

# Rainfall Prediction using Regression Model

J.Refonaa, M. Lakshmi, Raza Abbas, Mohammad Raziullha

**ABSTRACT---** Artificial intelligence is been widely used in all the applications and weather forecasting is not an exception. When it comes to weather forecasting, rainfall prediction is one of the most widely used research areas as numerous lives and property damages occur due to this. Intense rainfall has abundant impacts on society and on our daily life from cultivation to disaster measures. Previous rainfall prediction models that are widely used, makes use of many the complicated blend of mathematical instruments which was insufficient to get a higher classification rate. In this project, we propose new novel methods for predicting monthly rainfall using linear regression analysis. Rainfall predictions are made by collecting quantitative data about the current state of the atmosphere. Numerous machine learning algorithms can learn complex mappings from inputs to outputs, based solely on samples and require limited. Accurate prediction of rainfall is a difficult task due to the dynamic nature of the atmosphere. To predict the future's rainfall condition, the variation in the conditions in past years must be utilized. We have proposed the use of linear regressions by making use of various parameters such as temperature, humidity, and wind. The proposed model tends to forecast rainfall based on the previous records of a particular geographic area, therefore, this prediction will prove to be much reliable. The performance of the model is more accurate when compared with traditional rainfall prediction systems.

**Keywords —** Weather, Linear Regression, Classification, Accuracy, Parameters, Prediction

## I. INTRODUCTION

Immense growth in the research area of Artificial Intelligence has given birth to many new technologies that make use of numerous features of the and some of them are Machine Learning Approaches and Deep Learning Techniques. Machine learning techniques are widely used in numerous applications such as medical image analysis [1], robot path planning [2], flood detection in a particular city or area [3] and land cover classification [4].

Machine learning technique is a process of learning a specific task without any human intervention and improving the performance only by the continuous learning process. The learning is of two types: supervised learning [5], where the labeling is given for the features of the training dataset and unsupervised learning [6], where no labels are given and the system needs to label the features of the dataset. Feature extraction is a vital process in all the machine learning approaches [21,22]. The extracted features could then be used for various other approaches like classification or regression.

Some of the widely used classifiers in machine learning techniques are SVM classifiers [7], Decision Trees [8],

Naive Bayes [9] and Linear Regression [10] and Random Forest Trees [11]. When larger datasets are being used in an application then the use of Artificial Neural Networks (ANNs) is preferred for feature extraction as it produces more accurate results [12]. Use of ANNs is widely called Deep Learning Approaches as the neural network learns each and every layer very deeply and uses the output of a layer as the input of the next layer. ANNs serves as a classifier resembling the function of a biological neuron having numerous layers connected to each other through weights [13].

In image processing, the number of pixels (picture elements) depends upon the input image. The image is the replica of reality which provides as much information as possible about an object. The arrangement of neurons forming layers and the connection patterns formed within and between each layer is called Network Structures. ANNs are information processing structures that can solve any problem through learned examples rather than pre-specified algorithms [14].

In this paper, we have made use of linear regression for classifying the input data and predicting the occurrence of rainfall. The proposed model would serve a model predicting the rainfall beforehand to reduce various impacts on society and to make predictive measures for disaster reliefs. The Linear regression technique is used to classify the images and predict the rainfall. The rest of the section is as follows: Section II consists of Literature Survey, section III consists of the methodology used in the paper and section III consists of various results obtained. The paper is concluded in the last by mentioning the relevant future works that could be applied or added to the proposed work.

## II. RELATED WORK

Numerous research works are proposed by various researchers using various machine learning techniques. Deepak Ranjan Nayak [15] surveyed about rainfall prediction using Artificial Neural Network. Akash D Dubey [16], made a prediction on Pondicherry about the rainfall. A semi-automatic segmentation method proposed by Rui Lu et al. estimated the volume of tumor cells in the liver [15]. The boundaries of the tumor cells were localized from a CT image. Though the process consumed quite a lot of time for computation it claimed to be very efficient in finding the volume of the tumor cells by segmenting it into slices. In [16], Kostas Haris introduced a hybrid image segmentation using a morphological algorithm for watersheds that combined edge and region-based techniques. The technique found to be effective as it reduced the number of false edge detection. Emission of noise served

Revised Manuscript Received on July 10, 2019.

**J.Refonaa**, Assistant Professor, School of Computing, Sathyabama Institute of Science and Technology, Chennai, T.N, India.

