iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35032

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Machine Learning Technique for Rainfall Prediction

Dr. Vivek Waghmare[1], Miss. Nikita Patil[2], Miss. Purva Mahajan[3], Mr. Chetanya Patil[4], Mr. Shivansh Thakur[5]
[1, 2, 3, 4, 5]*Sandip Institute of Technology and Research Centre, Nashik, India*

Abstract: ***Rain prediction is one of the most challenging and uncertain tasks that has a profound effect on human society. Timely and accurate forecasting can help significantly reduce population and financial losses. This study presents a collection of tests involving the use of conventional machine learning techniques to create rainfall prediction models depending on the weather information of the area. This Comparative research was conducted focusing on three aspects: modeling inputs, modeling methods, and prioritization techniques. The results provide a comparison of the various test metrics for these machine learning methods and their reliability estimates in rain by analyzing weather data. This study seeks a unique and effective machine learning system for predicting rainfall. The study experimented with different parameters of the rainfall from various regions in order to assess the efficiency and durability of the model. The machine learning model is focused on this study. Rainfall patterns in this study are collected, trained and tested for achievement of sustainable outcomes using machine learning models. The monthly rainfall predictions obtained after training and testing are then compared to real data to ensure the accuracy of the model The results of this study indicate that the model has been successful in it predicting monthly rain data and specific parameters.***

Keywords: Rainfall Prediction, Correlation Based Feature Selection, Machine Learning, Modeling.

I. INTRODUCTION

Rain plays a vital role in shaping the animals and plants of natural health. It is not just significant for the human beings but also for animals, plants and all living things. It plays a significant role in agriculture and farming and undoubtedly; water is one of the most natural resources on earth. Changing weather conditions and rising temperature exclusion has made it difficult for humans and the planet earth to feel the required rain needed to meet human needs and its uninterrupted use in daily life .that is required to satisfy the human needs and its uninterrupted use in everyday life. Therefore, it is very important to analyze the changing patterns of rain and try to guess rain not only for human needs but also for predicting natural disasters that could be caused by unexpected heavy rains. To be more specific and aware of the devastating climatic changing and stay

Updated predicting rainfall has been the focus of computer scientist and engineers. In India, where the majority of agribusiness is dependent on precipitation as its standard wellspring of water, the time and measure of precipitation hold high importance and can impact the entire economy of the nation. Climate plays a vital role in our everyday life. From the earliest starting point of the human development, we are occupied with thinking about climatic changes. Weather forecasting is one of the most challenging issues seen by the world, in a most recent couple of centuries in the field of science and technology. Prediction is the phenomena of knowing what may happen to a system in the near future. Present weather observations are obtained by ground-based instruments and from the satellite through remote sensing. As the Indian economy relies heavily on horticulture, rain plays an important role. Rainfall is one of the most significant and complex weather conditions. It is as complex as it involves the interaction of several space systems and is essential for survival and ground support. Forecasting the rainy season is one of the most important issue from many real applications such as water resource planning and management [1], operation of lakes and floods prevention [2, 3], rain-fed agricultural activities, hydropower generation and drinking water supply [4, 5].

Due to the complexity of the atmospheric processes that create rainfall and the lack of available data for the time required as well local scales, it is usually not possible to predict rain using a physically based process model [2]. However, with the advent of a soft computer Strategies, neural network (ANN) implant methods have been used in rain forecast.

II. LITERATURE SURVEY

Rainfall prediction is not an easy job especially when expecting the accurate and precise digits for predicting the rain. The rainfall prediction is commonly used to protect the agriculture and production of seasonal fruits and vegetables and to sustain their production and quality in relation to the amount of rain required by them (Lima & Guedes, 2015). The rainfall prediction uses several networks and algorithms and obtains the data to be given to the agriculture and

production departments. The rainfall prediction is necessary and mandatory especially in the areas where there is heavy rainfall and it's more often expected (Amoo & Dzwairo, 2016).

