EasyChair Preprint

№ 7684

# A Nobel Approach to Identify the Rainfall Prediction Using Deep Convolutional Neural Networks Algorithm

Prachi Desai, Ankita Gandhi and Mittali Acharya



April 2, 2022

# A Nobel approach to identify the rainfall prediction using *Deep Convolutional Neural Networks* algorithm

*Prachi Desai[1], Ankita Gandhi[2], Mitali Acharya[3]
[1]Research Scholar, Department of Computer Science
Engineering Faculty of Engineering & Technology, Parul UniversityVadodara, India

1 prachidesai0131@gmail.com
[2-3]Assistant Professor, Department of Computer Science
EngineeringFaculty of Engineering & Technology,
Parul University Vadodara, India
2ankita.gandhi@paruluniversity.ac.in,
[3]mitali.acharya2873@paruluniversity.ac.in

## Abstract

It is unlikely to have a big impact on a country's economic progress. It also helps to reduce the loss of life and property caused by natural disasters. The study of rainfall prediction using machine learning techniques, with a special emphasis on India. In India, about 70% of the population is dependent on agriculture and related activities. Rainfall forecasting has been an issue of major technical and economic importance in the agricultural sector. This rain prediction model is still mostly based on artificial neural networks, and it has only been used in India so far. In the present study, a comparative analysis of the two rainfall forecasting techniques, and is more accurate. In today's technology, and the ability to predict the rainfall is not very well-informed about the complexity of the data. The methods used are the methods of statistical and mathematical methods that don't work, if there is a situation of a non-linear pattern. An existing installation may fail when the level of complexity of the information contained in it, increase it in the past. Now, this is the best way to get to the waiting, in the rain, it's a study of deep learning and neural networks, and genetic algorithms, is that it gives you more precise, it is used to predict the future. In order to come up with a good rain is a joint, an appraisal is not required. The simplest and easiest approach to get on a larger scale is to use weather forecasting. In the present study, it can be used in all types of weather stations, and the forecast will give access to some parts of the country. A technique for predicting month - to - month rain for a specific area in India using a deep CNN learning method as a replacement fortuning technique. The CNN was compared against an ANN.

***Keywords: ANN, CNN, Supervised Learning,***

## Introduction

Weather forecasting ensures the property development of society and economy. As a result, there is a growing interest in predictions. has started since 650 B.C., wherever Babylonians tried to predict weather-supported observations of clouds (observed patterns). Then, multiple philosophers projected numerous statement theories. Over time, it became clear that these hypotheses were insufficient. As a result, it was assumed that there was a desire to learn more about the weather in general. Measurement of the atmosphere was done after the invention of the most up-to-date devices. Many tools, “such as the telegraph and radiosonde, enabled for a better understanding of atmospheric conditions. These tools of measuring are now used to

record atmospheric conditions". Before the invention of the laptop, projections for current rain forecasts were made. As a result, scientists introduced new ways of developing new ways of developing new ways of developing new ways of developing new ways of developing new ways of developing new ways of developing new ways of developing new ways of developing new ways Today, scientists employ a variety of forecasting methodologies. Meteorology is used all across the world because of its connections to human existence and desires.

Rain greatly impacts our environment and surrounds; As a result, one of the necessary weather forecasting analyses is precipitation prediction. Natural events such as floods and droughts, as well as weather indicators such as ratios, are all severely harmed by precipitation. As a result, reliable and accurate precipitation prediction systems are of relevance. To anticipate quantitative rain, numerous attempts were made.

Rainfall forecast in nature is challenging due to the nature of meteorological knowledge. Because it comprises massive unique disciplines of expertise, the prognosis is advanced. Precipitation is influenced by a wide range of environmental factors, hence the scientific model for it is nonlinear. Numerical climate prediction frameworks have an unbroken run and rely on the method needed to simulate meteoric occurrences using a mesoscale or local forecast. Nonlinear systems now use neural networks in a variety of research since they are a lot more suitable for such systems. Prediction of precipitation is a critical demand for water supply management in nations like the Republic of India. This prediction is useful for government to know prior on droughts or heavy rainfall. The prediction of rainfall with high accuracy is inevitable in countries like India, as our economy majorly depends on agriculture, which in turn needs rainfall.