**M. Lakshmi**, Principal, Professor, Sri Krishna College of Technology, Coimbatore, T.N, India.

**Raza Abbas**, Student, School of Computing, Sathyabama Institute of Science and Technology, Chennai, T.N, India.

**Mohammad Raziullha**, Tamilnadu, India

to a part for indirectly increasing the processing time for the computation. FanmanMeng [17], designed an efficient and robust supervised image co-segmentation model comprising of a strategy called color reward and an active contour model. The model was evaluated on numerous images from a database and could efficiently pair the common objects with minimal error rate. Various research works have been carried out for weather forecasting too. Sharma et al. [18] have used machine learning techniques to predict solar generation from weather forecasts. In [19], the rain prediction was done using various classifiers to detect the amount of rainfall in a particular area.

III. PROPOSED APPROACH

The proposed system in the proposed system makes use of Linear Regression. In this paper, we have made use of the Chennai dataset. It is given as the input and the file is read by the toolbox. The dataset is then converted into separate data text where the 1<sup>st</sup> year's plot of rainfall is done. The dataset is configured by making use of Linear Regression where various parameters are given to initiate the entire network. The prediction values are measured by making use of the toolbox and the graphs are plotted for the obtained predicted values. The entire block diagram of the proposed model is shown in Fig. 1.

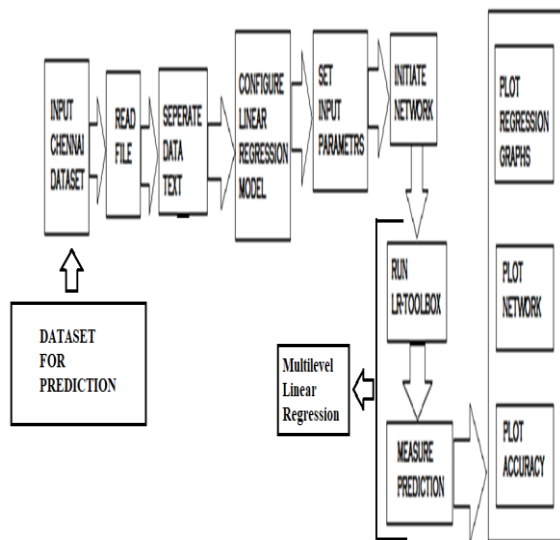


Fig. 1 Block Diagram of Proposed Model

When it comes to Linear Regression model, the rainfall dataset is used as the input where the pre-processing stage happens. The feature is extracted by making use of the Linear Regression Model.

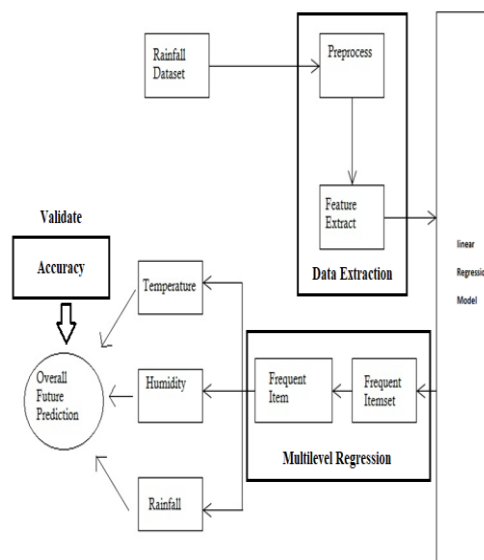


Fig. 2 Architecture of Proposed Model

The values or item set that occurs frequently in the model is identified and then separated for getting the list of the frequent items. Various parameters are obtained while using these model such as the temperature, the humidity and the rainfall of a given geographic area. The entire architecture and the flow of the proposed model are shown in Fig. 2.

Predicted values

$$\begin{aligned}
 & \text{minimize } \frac{1}{n} \sum_{i=1}^n (pred_i - y_i)^2 \\
 J &= \frac{1}{n} \sum_{i=1}^n (pred_i - y_i)^2
 \end{aligned}$$

We have used this expression for calculating the predicted values, firstly, we have get the difference value, then doing square for the difference. There are some values are not satisfied the system on getting the difference value. This type of operation was named as mean squared error.

Multilevel linear regression expression is

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \epsilon_i, \quad i = 1, \dots, n,$$

Here,

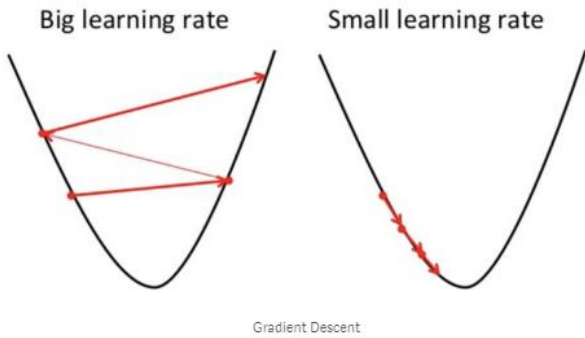
$y_i$  -  $i$ th response.