There are huge economies like those of Asia like India and China that that earn a large proportion of their revenue from agriculture and for these economies; rainfall prediction is actually very important (Darji, Dabhi, & Prajapati, 2015). The rainfall forecasting is prevailing as a popular research in the scientific areas in the modern world of technology and innovation; as it has a huge impact on just the human life but the economies and the living beings as a whole. Rainfall prediction with several Neural Networks has been analyzed previously and the researchers are still trying hard to achieve the more perfect and accurate results in the field of rainfall prediction (Biswas, et al., 2016). The prediction of seasonal rainfall on monthly basis by using the surface data to form annual prediction is also essential for the agricultural activities and therefore the production and supervision of the agriculture and crops. It could be done by recognizing the variations in the supply of moisture in the air. The case of African region illustrates that how this succeeded and how West Africa advantaged from the rainfall prediction in managing their agricultural activities (Omotosho, Balogun, & Ogunjobi, 2000). Similarly, the short-term streamflow forecasting for the rainfall is also reliable and bias-free. But they are not much effective in predicting the flood and post-processing of rainfall prediction. An approach called raw numerical weather prediction (NWP) was introduced in 2013, where the approach focused on the Bayesian joint probability model to formulate prediction data. The approach formed forecast possibility distributions for each location and it had prediction time for it; collaborative forecasts correlated with space and time was produced in the Southern part of
Australia (Khan, Sharma, Mehrotra, Schepen, & Wang, 2015). This approach focused on Schake shuffle to produce the forecast by the forecast possibility distributions (Robertson, Shrestha, & Wang, 2013). Furthermore, the short-term streamflow forecasting could also be used through the artificial neural networks as researched by Zealand, Burn and Simonovic in 1999. The study conducted outlined that ANNs ability to forecast for short-term stream flow and outlined some of the issues that the approach encountered with ANNs (Kumarasiri & Sonnadara, 2006). Although, ANNs with short-term stream flow can calculate and present complex and nonlinear relationship between input and output with an ability to outline the interface effect as well but has issues in processing some input data with certain type and number. The ANNs also encountered difficulty with dimensions of the hidden layers. This research outcome was represented by the data of Winnipeg River system in Ontario, Canada using the quarter monthly data. The outcomes of the study were encouraging with AANs performed quite well for the four prediction lead-times. The RMSE for the test data of 8 years outlined variation from 5cms to 12.1cms in a forecast from four-time step to two-time step ahead respectively (Zealand, Burn, & Simonovic, 1999). Also, the recent decade highlighted the significance of artificial intelligence and it has gained attention in water resource management and engineering as well. ANNs, ANFIS and GP are the driving simulations of AI and they are advantaged over other systems and approaches because of being more reliable and competitive. The adaptive neuro-fuzzy inference system (ANFIS) for time series and ANN for predicting streamflow in Apalachicola River, the United States with that of other neural network techniques like hybrid (Mittal, Chowdhury, Roy, Bhatia, & Srivastav, 2012); when compared to wavelet-gene expression' programming approach outlined the following results; ARMA model predicting accurate results for 1 day ahead time whereas, ANFIS forecasted the results for 2 days ahead time. The results from AI using ANFIS were more accurate and could predict 2 days ahead of time data rather than GEP and ANN (Nayak, Mahapatra, & Mishra, 2013). But for the 3 days forward data; ANN performed better than other models. For the monthly data; ANN, ANFIS and GEP outperformed as compared to ARMA models in the first part of the study (Karimi, Shiri, Kisi,&Shiri,2016). Water as is one of the most useful resources of the earth. There is no human and living thing on earth that can survive without water. As, this precious resource is running out because of the increasing temperature of the earth and the unexpected and unappreciated climatic conditions due to global warming. (Mittal, Chowdhury, Roy, Bhatia, & Srivastav,2012). In addition, the comparison among different neural models revealed that Non-linear autoregressive exogenous networks (NARX) and back propagation neural BPN) performed better than distributed time delay neural network (DTDNN) cascade-forward back propagation neural network (CBPN) in outlining more accurate and precise results for rainfall prediction (Devi, Arulmozhivarman, Venkatesh, & Agarwal, 2016). In comparison, statistical forecasting methodology can also be used for the rainfall prediction that outlines by using two different approaches like traditional linear regression and polynomial-based nonparametric; where nonparametric method outlined more competing results. Both the approaches could predict the 1-3 monthly rainfall forecasting data that could actually impact water resource planning and controlling (Singhrattna, Rajagopalan, Clark, & Kumar, 2005). The periodic and episodic rainfall data for the south-west peninsula of England has also exposed that atmospheric characteristics are key players of outlining the monthly and seasonal forecast (Mcgregor & Phillips, 2003). The rainfall prediction is also emphasized for its significance for the prediction of flood and consequently takes the precautionary measure to save the people from devastating destructions that a flood can cause (Hoai, Udo, & Mano, 2011). There are studies that outlined the significance of rainfall prediction in forecasting flood on the regions where there is heavy rain every year.

