Long term prediction of rainfall in Andhra Pradesh with Deep learning

Debnath Bhattacharyya1, Kalam Swathi2, N. Thirupathi Rao3, Nakke Marline Joys Kumari4

1. Koneru Laksmaiah Education Foundation, Greenfield, Vaddeswaram, Guntur, Andhra Pradesh, India
2. Vignan Institute of Information Technology, Visakhapatnam, Andhra Pradesh, India
3. Vignan Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, India

ABSTRACT

Rainfall is the major concern for almost types of people in the society. It helps different types of people in the society in different means. For some people, it’s the source of providing drinking water. For others like farmers, it’s the source of their livelihood and for all other human beings too. As farmers grow the farm and rice will be produced for the human being to eat. As a whole, the rainfall plays a key role for almost all kinds of people living in the society. Prediction of this rainfall is always an interesting and useful news for any kind of people in the society. Especially for the government agencies, it’s a very useful source as based on predictions, the harvesting and storing of water could be prepared well in advance. It also plays a key role in the production of power or generation power with the help of water flow in dams and reservoirs. In the article, an attempt has been made to build a system that takes input of previous years rainfall data from dataset which was collected from the Meteorology department of Andhra Pradesh and predict the average amount of rainfall of upcoming months in a specific year. We have initially separated the available data into training and testing data sets and made a model. We had applied various statistical and machine learning approaches like the linear model, Support Vector Machine algorithms etc. in predicting the results, along with a Neural Network model and made analysis over various approaches and compared their results with actual data. By these various approaches, the error of prediction was minimized and increased the accuracy of results predicted by the model.

Keywords: Linear Regression, Support Vector Machine, Artificial neural network, rainfall forecasting, model selection.

Received - 19/06/2021, Reviewed - 27/06/2021, Revised/ Accepted- 01/07/2021

Correspondence: Debnath Bhattacharyya* debnathb@kluniversity.in

INTRODUCTION

Agriculture is the backbone of the India’s work and economy (1). Most of the people who lives in rural India depends on the agriculture or agriculture was their main work and source of livelihood. The country depends mostly on agriculture and its related products development and usage. The provision of irrigation was getting developed slowly and it had become little faster than the previous decades. The sufficient number of irrigation projects are still not constructed or not available (2) farmers in the country. As a result, most of the water from rains will be passes to the sea and most of the water from rainfall could be used for agriculture and related sources. The prediction of rainfall is also one of the most important

Factors here to make the necessary arrangements for storing such rain water for both irrigation and for drinking water too. The saying a good and better (3) rainfall may give good storage of water for drinking and for agriculture too in the country.

The necessary of predicting the rainfall was very (4) serious due to several reasons. For a country like India where every few water storage capacities are available and most of the rural India depends on rain for drinking water. The information related to rainfall is always a good and interesting news for almost all types of people living in rural India. The economy and employment for rural India mostly depends on agriculture where they will produce most of the food and related crops to make feed to the other societies in the country. Hence, the prediction of rainfall is always a good work we felt and very few works were available in providing such information. The productivity (5) of the
crop also depends on the water available in ground and similarly the rainfall to be taken place during its growth of crop.

The other important factor to be noted was that the power generation was also another important factor to be considered for the rainfall prediction. As the rains are good, more amount of water was stored in reservoirs where the hydroelectricity was being generated with very low cost compared to other types of thermal and other sources of power generation. In the current work, a comparison between the various algorithms to minimize the error rate had performed. By taking the input of 100 years of rainfall data from Indian Meteorological Department (IMD) and applied various statistical, Data Preprocessing and Machine learning approaches. By using the previous data sets and the pattern of the rainfall to provide the most accurately predicted rainfall as a long-term forecasting. The data of Andhra Pradesh is taken to predict the average rainfall by separating data into training and testing. We have compared using classification algorithms like (Support vector machine (SVM), Linear Regression.) and analyzing monthly and yearly amount of rainfall in a specific location.