*1.1 Machine Learning Techniques:*

For an endless number of years, neural networks have been employed to predict weather in various regions throughout the world. All the activities that are available on neural network for the prediction of rainfall is done 15 minutes in advance. This also includes hour base, daily, monthly or annual report generation. The features that are taken into considerations for input are only rainfall values which are used for predictions. TDNN gave the maximum accuracy, among Multiple Neural Network Topologies which were evaluated and studied.

For the prediction of various weather attributes, we have used several Machine Learning Algorithms which includes ANFISs, SVMs, ELM, KNNs, RN & CNN.

- Adaptive Network-Based Inference Systems-ANFISs
- Support Vector Machines-SVMs
- Extreme Learning Machine -ELM
- K-Nearest Neighbours-KNNs
- Regression Trees-RT.

## 1. Related Work

These days, the water assets have been considered as the critical worry for any sorts of advancement program and arranging which incorporates successful water asset the executives, food creation area and flood control. The lopsided designation of water supply all through the country, due to the regular example of precipitation event which changes essentially in reality, is the principle obstacle for the successful water asset the board in India [1]. Factual and AI strategies will be prepared, or fitted, to foresee day by

day precipitation utilizing barometrical indicators. For the investigation, data from 17 downpour measurements will be used, with 80 percent of each station's data being used to train the models (preparation) and the remaining 20% being used to assess their expectation ability (testing). Every precipitation measure's 36-year-long unique time series is divided into two sets: the preparation set, which contains 30 years of data, and the testing set, which contains the excess data. For each station, models are created independently. [2].

For time series forecasting, inquiry, and determination, another technology is used, which was introduced by Box and Jenkins and consists of four procedures. Coming up next are the four stages utilized in the ARIMA model [3]. Step1: In the first stage, a sequence of reactions is identified, which is then used to calculate time series and autocorrelations using proclamation Distinguish Step2: In this step Estimation of the beforehand distinguished factors is done and furthermore, the boundaries are assessed utilizing the assertion ESTIMATE. Step3: This step contains above factors Diagnostics checking and boundaries. Step4: In this step the foreseeing upsides of time series are estimated which are future qualities, utilizing the ARIMA model utilizing the assertion FORECAST. The boundaries utilized in this model are p, d, q which depicts 'p' as the number of slack perceptions, 'd' as the degree of differencing and 'q' as the moving normal request. Plan and execute IoT based flood checking what's more, alarming framework and work out the time it would take for the flood to contact them and gives a period to individuals with the goal that they can empty appropriately. Foresee the seriousness of precipitation by AI models [4].

We use Scientific Python Development Environment scratch pad for dealing with this segment. Reverse Distance Weighting technique is used to get the upsides of all the causal and target factors at the particular city areas from its closest adjoining four network convergences. The range for scaling factor is [0,1], using minimum and maximum Scaler class from the Sklearn library used for pre-processing of data, which is accessible in the SciPy library to keep away from the scaling problem. For scaling, the whole dataset is divided into two parts, ex. preparing and testing. A k-fold cross approval is utilized to assess the expectation expertise of the model. Likewise, seven (k) around equivalent folds are thought of and the model is over and over prepared on the staying six (k−1) folds, and its exhibition is estimated on the excess overlap. This prompts roughly 85% of the information as the preparation set and rests 15% as the testing set. The upsides of the nine causal factors from past five continuous days are at the same time utilized as contribution to foresee the everyday precipitation with a lead season of 1-day to 5days [5].