The areas with high risk for flood are the vulnerable areas that need the rainfall forecasting not just to save a human life but to safe agriculture, water reservation and livestock (Fang & Zhongda, 2015). In comparison, the significance of rainfall prediction is also important for areas with high probability for the drought. The areas with high drought seasons are also vulnerable to high risk in terms of agriculture and livestock with an extreme threat to human life as a whole; the study conducted for Sakae River basin of Thailand (Wichitarapongsakun, Sarin, Klomjek, & Chuenchooklin, 2016). The artificial neural network model for rainfall prediction of 1to 6 hour ahead time is studied for Bangkok, Thailand by Hung, Babel, Weesakul, and Tripathi in 2008.

The study outlined that within artificial neural networks, using six models utilizing rainfall parameters like humidity, air pressure, wind direction and wind speed can give more accurate and precise prediction when previous forecasting data is also used with these parameters as an input as well (Hung, Babel, Weesakul, & Tripathi, 2009). Nevertheless, land sliding is another natural hazard that could be caused due to heavy rainfall. The rainfall prediction could assist in combating the devastation caused by land sliding. The rainfall prediction for the areas vulnerable to land sliding is an essential part of artificial intelligence within engineering and management fields (Schmidt, Turek, Clark, Uddstrom, & Dymond, 2008). The metrological and hydrological centers are struggling hard to produce the more competitive and precise rainfall prediction in order to overcome these issues that the rainfall can cause and their efforts have marked quite an improvement in the rainfall prediction and forecasting data for many models using the neural networks. The prediction for extreme rainfalls is useful for not just the metrological departments in sharing in time alerts but also for the hydrological departments in order to form better safety measures for example the flood prediction in Australia (White, Franks, & McEvy, 2015). The rainfall prediction systems are much popular with artificial neural networks and the rainfall prediction departments like the metrology and hydrology engineering with management (Abhishek, Kumar, Ranjan, & Kumar, 2012). The rainfall prediction using the neural network aims at predicting more efficient and more accurate results and precise predictions for a more useful and reliable output that could be used by the management and engineering departments in designing the plans and policies that will not only increase efficiency but it will also enhance the management systems from a quality data produced by using the Artificial Neural Networks. The study conducted with the different networks highlighted different results by operating within same training functions and outlined that back propagation neural network is capable of obtaining more precise predictions. Also, that increased neurons can decrease errors (MSE) Sharma & Nijhawan, 2015). Neural networks have proved capability for the rainfall prediction and in obtaining accuracy with precision among the other networks with other modelling techniques (Narvekar & Fargose, 2015)

III. PROBLEM DEFINITION

An accurate and accurate forecast of rainfall is still needed which can help in various fields such as agriculture, water storage and flood forecasting. The problem with building rain statistics a prediction that would be based on previous findings and similarities and that would provide predictable output i.e. reliable and appropriate. A vague and inaccurate prediction is not only a waste of time but also a loss of resources and lead to the effective management of problems such as poor agriculture, poor water resources and poor management of these floods. Therefore, the need is not only to build a rain forecast system but also a more efficient system accurate and accurate compared to available rainfall forecasts.

IV. PROPOSED METHODOLOGY

The forecast model is used to predict rainfall. The first step is to convert the data into the appropriate format to perform the tests and then do a good data analysis and see the variation in rain patterns. We predict rainfall by dividing the data into a training set and a test set and then using Linear regression and mathematical learning methods and comparing and drawing analyzes of the various methods used. With the help of many methods we try to reduce the error.