As the data needs to be processed for better results and for better predictions, the data set needs to be made for two purposes. The data was made to work as training purpose and for the testing of data. This phenomenon had done to evaluate the models which were considered with the help of machine learning algorithms. The dataset which was considered for training covers almost 75% f the data available in the dataset and the remaining 25% of data will from the dataset will be used for testing purpose f the model one the model was built successfully. The dataset was preprocessed by removing any unwanted data or any data errors presented in the dataset before using for the implementation. The values related to the rainfall were clustered by using the clustering algorithm and the states of the rainfall was noted as various stats like the actual rainfall state, predicted state of rainfall and submitted to the model for further training purpose of the model. In general, the Histogram formation was considered here and for the same histogram both regression and classification had used for better results. The major advantage with this step will be achieved was that the computation complexity will be reduced a lot. The process continues until with a status had reached. The status might be two types like the model reaches an error which could be an acceptable error or the threshold had crossed by the iterations count in the model. The artificial neural network model also implemented and tested for the generation of better results.

As a whole, three of the approaches may be used within the framework of a real time forecasting system, though with different levels of reliability. For better understanding of the problem, an example had provided as follows:

For example

Input
Prediction year, and month. (It undergoes the all-classification algorithms and data preprocessing techniques).

Output
Predicted amount of rainfall in mm (millimeters) (the comparison of algorithms in the result e.g. the prediction accuracy of neural network is 77.17%, SVM is 32.5% and linear model is 48.8% as per considering a year).

So, we finally consider neural network is performing the better result than remaining models.

LITERATURE SURVEY

Several authors had considered the problem of rainfall prediction with the help of different models. Some of those models are,

In (11), had discussed about the rainfall prediction. They had used various combined preprocessing techniques for cleaning the dataset before being implemented the actual processing of the dataset. They had used the feed forward neural network model for the prediction of rainfall. The results generated are shown in the tables for better understanding of the performance of the model considered here.

In (12) had developed and discussed about the rainfall prediction system based on the local data in Japan. They had collected the dataset of rainfall from the local Japan Meteorological department for analyzing the data for predicting the rainfall. They had used the neural network model to forecast rainfall. The data was a local dataset which had considered and results displayed in the form of tabular and graphical representations.

In (13) had developed and discussed about a hybrid model of using the machine learning techniques to predict the weather. They had used several machine learning techniques for the same to process and analyze the results. The results had tabulated and presented in a well-organized manner. C. Z. Basha ET. al. (4) also developed and discussed about the prediction of rainfall. They had used the neural network model to analyze and predict the rainfall.

In, (14) had considered several models and machine learning techniques for the prediction of rainfall. They had used various new approaches and techniques to enhance the performance of the model for the prediction. They had also used some forecasting techniques such that to predict the rainfall in an earlier time to make the people alert. Kaushik Dutta et.al. (6) had developed a model through which the prediction of rainfall could be predicted. The model was equipped with machine learning algorithms and other data mining techniques such that to make the prediction model more accurate and more efficient. The results generated are compared with the precious models which were already developed and current model had good improvement in the performance when compared with the previous models.

In (15) had developed a model through which the rainfall can be predicted in the Queensland area of Australia. They tried to forecast the rainfall in this area by using the artificial neural network models.
(16) had discussed and developed a model to predict the rainfall. He used the model of rainfall prediction with the help of time series model. He also used the multilayer perceptron network model such that to predict the rainfall. The results are very encouraging and the current model’s performance was compared with that of the other previous existing models. The model performance was very good when compared with the other existing models.

From all those above various works considered, the identification of rainfall was not same all the time and also the identification of rainfall various from time to time and also location specific. The amount of rainfall might take places may change from place to place. As a result, we cannot assume the same method or same model may process the better results with different locations. The rainfall prediction might depend on different types of features like the type of soil, environmental conditions etc. As a result, we need to predict the rainfall at various locations by considering the local atmospheric and environmental conditions too. Very few works had observed in rainfall prediction using the linear regression models, machine learning algorithms and the convolutional neural network models. As a result, in our current we tried to identify, predict and analyze the data by using the models like the Linear Regression Model, SVM algorithm and the Convolutional Neural Network models. As we had already discussed, the identification rainfall prediction based on location specific and depends of environmental conditions, we had chosen the area of which the prediction had to be identified was the Andhra Pradesh State of India. Where there is a huge difference of rainfall in different areas in the state of Andhra Pradesh.

