

Survey on Rainfall Prediction System

¹Shravani Kadam, ²Dhanesh Salgaonkar, ³Bhishmesh Chaudhari, ⁴Prof. Santosh Tamboli

^{1,2,3}UG Student, ⁴Assistant Professor, Vidyalankar institute of Technology, Mumbai, India,

¹shravani.kadam@vit.edu.in, ²dhanesh.salgaonkar@vit.edu.in, ³bhishmesh.chaudhari@vit.edu.in,

⁴santosh.tamboli@vit.edu.in

Abstract — Rainfall prediction is one of the challenging and uncertain tasks which has a significant impact on human society. Agriculture, which acts as major economic sector for various countries and is being the livelihood of many people, is heavily dependent upon the rainfall received by the region. Apart from the agriculture sector, there are many other professions and lives of people that are dependent on rainfall and can be severely affected by the damage caused by it. Also, due to untimely and irregular rainfall pattern observed around the world, it has become difficult to predict rainfall even by high end forecasting systems. Hence, accurate and precise rainfall prediction should be made that can help in understanding & analyzing rainfall patterns & can provide prior information on the rainfall expected in the regions. In this paper we shall see how various other rainfall systems are already developed using machine learning that are able to predict the rainfall upto a certain accuracy with the help data of the data obtained. We shall also learn about which machine learning algorithms and methodologies they have used for the systems. Names of the methodologies include artificial neural network, convolution neural network, linear regression, multiple linear regression, support vector machine classification, random forest, catboost, xgboost. We will compare all those algorithms one by one considering parameters which come under machine learning such as accuracy.

Keywords — accuracy, catboost, machine learning algorithms, prediction, rainfall, xgboost

I. INTRODUCTION

Rainwater is a sustainable resource which plays a useful but vital role in agricultural division of a country. Extreme rainfall can cause crop corruption mainly decreasing crop yields leading to long term farming losses. But, judging from positive aspects, emergence of some drizzles is necessary to meet people's need of food manufacturing. Rainfall prediction is considered among the challenging and non-linear tasks which can have a significant effect on us. Intime and perfect predictions can help to decrease human and economic loss. As we know heavy rainfall can cause to many disasters, predicting rainfall can result very vital. Efficiency of rain forecasting has great influence for countries like India whose economy is broadly reliant on farming. Because of dynamic nature of atmosphere, Statistical techniques cannot provide good accuracy for rainfall forecasting. Uncertainty of rainfall data makes ML and Artificial Intelligence a better approach. The forecasting helps people to take precautionary measures. Precipitation forecasting remains a severe matter and has gathered the attention of industries, risk management entities. It is a climatic aspect that affects numerous human actions like, construction, forestry and many more. Looking towards all these factors we can say that rainfall prediction is essential because this is the one having the most correlation with negative natural events such as flooding, mass movements

and landslides. These have adversely affected civilization for years . Therefore, having an fitting way for rainfall prediction makes it feasible to take precautionary and mitigation measures.

Scientists across the globe have come up with various models to forecast the rainfall mostly using random numbers which are relevant to climatic data. These models are hard to frame due to dynamic differences in meteorological processes, resulting in it making one of the challenges for the atmospheric department despite much improvement in weather forecasting lately. In this paper we will examine various papers from different authors and see what all techniques methods they have used.

II. LITERATURE SURVEY

A. Kala and Dr. S. Ganesh Vaidyanathan in [1] have used Artificial Neural Network for implementing their idea. Their procedure for rainfall prediction includes gathering the weather data, then preprocessing it, building the Feed Forward Neural Network (FFNN) model with training data and then validating it with testing data and in the end evaluating the model by comparing desired and actual output. R. Kingsy Grace and B. Suganya in their paper has said that they have implemented their model using machine learning [2]. They have compared various models such as Deep Convolutional Neural Network, Genetic

Programming, ANN, Linear Regression, Hybrid Neural Network, Likelihood, LSTM and ConvNet. Dataset they have used for the prediction had 4116 number of rows.