The streamlining strategy for M1 utilizes cross-approval with tuning. This improvement method gives the best accuracy also, assists with discovering issues with datasets by separating information into some overlay, then, at that point building and testing the models on each overlap. Quantities of information additionally assumed a part in light of the fact that the more info being created, the most

noteworthy consequence of the coefficient of assurance [6]. Straight relapse is the most generally utilized calculation in measurements and AI for assessing the connection between the ward (wanted yield) and the autonomous factors (input boundaries). It centres on the contingent likelihood conveyance of the ward variable given the upsides of the autonomous factors. The centre guideline behind direct relapse is to get the ideal bend that best fits the noticed information [7]. In this task first of all user has to input trainee data for the reference. This data can be in text form which can be duplicated or consisting so many noises, so first user has to clean it. Then the cleaned data gets split for training the dataset.

This data will be train by using machine learning algorithm and then store it as a reference for other input data in future. After that whenever any data will be there as an input then that data will be compared with our trained data by applying algorithms like ANN, CNN etc and predict the result.

*2.1 Loophole of Existing Solutions*

- The existing research does not focus on monthly rainfall data.
- Less accuracy on prediction through machine learning approaches like ANN.

## 2. Proposed Methodology

*2.1 Collection of data*

For the quality data selection and better analysis of the data to get expected output there is a process known as data collection. For this purpose we have used rainfall dataset from the year 1901 to 2018 which are gathered from various regions. The dataset contains ID, region name, year, monthly data of rainfall and sum of annual rainfall data. The role of data analyst is to get the ways out and find sources for the collection of relevant and detailed data, interpret them and analyse the results which are done with the help of certain statistical equations.

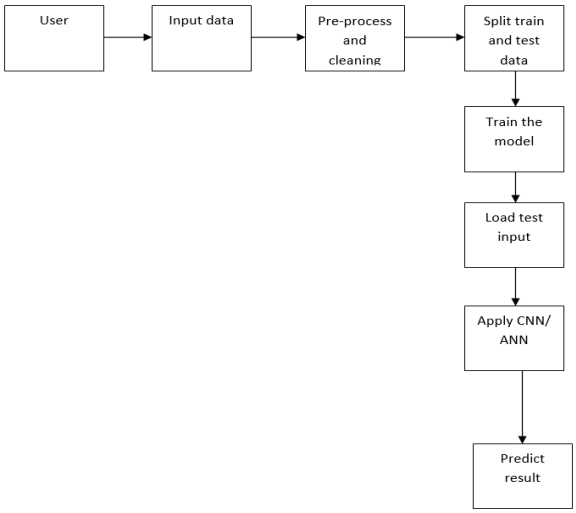


*Fig 1. Rainfall Prediction Model*

*2.2 Pre-processing of data*

Raw data is converted to machine learning suitable data with the process known as data pre-processing. The claimed machine learning model uses structured and cleaned form of data in order to get more accurate results. The methods used for this process is formatting the data, cleaning them and then sampling them. Which also includes Dataset splitting the dataset which is applied for this machine learning algorithm should be divided into 3 subsets which includes training, test, and validation datasets. In training dataset a data scientist defines the optimal parameters of training dataset to be

used from the data along with using training dataset to train a machine learning model. In Testing set there is need to have test set to evaluate the before trained model and also evaluate the capabilities for generation for that trained model. The next step is to identify the patterns in the new data based on the ability of model which are trained over the training data. To solve the problem of incapacity for generalization as given above it is very important to use different data subsets for the training and the testing data model and it's overfitting issue.

*2.3 Data visualization*

The processes like data collection, pre-processing, selection and transportation is done now it's time for visualization. Among all the parameters the responsible one's for rainfall forecasting is monthly rainfall data. All of this information is gathered from meteorologists at the meteorological department centre, who then use various algorithms to predict future weather on a day-by-day, month-by-month, or year-by-year basis.