**Proposed System**

Our proposed model predicts the long-term average amount of rainfall in Andhra Pradesh by considering the previous big data taken from India Meteorological Department. Our model also overcomes the problem of inconsistency of data along with low error rate in predictions. Rainfall prediction poses right from the ancient times as a big herculean task, because it depends on various parameters to predict. The Models for monthly rainfall prediction in which we evaluations, the anticipated models were compared by using Machine learning, Data preprocessing techniques like Artificial neural networks (ANN), classification algorithms such as (SVM, Linear regression.) to predict. The present research is focused on using the previous data sets and the pattern of the rainfall to provide the most accurately predicted monthly and yearly rainfall as a long-term forecasting.

**ALGORITHMS USED IN THE CURRENT MODEL**

Some of the algorithms or the techniques used in the current work to predict the rainfall are discussed here like,

**Linear Regression**

Linear regression is a supervised learning technique which comes under the model of machine learning algorithms. The regression task is being performed by this model. Several other types of regression techniques are available like the logistic regression etc. Basically, this regression had evolved or taken from the Statistics and being used in the Machine learning. In other words, this regression was borrowed from the statistics and being used in machine learning based problems. These regression model types vary from data to data based on the type of data like the dependent or independent data points. They also differ based on the relationship between the dependent and independent variables. In general, these models count the number of independent variables is used in the designed or executing model for better results (9).

In general, the machine learning models always tries to work on reducing the error in the data and tried to increase the more accurate predictions. These two things are interrelated to each other. As errors in the data of a dataset are less, the results expected from the dataset might give better output as a result the accuracy of the model could be enhanced a lot. This linear regression mode works to identify a relationship between more input variables and a single output variable.

**Support Vector Machine Algorithm**

The support vector machine algorithm is also one of the supervised learning model-based algorithms. This algorithm is also mainly used for the classification type of problems. Once two sets of data have been given to this algorithm as an input, the next level type of data will be predicted by this algorithm. It mainly works on two terms namely the hyper plane and the support vector. In general, the hyper planes are considered the points of data through which the data can be classified. The other points are the support vectors are other type of data points which lies very closer to these hyper planes and tried to influence the orientation and position of these hyper planes (10).

**Convolutional Neural Networks in Keras**

A convolutional neural network is a deep learning-based algorithm which could be used for the prediction of new data labels from the existing previous data. The normal working of a convolutional algorithm that it takes an image as input and processes it. The weights or the importance will be assigned for the various labels or the various locations on the image based on the importance. This assignment may vary from one point to the other point. The data needs to be pre-processed for analyzing and yielding better results in any case. But, in the case of this CNN models, the early pre-processing of much data is not required. The filters for filtering the data also embedded in these models such that to get the better results from the models considered.

The entire process of the current model works on two stages and those are the training and testing of the dataset chosen for the prediction of rainfall and the second phase was the fitting of the model and evaluates the model for getting the accurate results.

**DESIGN AND IMPLEMENTATION OF THE MODEL**

For predicting the rainfall at various locations, different strategies had followed by various authors. From the literature we had
understood that the prediction of rainfall and the conditions to be considered for the prediction might vary from location to location. The points on which the rainfall may depend like the type of the land or the soil, the climatic conditions, temperature and other environmental based conditions. As a result, several works of the authors had considered the database of rainfall of previous years from local government agencies for accurate and useful predictions. In the current work also, we had collected the dataset of rainfall for the last 100 years belongs to the state of Andhra Pradesh of India. This state comprises of three types of areas which varies in both type of lands and climatic conditions.

Among three regions of the Andhra Pradesh state, one was the “Coastal Region” which was full of agricultural lands and full of ground water. The land here was the very fertile land the farmers here will grow the rice twice in a year. The other part of the State was the “Uttar Andhra” which was very under developed and heavily cyclone hit areas. The people live here in this area always sees nearly one to two cyclones crossing thee districts every year. The land here also fertile land. The third part of the State was the “Rayala seema Region”, where the rainfall was very low. In some years, this region had got the very less rainfall in entire country also. The average rain fall of this region was less than the entire average of India’s rainfall. So, this made us to select this state of Andhra Pradesh to predict the rainfall as there is a diverse type of rainfall was noted in realistic conditions and people are very affected with this less amount of rainfall.

In order to use the current application for farmers any other normal person of a society, a user interface had been developed and provided for better utilization. The user interface had shown at figure 1.

**Purification of fungal sample**

Acetone purification done in dialysis membrane and this purified sample were taken after 24 hour incubation as shown in Figure 1.