Hiyam Abobaker Yousif Ahmed and Sondos W. A. Mohamed in [3] have implemented their model using linear regression. Their whole idea is done in 2 major parts i.e., first is data collection and selection and second is data cleaning and transformation. They have used MLR model for precipitation forecasting for Khartoum state dependent on respective dataset. Dataset includes attributes like maximum and minimum temperature, dewpoint, wind speed, etc. After implementing their model, they got to know that the mean square error between real and desired values has been convincingly diminished while implementing. CMAK Zeelan Basha, Nagulla Bhavana, Ponduru Bhavya, Sowmya V. has implemented a model on the same topic using machine learning and deep learning techniques [4]. They have taken the help of Auto-Regressive Integrated Moving Average (AREMA Model), Artificial Neural Network, Support Vector Machine and Self Organizing Map. Also, two major deep learning methods used while implementing the models are Auto Encoders (AE) and Multilayer Perception (MP). AE has 3 layers which are Input layer, Hidden layer and Output layer. MP takes the input and forecasts subsequent information from the prior information. They have also used Neural Network Training tool. They have also listed various methodologies to anticipate the shower and problems that could be seen using various methods to anticipate shower in their paper.

B. Vasantha, R. Tamilkodi, L. Venkateswara Kiran in [5] have forecasted the rainfall applying actual time global climate parameters. Their proposed system algorithm consists training the mode, evaluating and optimizing the model with test set, 4-fold cross validation of the model with blind set, minimizing cost function, feeding recent temperature values, and anticipating temperature values for tomorrow. They have mainly used convolution neural systems to anticipate weather parameters which in result will give meaningful designs to understand the forecasting. Anjali Samad, Bhagyanidhi, Vaibhav Gautam, Piyush Jain, Sangeeta, Kanishka Sarkar has done precipitation forecasting applying Long Short Term Memory Neural Network [6]. They have used data study and processing, seasonal decomposition, artificial neural network, recurrent neural network, long short term memory, making of the actual model, model evaluation. They have also shown the data analysis and model prediction results. This model is based on Australian dataset which covers seasonal decomposition methods.

Rose Ellen N. Macabiog, Jennifer C. Dela Cruz have done the model building on rainfall predictive approach applying classification [7]. In classification data is grouped into given number of classes. They have done data collection, data preprocessing, building the predictive

model, model evaluation and selection and finally testing and evaluation. Choosing proper evaluation metrics is important while selecting the model algorithm. They also think that this study may be added in smart flooding system of the government. Arief Bramanto Wicaksono Putra, Rheo Malani, Bedi Suprpty, Achamad Fanany Onnilita Gaffar have implemented Deep Auto Encoder using Semi CNN [8]. Their method of working includes Auto-Encoder Neural Network (AENN), Convolution Neural Network (CNN), proposed method, model data time series autoregressive (AR model), dataset. Proposed method consists of two phases, one is Deep Auto Encoder Semi Convolution Neural Network (DAESCNN) and second is training strategy. DAESCNN furthermore consists Forward process, Net error and error function, Backward process.

Yuana Ratna Sari, Esmeralda Contessa Djamal, Fikri Nugraha in [9] have done the precipitation forecasting using 1-D CNN. They have done preprocessing and convolutional neural network (CNN). Preprocessing consists of data loss treatment, segmentation, normalization. CNN has two layers namely feature extraction layer and prediction layer. Furthermore, feature extraction layer consists convolutional layer, Rectified Linear Unit Activation process and Max pooling layer. Eslam Hussein, Mehrdad Ghaziasgar, Christopher Thron in [10] have applied Support Vector Machine Classification. The methodology used consists data gathering, preprocessing, classification, comparison between different SVM inputs, comparison between regional predictions. They have used images of maps and it is the backbone of the model. The whole Australian region is divided into 5*5 matrix area and data is collected accordingly and is further manipulated for the prediction.

Nikhil Oswal in [11] has used machine learning techniques to implement his model. His methodology includes data exploration and analysis, data preprocessing, modelling and evaluating. He also has applied the data to different classifiers of different family which includes decision tree, logistic regression, K-nearest neighbour, decision table, random forest, adaboost, gradient boosting. In the evaluation phase he checked accuracy, area under curve, precision, recall, f1 score, confusion matrix, stratified k-fold, statistical testing. He also supplied original dataset, undersampled dataset, oversampled dataset to the model.

III. FINDINGS AND ANALYSIS

A. Kala and Dr. S. Ganesh Vaidyanathan in [1] observed that there was 93.55% accuracy using Artificial Neural Network. Hence, they were sure that FFNN model can be used for the rainfall prediction. They have also said that in future, model accuracy can be made better by implementing various algorithms and increasing the dataset. R. Kingsy Grace and B. Suganya found that there was 99% accuracy of the Multiple Linear Regression. They have concluded their paper saying their proposed model forecast the rainfall for Indian dataset providing better results

considering the factors that comes under machine learning models [2]. Hiyam Abobaker Yousif Ahmed and Sondos W. A. Mohamed in [3] have seen that there was 85% accuracy given by the Multiple Linear Regression model.

