017). Open Access Study of5Various Rainfall Estimation & Prediction Techniques5Using Data Mining. American Journal of Engineering Research (AJER), 7, 137– 139. http://www.ajer.org/papers/v6(07)/Q0607137139.pdf

5. Aakash Parmar, Kinjal Mistree, M. S. (2017). Machine5Learning Techniques for rainfall prediction: A Review. International5Conference on Innovations in Information Embedded and Communication Systems (ICIIECS). https://www.researchgate.net/profile/Aakash_Parmar4/publication/319503839_Machine_Learning_Techniques_For_Rainfall_Prediction_A_Review/links/59afb922458515150e4cc2e4/Machine LearningTechniques-For-Rainfall-Prediction-A-Review.pdf

6. Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I., & Nawaz, Z. (2018). Rainfall prediction using data mining5techniques: A systematic literature review. International5Journal of Advanced Computer Science and Applications, 9(5), 143–150. https://doi.org/10.14569/IJACSA.2018.090518

7. Meng – Hua Yen , Ding – Wei Liu , Yi – Chia Hsin, C. – E. L. and C. – C. C. (2019). Application5of the deep learning for the5prediction of rainfall in Southern Taiwan.5Scientific Reports, 9(1), 1–9. https://doi.org/10.1038/s41598-019-49242-6

8. Shi, X., Chen, Z., & Wang, H. (2015). Convolutional LSTM Network. Nips, 2–3. https://doi.org/[]

9. Wahyuni, E. G.Fauzan, L. M.F.Abriyani, F.Muchlis, N. F., & Ulfa, M. (2018). Rainfall prediction5with backpropagation method.5Journal of Physics: ConferenceSeries,983(1). https://doi.org/10.1088/17426596/983/1/012059