$\beta_k$  -  $k$ th coefficient, where  $\beta_0$  is the constant term in the model.

$X_{ij}$  -  $i$ th observation

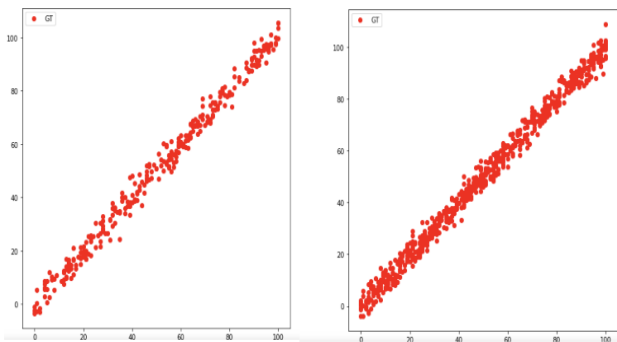
This multilevel expression has many types' of regression types for predicting the good values. This system has also the system prediction scheme for the multilevel operations and valuation. By using scheme, we have achieved the predicted values for rainfall and implemented as automated in the proposed system.





**Fig. 3 Learning Rate**

In Fig. 3, we can see the difference between the big and small learning rate, in the first difference, we can see the zig zag conversation happens on the first model for the larger dataset in this model. If there are more difference in the prediction values, then it will be shown as graph in our proposed system. In the second difference, we can see the smaller difference in the predicted values and the difference values. It is very fast to reach the predicted values in the rainfall. Rainfall is predicted based on linear regression for showing the bigger difference between the areas.

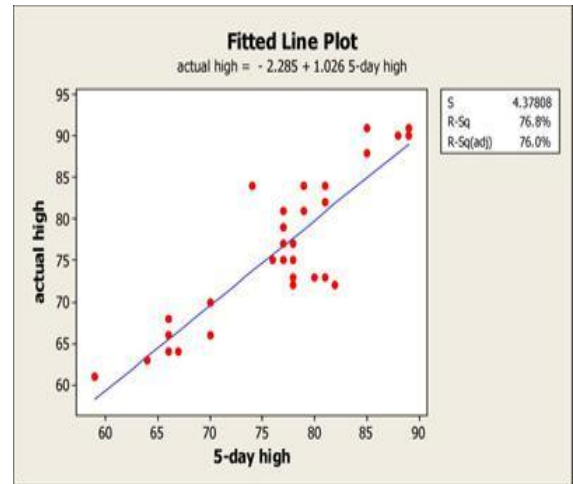


**Fig. 4 Difference of larger rate**

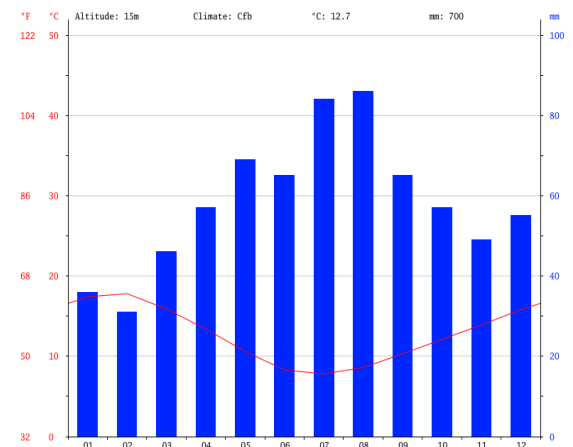
In Fig.4, we can see the smaller prediction and the large prediction. There some system is implemented in the proposed system. Some system is very unique, namely graph representation, it will show the values in diagrammatic view, so that user or admin can choose the correct place for prediction in future, it also automated for predicting the rainfall values.

**IV. EXPERIMENTAL RESULTS**

The experimental results were obtained for the proposed model by making use of Matlab. The dataset that was used to train the model consisted of previous datasets of rainfall in Chennai district. It also consisted of various other weather-related information of a city for a period of time.



**Fig. 5 Linear Regression Curve of Proposed Model**



**Fig. 6 Predicted values of the Rainfall Prediction**

Linear regression model proposed in the paper has predicted the rainfall into various classes by classifying it as shown in Fig. 5. It can be noted that the areas with rainfall are separated from the areas that do not have rainfall. In Fig. 6, the accuracy of the proposed model is shown. From the graph it can be observed that the prediction of the rainfall in Chennai city is done predominantly with very less misclassification. The proposed model also serves to be highly efficient when compared to other systems.

The model made use of Linear regression to classify and predict the rainfall. The Computer Vision toolbox present in the MAT Lab was used to generate the entire model. Once the model is created using various tools in the software, the dataset was loaded to perform the training. The training was given on the model where the features were extracted. The performance of the network was very accurate and the results are shown in the following figure. The model was able to efficiently detect the rainfall prediction for a specific given time.