![Figure 1: The home page of user interface for the current application](image1.png)

The workflow or the model diagram of the current work considered here was sown at figure 2.

The entire work in the current model had processed in two phases. The first phase was the collection of datasets and passes the dataset for pre-processing. In the stage of preprocessing the data, the errors in the data of the dataset or the values with missing values are identified and corrected in the first phase of the dataset. Similarly, if any unwanted data also present in the dataset also cleared in this stage. Then, the dataset was processed by using the chosen algorithm in the model and the results will be generated as per the model’s output.

![Figure 2: The architecture model of the proposed work](image2.png)

The working of the current model was shown above at figure 2 in detail. The rainfall data which was collected from the meteorological department for the last 100 years was fed to the model. The dataset given to the model was pre-processed and the data forwarded for the further analyzing and results generation. Then, the dataset will enter to the phase of feature selection. Here some of the important features to be considered here are like the area of the location where the rainfall needs to be identified or predicted and other related features. After the feature selection process had completed, the work will be classified into two parts namely the training of data and the testing of the data. In the phase of training the data, the CNN model was used and the training phase will be completed for the model and then the model was built and the accuracy of the model built will be arrived. The testing of the model also implemented and the results will be generated for further processing to identify the accuracy of the model. From this phase, the output of the predicted rainfall will be displayed as output. The output may be considered here for two types namely the high accuracy and the low accuracy. If the accuracy of the model was high, the output is considered, if the accuracy is lee or low, then again, the model was running for further times and the result were generated.

The considered techniques like the linear regression models, Support vector machine algorithm and the convolutional neural networks models had implemented for the considered dataset for predicting the rainfall and the results and entire process had discussed as follows. For better understanding of the influence of each value of the dataset or some selected values a histogram had been generated such that to understand the frequency distribution of various set of values in a dataset occurs. In the current work also, a histogram had been generated and shown at figure 3 for the better understanding of the distribution of various values in the dataset with respect to the frequency distribution.

From the below histogram at figure 3, the distribution of rainfall for several months in a particular year had displayed. From the figure, it is understood that the rainfall had increased gradually in the
From the data available in the dataset, it is observed that the year of 1954 had the highest rainfall since last 100 years. This gives the highest rainfall since last 100 years as per the model built as an output. For better understanding of the model, the results and the rainfall had displayed in the form of a table for better understanding of the users.

From above figure 5, it is clear that if amount of rainfall is high in the months of July, August, September then the amount of rainfall will be high annually. It is also observed that if amount of rainfall in good in the months of October, November, December then the rainfall is going to be good in the overall year.

In order to verify and test the performance of the current model developed, the model had ran for the year 2005 to predict the rainfall and compared with the actual rainfall noted by the meteorological department of India.

The ground results and the models output results had compared for better understanding of the model and almost the model is performing not bad in rely with the actual data. Also we had chosen the year 1996 as a random year and tried to analyze the performance of the model with the three techniques considered here. The performance of the had presented at various figures like figure 7, 8 and figure 9.

The output of the model for rainfall prediction during the year 1996 and the actual rainfall noted with respect to the linear regression model had noted and shown at the above figure 7.

The output of the model for rainfall prediction during the year 1996 and the actual rainfall noted with respect to the support vector method algorithm had noted and shown at the above figure 8.

The output of the model for rainfall prediction during the year 1996 and the actual rainfall noted with respect to the CNN algorithm had noted and shown at the above figure 9.
The mean absolute error that had for the three models considered was shown at table 1 as,

| Algorithm                  | Mean Absolute Error |
|----------------------------|---------------------|
| Linear Regression Model    | 65.67941720064871   |
| SVM Algorithm              | 77.94353245327993   |
| Convolutional Neural Network| 51.45843444837714   |

From the above table 1, it is understood that the error rate for the three models had calculated and tabulated. The SVM algorithm is having the highest error rate when compared to the other two algorithms and the Convolutional neural network is having the least error rate from which we can get the best predictions. As the error rate is less, the predictions from the model might be good and expected from the actual rainfall being happen at the coming years or the coming months of the period.

**CONCLUSION**

In the current article an attempt had been made to develop a model by using the deep learning and machine learning techniques to predict the rainfall in the state of Andhra Pradesh in India. A simulation-based model was running with different algorithms and the results were presented in the implementation and results section.