CMAK Zeelan Basha, Nagulla Bhavana, Ponduru Bhavya, Sowmya V. has implemented a model on the same topic using machine learning and deep learning techniques [4]. The outcomes intend that in terms of MSE and RMSE, their architecture outruns other approaches. B. Vasantha, R. Tamilkodi, L. Venkateswara Kiran in [5] have forecasted the rainfall applying actual time global climate parameters. They believe that the results precision is expected to be extended by 70%. Anjali Samad, Bhagyanidhi, Vaibhav Gautam, Piyush Jain, Sangeeta, Kanishka Sarkar have seen that their model has been forecasting accurately for check cases but there was some increase in error rate while handling the outliers. Otherwise, the model works perfectly fine [6].

Rose Ellen N. Macabiog, Jennifer C. Dela Cruz have observed that we can get most accurate results if we use all 5 attributes for the model. Course KNN, Fine Gaussian SVM, Neural Network given the best 0.811 accuracy [7]. Arief Bramanto Wicaksono Putra, Rheo Malani, Bedi Suprpty, Achamad Fanany Onnilita Gaffar have found that this approach is one of the recent ones for rainfall prediction with 99% performance [8]. When Yuana Ratna Sari, Esmeralda Contessa Djamal, Fikri Nugraha checked the 1-D CNN model for 14 days, they observed that there was 0.9463 accuracy of training data and 0.8146 accuracy of testing data. Also, they configured that more the layers you use in the model, more is the accuracy [9].

Eslam Hussein, Mehrdad Ghaziasgar, Christopher Thron in [10] are able to point out few of the drawbacks and potential difficulties using SVMs for climate forecasting up to a month in advance. Nikhil Oswal, after applying several preprocessing steps and understanding their impact he has concluded that Australian rainfall is not certain there is not a particular relation between time and shower. Still, he is able to find certain pattern and develop high performance model [11].

After analyzing, we can state that the MLR and the deep learning models such as CNN, ANN perform better and yield better results than the other ML models.

IV. CONCLUSION

After understanding and analyzing the papers and their findings, we can deduce that it is possible to effectively predict Rainfall by using machine learning techniques. The selection of the best machine learning algorithm out of the many, for a particular dataset, depends upon the nature of the data contained, attributes considered and the size of the data. The best suited ML algorithm, by considering the above factors will provide us with the best accuracy, AuC, ROC score. To further improve the accuracy, data

preprocessing and cleaning techniques can be applied which will help in achieving precise result for the implemented model.

REFERENCES

- [1] Kala, A., & Vaidyanathan, S. G. (2018, July). Prediction of rainfall using artificial neural network. *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*.
- [2] Grace, R. K., & Suganya, B. (2020, March). Machine Learning based Rainfall Prediction. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*.
- [3] Ahmed, H. A. Y., & Mohamed, S. W. A. (2021, February 26). Rainfall Prediction using Multiple Linear Regressions Model. *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEE)*.
- [4] Basha, C. Z., Bhavana, N., Bhavya, P., & V, S. (2020, July). Rainfall Prediction using Machine Learning & Deep Learning Techniques. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*.
- [5] Vasantha, B., Tamilkodi, R., & kiran, L. V. (2019, March). Rainfall pattern prediction using real time global climate parameters through machine learning. *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (VITECoN)*.
- [6] Samad, A., Bhagyanidhi, Gautam, V., Jain, P., Sangeeta, & Sarkar, K. (2020, October 30). An approach for rainfall prediction using long short term memory neural network. *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*.
- [7] Macabiog, R. E. N., & Dela Cruz, J. C. (2019, November). Rainfall Predictive Approach for La Trinidad, Benguet using Machine Learning Classification. *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*.
- [8] Wicaksono Putra, A. B., Malani, R., Suprpty, B., & Onnilita Gaffar, A. F. (2020, July). A deep auto encoder semi convolution neural network for yearly rainfall prediction. *2020 International Seminar on Intelligent Technology and Its Applications (ISITIA)*.
- [9] Sari, Y. R., Djamal, E. C., & Nugraha, F. (2020, September 15). Daily rainfall prediction using one dimensional convolutional neural networks. *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)*.
- [10] Hussein, E., Ghaziasgar, M., & Thron, C. (2020, July). Regional rainfall prediction using support vector machine classification of large-scale precipitation maps. *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*.
- [11] Oswal, N. (2021). *Predicting Rainfall using Machine Learning Techniques*. Institute of Electrical and Electronics Engineers (IEEE).