A. Data Description

Database [10] contains estimates of rainfall from 1916-2019 in each province.

1) The information contains 19 attributes (each month, year, and combination of 3 consecutive months) of 36 categories.
2) Details are only available from 1950 to 2015 in some sections
3) Attributes are the amount of rain measured in mm.

As the database is very large, it reduces the feature to improve authenticity, reduces calculation time and retention. Principal Component Analysis (PCA) the flexible output process required for a large set of variables. It releases a set of low-volume for the purpose of capturing a high amount of data. With a few exceptions, visual contact is very important. It is done using the covariance matrix and finding Eigen values in it. In our database through PCA we have reduced the qualifications by looking only at the rain data for the three consecutive months and annual data from all episodes. Technique Linear Regression. The full block diagram of the proposed model is shown in Fig. 4.1.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

B. System Architecture Diagram

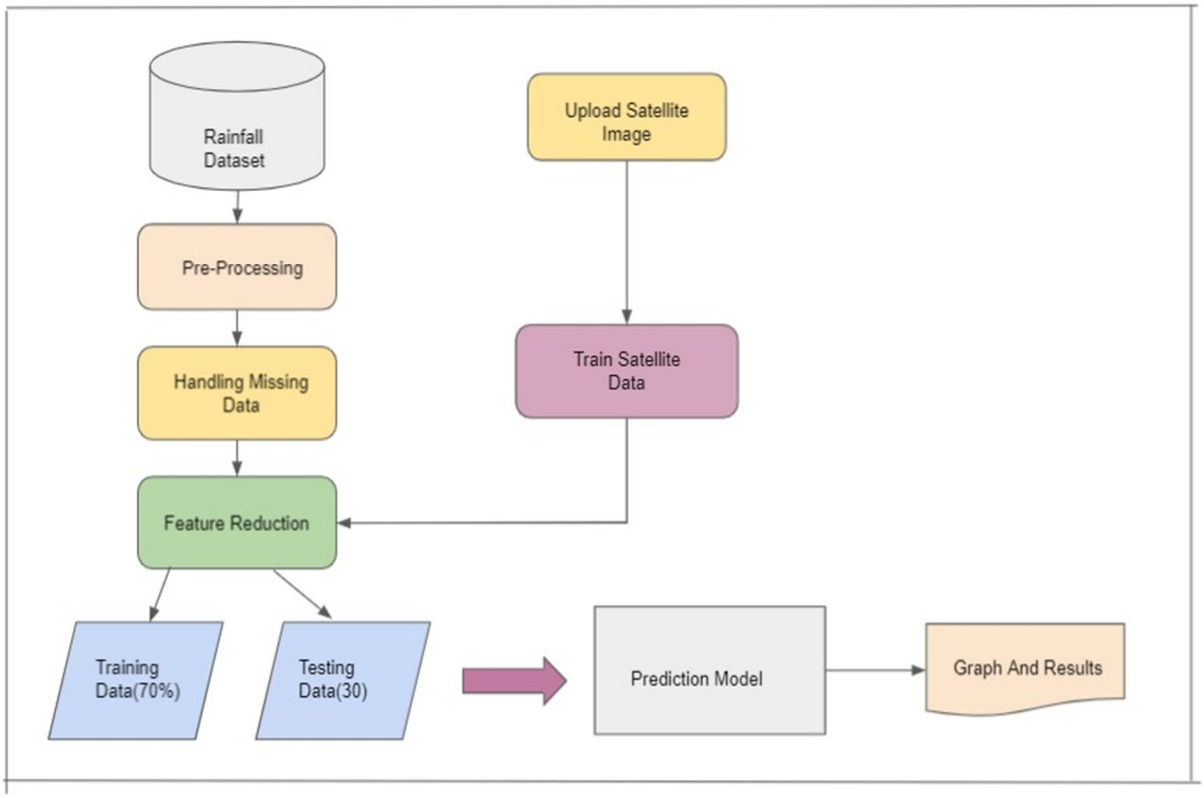


Fig 4.1 rainfall prediction model

C. Linear Regression Analysis

Linear Regression is a machine learning algorithm based on supervised learning. Do the back-up function. Regression models are targeted predictors based on independent variables. It is widely used to find relationships between variables and predictions. The different types of regression vary depending on the type of relationship between dependent and independent variables, taking into account

Line reversal makes it possible to predict a different amount of dependence (y) depending on the given independent variation (x). Therefore, this regression process finds a direct relationship between x (input) and y (output). Therefore, the name is Linear Regression.