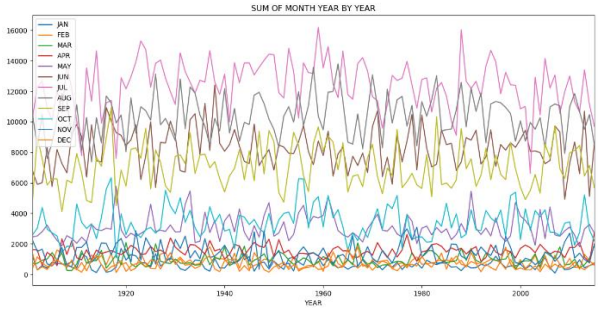


*Fig 2: Data Visualization of Rainfall data Year-wise*

*2.4 Model training*

A model is to be trained when a data scientist has pre-processed the acquired data and divided it into train and test groups. This procedure primarily works with the algorithm's training dataset. An algorithm will analyse data and produce a model that can locate a target value (attribute) in fresh data and provide the answer you're looking for via predictive analysis. Building a good machine learning model is the main focus of model training.

*2.5 Testing dataset and evaluation of model*

The main goal of evaluation and testing the model is to develop the simple form of model which is able to formulate the target value fast and in a accurate form. Model tuning is the one through which data scientist can achieve this goal. The motive is to get better and accurate performance measures for algorithm by improving and enhancing the performance of the given model features.

## 3. Experimental Results & Equations

The algorithm that we have used for this rainfall prediction is performed using CNN. The convolutional neural network is divided into two sections: **The convolution layers** is the one which Extracts features from the input. **The fully connected (dense) layers** is the one which uses data from convolution layer to generate output.


*Fig 3. Convolutional Neural Network (CNN) Architecture*

The training of any neural network involves two fundamental stages, as we explained in the previous section:

1. **Forward Propagation:** This type of propagation receives the input data, it processes the data and precisely gives as output results.
2. **Backward Propagation:** Backward Propagation calculates error that are generated and also updates the features of the neural network.

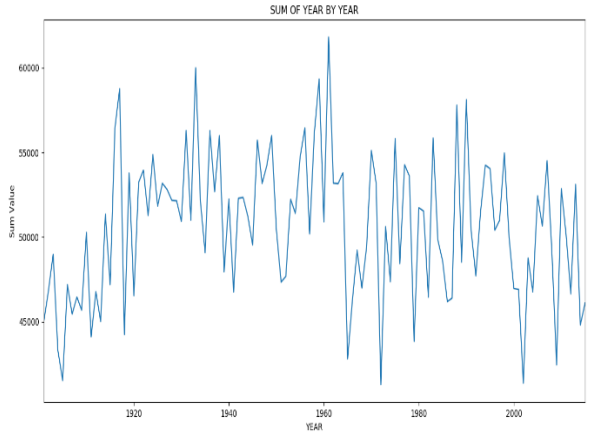


*Fig 4. Line graph for sum of year by year*

Figure 4 represents the line graph for sum of year by year from 1901 to 2018

Figure 5 shows the bar graph for the amount of rainfall in all the subdivision of our country on the monthly basis as this monthly basis.

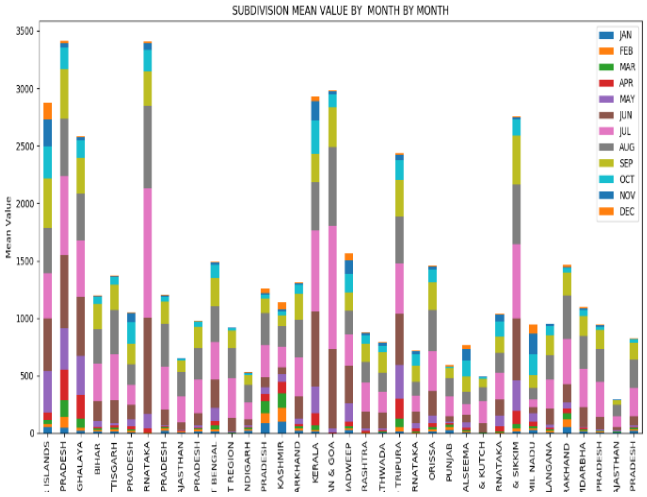


*Fig 5. Bar graphs for the amount of rainfall in all subdivisions on monthly basis*

Figure 6 represents scatter plot for the predictions and the testing datasets that is done on the basis of dataset used from the weather forecast.