During the simulation of the algorithms, in the graph some points were not forming cluster rather these points were away from the clusters and the straight-line distance (Euclidean Distance) was greater. This problem had found in some month’s data. As a result, the result was giving quite different value than the actual rainfall amount in that month. In other cases, the simulation results were closer value to the actual data of rainfall. The current article and work could help normal people and also the farmers whose farming depends mainly on the rainfall. The farmers may come to know about the possibility of rainfall long before it will take place and a huge number of disasters due to the heavy rainfall and flood will be shorten in future.

**REFERENCES**

1. Duong Tran Anh, Thanh Duc Dang and Song Pham Van, 2019. Improved Rainfall Prediction Using Combined Pre-Processing Methods and Feed-Forward Neural Networks.
2. Tomoaki Kashiwaoa, b, Koichi Nakayamaa, Shin Andoe, Kenji Ikeda d, Moonyong Lee, Alireza Bahadori, 2017. A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency.
3. Aditya Grover, Ashish Kapoor, and Eric Horvitz, 2015. A Deep Hybrid Model for Weather Forecasting, Pages 379–386, Sydney.
4. C Z Basha, N Bhavana, P Bhavya and S V, 2020. "Rainfall Prediction using Machine Learning & Deep Learning Techniques," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, pp 92-97.
5. 5U Shah, S Garg, N Sisodiya, N Dube and S Sharma, 2018. "Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques," 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan, India, pp 776-782, doi: 10.1109/PDGC.2018.8745763.
6. 6Kaushik Dutta, Gouthaman P, 2020. Rainfall Prediction using Machine Learning and Neural Network, International Journal of Recent Technology and Engineering (IJRTE), Volume-9 Issue-1, Pp,1954-1961.
7. Abbot J and Marohasy J, 2012. “Application of ANN to rainfall forecasting in Queensland”, Australia, Advances in Atmos, Sci, vol.29, no.4, pp.717-730.
8. Deshpande R R, 2012. “On the rainfall time series prediction using Multilayer Perceptron Artificial Neural Network,” Int. Jr. of Emerging Technology and Advanced Eng, vol 2, no1, pp.148-153.
9. V Singh, Priya, Shilpi, Vashistha, 2014. “Time Series Analysis of Forecasting Indian Rainfall,” Int. Jr of Innovations & Advancement in Comp Sci, vol 3, no 1, pp 66-69.
10. D. Pinky Saikia and Tabbilder Hitesh, 2014. “Prediction of Rainfall Using Data Mining Technique over Assam”, J I C S E, Vol 5 No 2.
11. Anguraj, DK, Mandhala, V N, Bhattacharyya, D, 2021. Hybrid neural network classification for irrigation control in WSN based precision agriculture, Journal of Ambient Intelligence and Humanized Computing.
12. Sk Meeravali, Debnath Bhattacharyya, N Thirupathi Rao & Yu-Chen Hu, 2021. Performance analysis of an improved forked communication network model, Connection Science.
13. Bhattacharyya, D, Kurnam, N M J, Joshua, E S N, & Rao, N T, 2020. Advanced Empirical Studies on Group Governance of the Novel Corona Virus, MERS, SARS and EBOLA: A Systematic Study. Int J Cur Res Rev| Vol, 12(18), 35.
14. Kumari N M J, Krishna K K, 2018. Prognosis of Diseases Using Machine Learning Algorithms: A Survey. In 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT) (pp. 1-9). IEEE.
15. Joshua ESN, Chakkravarthy M, Bhattacharyya D, 2020. An extensive review on lung cancer detection using machine learning techniques: A systematic study. Revue d'IntelligenceArtificielle, Vol 34, No 3, pp 351-359.
16. Eali Stephen Neal Joshua, Debnath Bhattacharyya, Midhun Chakkravarthy, Yung-Cheol Byun, 2021. "3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted Class Activation”, Journal of Healthcare Engineering, vol 2021, Article ID 6695518, 11 pages.

How to cite this article
Debnath Bhattacharyya, Kalam Swathi, N. Thirupathi Rao, Nakka Marline joys Kumari, 2021. “Long term prediction of rainfall in andhra pradesh with deep learning” Jour. of Med. P’cuteal & Alli. Sci. V 10 - I 4, 1397 P-3132-3137. DOI: 10.22270/jmpas.V10I4.1397.