IX (input) is a work experience and Y (output) is a personal salary. The return line is the most appropriate line in our model.

Linear Regression Hypothesis Function shown in eq
(4, 1):

$$y = \theta_1 + \theta_2 . x \qquad \textbf{(4,1)}$$

While training the model we are given:

x: input training data (univariate — single input variables (parameters))

y: data labels (supervised reading)

When training a model — it equals the best line to predict the value of y by a given number of x. The model gets the best line back by getting the best values best1 and θ2.

Θ1: avoid

e2: Coefficient of x

Once we get the best values θ1 and θ2, we get the best line. So when we finally use our model to predict, it will predict the value of y by the input value of x.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

D. How can you Update θ1 and θ2 values to find the Correct Line?

Cost of Work (J): By achieving a well-proportioned regression line, the model aims to predict the value of y as that the difference in error between the predicted value and the actual value is minimal. Therefore, it is very important to renew the values of θ1 and θ2, to achieve a positive value that reduces the error between the predicted value y (prep) and the true value y (y).

$$minimize \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2 \quad (4,2)$$

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2 \quad (4,3)$$

Linear Regression cost (J) shown in eq (4,3)s a Root Mean squared error (RMSE) between the predicted value y (prep) and the real value y (y).

E. Gradient Decline

Renew amanani1 and θ2 values to reduce operating costs (reducing the value of RMSE) and achieve the line that fits best with the model using Gradient decreases. The idea is to start with random and1 and -2 values and then update prices by price, reaching the lowest cost.

V. ALGORITHUM

1) Rain Forecast Input: Preset rain data
2) Output: Accuracy / forecast error
a) Step 1: Import the rain data that sets the CSV file.
b) Step 2: Fill in the missing values with the mean amount of data.
c) Step 3: Measuring features- measuring data at a fixed rate.
d) Step 4: Feature reduction- PCA is used to reduce data.
e) Steps: Data is divided into training set (70%) and testing set (30%).
f) Step6: Linear line and Mean Absolute Error line, r2 score calculated.
g) Step 7: Distribution sites are built between predicted data and testing of used models and errors are compared, and the best model is selected.
h) Step 8: Show the results in the form of graphs.

VI. MATHEMATICAL MODEL

A mathematical model is a description of a system that uses mathematical concepts and language. The process of building a mathematical model is called something mathematical modeling. Mathematical models are used not only in the fields of natural sciences and engineering but also in social sciences. Physicists, engineers, mathematicians, performance and economic research analysts use mathematical models extensively. The model can help explain the system and learn about the effects of various factors, and make predictions about behavior.

Mathematical genres can take many forms, including but not limited to flexibility, statistical systems, different scales or game theory models.

A. Set Theory

Set theory is a branch of mathematical concept that researches sets, which are collections of objects. While any type of object can be collected in a set, a set of theories are often used in mathematical objects. A set is a set of so-called members or elements of that set. If we have a set and say some things may (or may not be) belong to this set, they are (or are not present) in the set. We also say that those sets contain their own stuff.

B. System Description

Let's be a system with parameter sets Set S = {{I}, {R}, {P}, {O}}
Where, S again
I am set to all input that provides the system.
R a set of rules that drive your input set. P is set for all processes in the system.
O the expected product set from the system.

• Input (I): {I1, I2} Where,
I1 = User Information.
I2 = Rain Information.

Rules (R): {R1, R2}
Where,
R1 = Special information should only be displayed R2 = Continued Internet connection required.

Procedures (P): {P1, P2, P3} Where,
P1 = User Details must be updated.
P2 = Special Module should only view rain details
P3 = Predictability Information. should continue to be available.
Output (O): {O1, O2, O3}
Where,
O1 = City Weather.
O2 = Rainfall values.
O3 = Rain forecast.