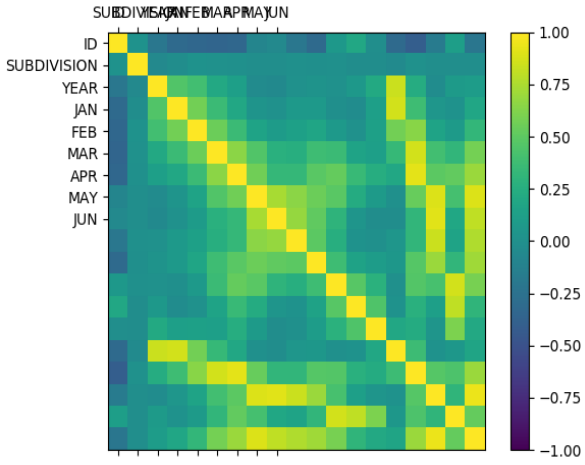


*Fig 6. Scatter plot between the predictions and testing set*

You can calculate the **CNN accuracy** of your model with:

$$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}}$$

Figure 7 represents all the five evaluation parameters of our trained model for rainfall prediction that is MSE, MAE, R-squared, RMSE and Accuracy for the CNN Algorithm Output.

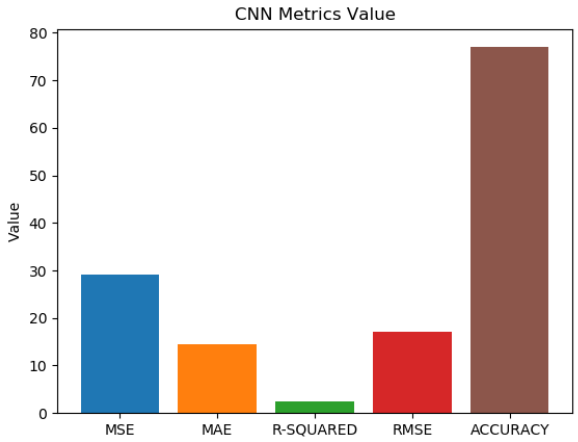


*Fig 7. CNN Metrics Value*

The accuracy from the performed algorithms is shown below with **CNN** giving the maximum accuracy and better performance for the rainfall prediction dataset.

| Algorithm | Accuracy (%) |
|---|---|
| Neural Networks | 76% |
| **Convolutional Neural Networks** | **77%** |

*Table 1. Comparison of accuracy for different algorithms*

## 4. Conclusion & Future Scope

The modern-day state of ML modelling for flood prediction is pretty young and in the early level of development. This paper provides us with the outline of different ML Methodologies used in rainfall prediction and classifies them for the analysis of existing surveys. This Survey focuses mostly on current Machine Learning Models Performance and how much accurate results are being provided. There have been significant research and evaluation for development and improvement in popular machine learning technologies which includes CNN, ANFIS, WNN, ANNs and DT. To improve the prediction quality there are mainly four trends resulted from the past analysis. In the first one it is combination of two or more Machine Learning Technologies or combining a software with a machine learning methodology. The second part includes data extracting part for improvement of the quality of data, whose contribution can improve the prediction accuracy. The third one includes the combination of different methodology which would result in increase of common capability of the models and decrease in uncertain predictions. The fourth would work for the improvement of Machine Learning Algorithms quality and optimizing it. eg. for achievement of better neuronal structures, we use better algorithm CNN. We can expect from these four trends that rainfall forecasting would significantly improve in both ways short-term prediction and long-term predictions. Also, the development of Machine Learning Methodology is highly focused on the correct use of soft-computers technology

and learning algorithms basis on that. Therefore, we can conclude that improvement in machine learning techniques may lead to successful rainfall predictions.