**V. CONCLUSION**

Rainfall plays a vital role in all aspects. It is better to predict future rainfall and take if any necessary steps need to

be taken to minimize the amount of life loss. In our proposed work, we have proposed a model that could predict or forecast the rainfall of any particular location where the classifier is trained with numerous previous weather information. The classification was performed using Neural Networks using Linear Regression algorithms. The proposed model had good accuracy. The algorithm extracted the features of the dataset from the training set and was able to accurately predict the weather well in advance. Various other parameters such as efficiency and computational time for the process was also considered which was found to be better than the traditional methods. The future works may include the use of other deep learning techniques and also working on the security concerns of the model.

REFERENCES

1. Rahman, M. M., Bhattacharya, P., & Desai, B. C. (2007). A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Transactions on Information Technology in Biomedicine*, 11(1), 58-69.
2. Morales, M., Tapiá, L., Pearce, R., Rodríguez, S., & Amato, N. M. (2004). A machine learning approach for feature-sensitive motion planning. In *Algorithmic Foundations of Robotics VI*(pp. 361-376). Springer, Berlin, Heidelberg.
3. Ireland, G., Volpi, M., & Petropoulos, G. P. (2015). Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: a case study from a Mediterranean flood. *Remote sensing*, 7(3), 3372-3399.
4. Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, 23(4), 725-749.
5. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160, 3-24.
6. Cheriyyadath, A. M. (2014). Unsupervised feature learning for aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1), 439-451.
7. Bauer, S., Nolte, L. P., & Reyes, M. (2011, September). Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 354-361). Springer, Berlin, Heidelberg.
8. Liu, K., Tong, M., Xie, S., & Zeng, Z. (2014, August). Fusing decision trees based on genetic programming for classification of microarray datasets. In *International Conference on Intelligent Computing* (pp. 126-134). Springer, Cham.
9. Dumitru, D. (2009). Prediction of recurrent events in breast cancer using the Naive Bayesian classification. *Annals of the University of Craiova-Mathematics and Computer Science Series*, 36(2), 92-96.
10. Chou, Y. H., Tiu, C. M., Hung, G. S., Wu, S. C., Chang, T. Y., & Chiang, H. K. (2001). Stepwise linear regression analysis of tumor contour features for breast ultrasound diagnosis. *Ultrasound in medicine & biology*, 27(11), 1493-1498.
11. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18-22.
12. Sarhan, A. M. (2009). Cancer classification based on microarray gene expression data using DCT and ANN. *Journal of Theoretical & Applied Information Technology*, 6(2).
13. Steven K. Rogers, Dennis W. Ruck, Matthew Kabrisky, *Artificial Neural Networks for early detection and diagnosis of cancer*, Elsevier Scientific Publishers, pp. 79-83, 1994.
14. Md. BadrulAlam Miah, Mohammad Abu Tousuf, *Detection of Lung Cancer from CT Image Using Image Processing and Neural Network*, IEEE, In *Proceedings of 2nd Int'l Conference on Electrical Engineering and Information & Communication Technology*, 2015.
15. Nayak, D. R., Mahapatra, A., & Mishra, P. (2013). A survey on rainfall prediction using an artificial neural network. *International Journal of Computer Applications*, 72(16).
16. Dubey, A. D. (2015). Artificial neural network models for rainfall prediction in Pondicherry. *International Journal of Computer Applications*, 120(3).
17. Lu, R., Marziliano, P., & Thng, C. H. (2006, January). Liver tumor volume estimation by semi-automatic segmentation method. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*(pp. 3296-3299). IEEE.
18. Haris, K., Efstratiadis, S. N., Maglaveras, N., & Katsaggelos, A. K. (1998). Hybrid image segmentation using watersheds and fast region merging. *IEEE Transactions on image processing*, 7(12), 1684-1699.
19. Meng, F., Li, H., Liu, G., & Ngan, K. N. (2013). Image cosegmentation by incorporating color reward strategy and an active contour model. *IEEE transactions on cybernetics*, 43(2), 725-737.
20. L.Mary Gladence, Karthi, M., Ravi (2016), T "A novel technique for Multi-class ordinal regression- APDC" *Indian Journal of Science and Technology* Vol.9/ No.10/1-5
21. Ancy, S., Kumar, R., Asokan, R., Subhashini, R., "Prediction of onset of south west monsoon using multiple Regression", *Proceedings of ICCCS 2014 - IEEE International Conference on Computer Communication and Systems*, 2014.
22. Mathew, S., Saravanan, M. Improving an automation reference tool GRASP using active data handling approach , *Proceedings of 2015 IEEE 9th International Conference on Intelligent Systems and Control*, ISCO.