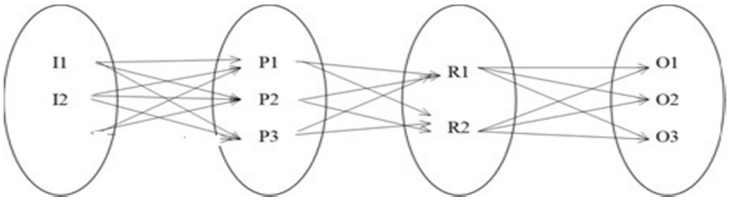


Fig 6.1 Venn diagram

VII. APPLICATION

A. Government
B. Public Domain
C. Security Domain
D. Weather Forecasting Department
E. Society

VIII. FUTURE SCOPE

In the future we can use advanced algorithms in the project domain, and we can try to improve the accuracy of the result.

IX. CONCLUSION

Rainfall is one of the most important natural phenomena that is important not only for humans but for living creatures. As a result of climate change, rainfall cycles also change, and global temperatures rise. Changing temperatures also affect agriculture, industry and can sometimes cause floods and landslides. Therefore, it is important for people to keep looking at this natural phenomenon in order to survive. Water is a scarce natural resource without which human life is impossible and nothing else can replace this natural resource. Therefore, it predicts agricultural rainfall and water storage areas, and it is also important to keep people alert to natural disasters such as floods and landslides. However, to overcome these problems and meet the needs, the rainfall forecasting system is essential using the artificial neural ingenuity popularized in modern technology

REFERENCES

[1] Urmay Shah, Sanjay Garg, Neha, Nitant, Shashikant, "Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques", 5th IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC-2018), 20-22 Dec, 2018, Solan, India, 978-1-5386-6026-3/18/$31©2018 IEEE.

[2] M. N. Ahuna, A. A. Alonge, "Rate Attenuation Prediction Using ANN for Dynamic Rain Fade Mitigation", South African Institute of Electrical Engineers, Vol. 110(1), pp. 11-18, Mar 2019.

[3] POOJA VERMA, DR. SWASTIKA CHAKRABORTY, "NOW CASTING OF OROGRAPHIC RAIN RATE USING ARIMA AND ANN MODEL", URSI RCRS 2020, IIT (BHU), VARANASI, INDIA, 12 - 14 FEBRUARY, 2020.

[4] R. Arya, Maya L. Pai, "Rainfall Prediction using an optimized Genetic-ANN MOdel", International Journal of Pure and Applied Mathematics, Volume 119 No. 10 2018, 669-678, ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version).

[5] Xiabo Zhang, Sachi Nandan Mohanty, "Annual and Non-Monsoon Rainfall Prediction Modelling Usign SVR-MLP: An Empirical Study From Odisha", IEEE Access, Received January 7, 2020, accepted January 21, 2020, date of publication February 10, 2020, date of current version February 18, 2020. Digital Object Identifier 10.1109/ACCESS.2020.2972435.

[6] S. Renuga Devi, C. Venkatesh, "Performance Comparison of ANN Models for Daily Rainfall Prediction", International Journal of Automation and Computing, DOI: 10.1007/s11633-016-0986-2.

[7] Senthamil Selvi, Seetha N, "Rainfall Prediction by using Time-series data in analysis of ANN Mmodels", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072.

[8] Mohini Darji, Vipul Dabhi, Harshadkumar, " Rainfall Forecasting using Neural Network: A Survey", Conference Paper · March 2015 DOI: 10.1109/ICACEA.2015.7164782

[9] Hassanuddin Noor, David, Yang, Noor Safar, "Rainfall-based River Flow Prediction Using NARX in Malaysia", 2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA 2017), 10 - 12 March 2017, Penang, Malaysia, 978-1-5090-1184-1/17/$31.00 ©2017 IEEE.

[10] H D Purnomo, K D Hartomo, SYJ Prasetyo, "Artificial Neural Network For Monthly Rainfall Rate Prediction", 1st Annual Applied Science and Engineering Conference IOP Publishing IOP Conf. Series: Materials Science and Engineering (2017) 012057 doi:10.1088/1757-899X/180/1/012057I.