**References**


[1] Bushra praveen, Swapan talukdar, Shahfahad, Susanta Mahato, Jayanta Mondal, Pritee Sharma, Abu Reza Md. Towfiqul islam & Atiqur Rahman, “Analyzing trend and forecasting of rainfall changes in India using nonparametrically and machine l-learning approaches”, Scientific RepoRtS, 2020

[2] Javier Diez-Sierra, Manuel del Jesus, “Long-term rainfall prediction using atmospheric synoptic patterns in semiarid climates with statistical and machine learning methods”, Journal of Hydrology 586 (2020) 124789

[3] CMAK Zeelan Basha, Nagulla Bhavana, Ponduru Bhavya, Sowmya V, “Rainfall Prediction Using Machine Learning & Deep Learning Techniques”, 978-1-7281-4108-4/20/$31.00 ©2020 IEEE

[4] Dola Sheeba Rani, Dr. Jayalakshmi G N, Dr. Vishwanath P Baligar, “Low Cost IoT based Flood Monitoring System Using Machine Learning and Neural Networks”, 978-1-7281-4167-1/20/$31.00 ©2020 IEEE

[5] MOHD IMRAN KHAN, RAJIB MAITY, “Hybrid Deep Learning Approach for Multi-Step-Ahead Daily Rainfall Prediction Using GCM Simulations”, Digital Object Identifier 10.1109/ACCESS.2020.2980977

[6] Wanie M. Ridwan, Michelle Sapitang, Awatif Aziz, Khairul Faizal Kushiar, Ali Najah Ahmed, , Ahmed El-Shafie, ” Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia”,

[8] D. Hudson, A. G. Marshall, Y. Yin, O. Alves, and H. H. Hendon, ``Improving intraseasonal prediction with a new ensemble generation strategy,'' Monthly Weather Rev., vol. 141, pp. 44294449, Dec. 2013.

[9] W. Drosdowsky and M. C. Wheeler, ``Predicting the onset of the north Australian wet season with the POAMA dynamical prediction system,'' Weather Forecasting, vol. 29, no. 1, pp. 150161, 2014.

[10] A. Cottrill et al., ``Seasonal climate prediction in the Pacific using the POAMA coupled model forecast system,'' CACWCR, Australia, Tech. Rep. 048, 2012.

[11] D. Hudson et al., ``ACCESS-S1: The new bureau of meteorology multiweek to seasonal prediction system,'' J. Southern Hemisphere Earth Syst. Sci., vol. 67, no. 3, pp. 132159, 2017.

[12] Z. M. Yaseen, A. El-Shae, O. Jaafar, H. A. Afan, and K. N. Sayl, ``Artificial intelligence-based models for stream-ow forecasting: 2000 2015,'' J. Hydrol., vol. 530, pp. 829844, Nov. 2015.

[13] C.-F. Tsai, Y.-C. Lin, D. C. Yen, and Y.-M. Chen, ``Predicting stock returns by classifier ensembles,'' Appl. Soft Comput., vol. 11, no. 2, pp. 24522459, 2011.

[14] L. Zhang, B. Verma, and D. Stockwell, ``Class-semantic colour-texture textons for vegetation classification,'' in Proc. Int. Conf. Neural Inf. Process. (ICONIP), 2015, pp. 354362.

[15] S. S. Talathi, ``Hyper-parameter optimization of deep convolutional networks for object recognition,'' in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2015, pp. 39823986.

[16] Amir Mosavi , Pinar Ozturk, Kwok-wing Chau "Flood Prediction Using Machine Learning Models"

[17] "Rainfall Prediction Using Machine Learning Algorithms." https://ukdiss.com/examples/rainfall-prediction-machine-learning.php ©2018

[18] M. A. Ahmadi, R. Soleimani, M. Lee, T. Kashiwao, and A. Bahadori, ``Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool,'' Petroleum, vol. 1, no. 2, pp. 118132, 2015.

[19] CMAK Zeelan Basha, Nagulla Bhavana, Ponduru Bhavya, Sowmya V, "Rainfall Prediction Using Machine Learning & Deep Learning Techniques", 978-1-7281-4108-4/20/$31.00 ©2020 IEEE

[20] Javier Diez-Sierra, Manuel del Jesus, "Long-term rainfall prediction using atmospheric synoptic patterns in semiarid climates with statistical and machine learning methods", Journal of Hydrology 586 (2020) 124789 .