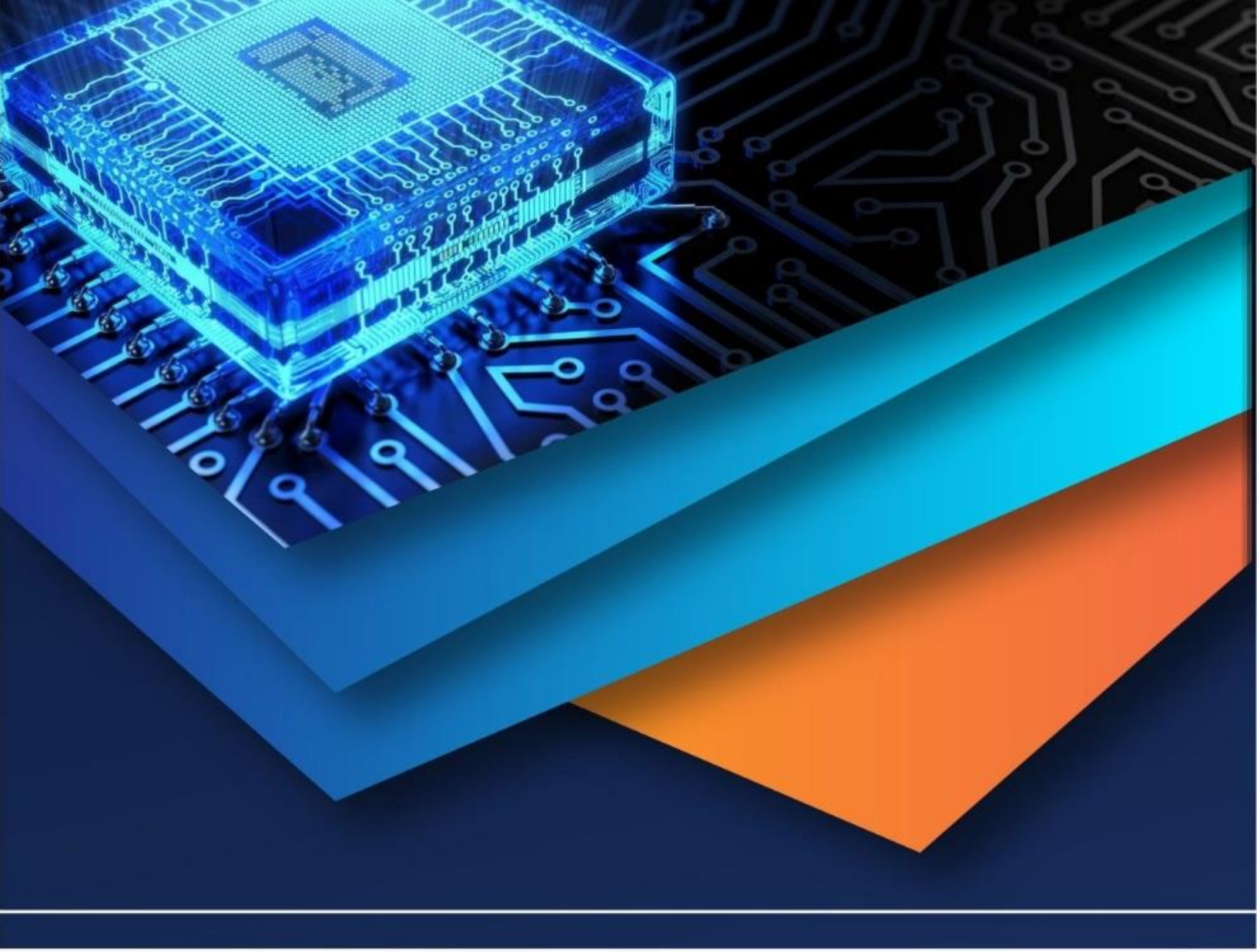
[11] https://data.gov.in/resources/subdivision-wise-rainfall-and-its-departure-1901-2015

[12] 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.

[13] Singh, Gurpreet, and Deepak Kumar. "Hybrid Prediction Models for Rainfall Forecasting." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2019.

[14] 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.

[15] 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.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY