STM JOURNALS

International Journal of
Satellite Remote Sensing

Volume 1, Issue 1, 2023
January–June
DOI (Journal): 10.37591/IJSRS

https://journals.stmjournals.com/ijsrs

Research IJSRS

Implementation of Anticipating Rainfall Using Machine Learning

Shaik Shafeeq Ahmed[1]*, Simran Pal R.[2], Shivaraj[1], Suneet Lionel Dsouza[1]

Abstract

Rainfall forecasting is crucial for many aspects of our national economy and should help prevent major seasonal droughts. Since agriculture is a beloved profession in many states, some Asian countries are economically hooked to decline. Previous precipitation info is beneficial. Farmers are cancerous in managing their crops, resulting in economic progress for the country. downfall prediction is hard for earth science scientists because of unordering time and unordered quantity of downfall. This enables the development of fresh herb designs and the efficient use of aquatic resources for the crops. Linear and non-linear style square measures are widely used to forecast seasonal decline. Several algorithms that are computer-aided rule-based techniques include the support vector machine (SVM), genetic algorithm (GA), and CART. Regression, Naive Bayes, Random Forest, and other analytical techniques are frequently used in this study. Overall, we tend to look at how the algorithmic rule that can be applied qualitatively is anticipating failure. The task appears challenging and sophisticated since it requires a vast array of specialist workers, and every person's choice is enthralled with no assurance of success.

Keywords: Rainfall detection, Anticipation, Prediction, Machine Learning, Genetic Algorithm.

INTRODUCTION

Rainfall prediction is beneficial for avoiding floods, which saves property. Moreover, it aids in the management of water resources. Previous precipitation info is beneficial. Farmers are cancerous in managing their crops, resulting in economic progress for the country. downfall prediction is hard for earth science scientists because of unordering time and unordered quantity of downfall [1–6].

Weather forecasting is the most important service provided by the earth science department for all nations on the planet [7]. The work looks difficult advanced because it necessitates associate outsized variety of specialized personnel, and everyone's selections square measure infatuated with no guarantee of success. Section two discusses the various ways for predicting downfall forecasting, nevertheless their limitations.

***Author for Correspondence**
Shaik Shafeeq Ahmed
E-mail: 1hk19cs138@hkbk.edu.in

[1]Student, Department of Computer Science and Engineering, H.K.B.K. College of Engineering, Nagawara, Bangalore, India
[2]Assistant Professor, Department of Computer Science and Engineering, H.K.B.K. College of Engineering, Nagawara, Bangalore, India

Received Date: July 27, 2023
Accepted Date: September 20, 2023
Published Date: October 17, 2023

Citation: Shaik Shafeeq Ahmed, Simran Pal R., Shivaraj, Suneet Lionel Dsouza. Implementation of Anticipating Rainfall Using Machine Learning. International Journal of Satellite Remote Sensing. 2023; 1(1): 1–8p.

In the current article, an effort has been made to extract the best data-driven machine learning techniques for predicting the state of India's average rainfall throughout the rainy season (June to September 1901–2018). This comparative analysis is based on three aspects: pre-processing techniques, input modeling and modeling methods. Comparisons among the four linear regression analysis, random forest method, Decision Tree and Naive Bayes method have been considered to find out the best prediction technique. In this study it has been observed that the maximum rain falls 385.3 mm occurred in year 1961 and minimum197.2 mm rain fall occurred in year 1974 [8–15].

It has been found that maximum accuracy for Random Forest regression 91% and mean absolute error was 2.3%, for Navie bayes accuracy was 89.64% and mean absolute error was 3.2%, for Linear regression accuracy was 87.28% and mean absolute error was 3.8%. From the present study it has been observed that Random forests are a viable model to be used in the field of rainfall prediction.

EXISTING SYSTEM

The rest of the paper is as follows: discussed briefly about the study area and the rainfall series used in this present study. It has been described the modelling approach which includes the input selection technique and the variable selection method, and how the weights are extracted. This is followed by discussions about the experimental setup and result and conclusion of the paper appear.

The rainfall data have taken from secondary source IMD (Indian Meteorological department rainfall updates from 1901 to 2018). In this paper; it has taken annual rainfall series of rainy season (June to September) data of State. Objectives are as follows:

1. Prediction of Rain fall for the study area.
2. Comparative analysis and to provide the best suitable model for prediction of rain fall [15–20].

Logistic Regression

A categorical dependent variable and one or more independent variables are analyzed and modelled using the statistical approach known as logistic regression. It is frequently employed in applications involving data analysis and machine learning, particularly in the discipline of predictive analytics. Logistic regression works by fitting a logistic function to the data, which models the probability of the dependent variable as a function of the independent variables. A value between 0 and 1, which the logistic function outputs, shows the likelihood that the event will occur [16].

Random Forest

The bagging method is extended by the random forest algorithm, which uses feature randomness in addition to bagging to produce an uncorrelated forest of decision trees. The random subspace method, also known as feature bagging, creates a random subset of features that guarantees low correlation between decision trees. The main distinction between deciding trees and random forests is this. Random forests merely choose a portion of those feature splits, whereas decision trees consider all possible feature splits.

Decision Tree

An easy technique for categorizing objects based on their features is a decision tree model. For instance, you might create a decision tree that determines whether your object is an apple based on its color, size, and weight. A decision tree functions by descending from the root node to the decision node. There are branches from the decision nodes that lead to either leaf nodes or further decision nodes. As its name implies, leaf nodes are terminal nodes that give a conclusion [17].

Naive Bayes

The definition of naive bayes in the context of machine learning is a probabilistic model in the supervised learning genre that is employed in a variety of use cases, mostly classification but also applicable to regression. The algorithm is designated as naive because it presumes that each characteristic that makes up the model is independent of the others, or, to put it another way, that changing one variable has no impact on any other variables. Even though this assumption is fundamental and powerful, the effect of the variable is still present. This approach makes Naive Bayes a straightforward but incredibly effective algorithm. This is the go-to method for any algorithm where one wants to either react to a request quickly or perform some calculations to produce some simple yet effective insights from the data due to its lower level of complexity [18].

PROPOSED SYSTEM

The brief methodology gives step by step procedure for prediction of rain fall. Because of the greater processing costs and necessary memory utilization associated with high dimensional data, classification techniques can be challenging to apply. To address this, dimensionality reduction techniques like as feature reduction and feature selection have been used. To enhance performance, estimated accuracy, visualization, and knowledge comprehension, feature extraction and selection procedures are either utilized separately or merged. The benefit of employing feature selection is that any small amount of information is not regarded as unimportant, but when considering a vast pool of information with a variety of features, some features may be left out. While performing feature extraction, the feature space's size is reduced with little information from the original feature space lost. The exact task's type and domain must be considered while deciding one of these two strategies to use.

Consisting of the following steps below mentioned:

1. Importing datasets.
2. Extracting independent and dependent variables.
3. Splitting the dataset into training and test set.
4. Feature scaling.

In training linear regression, decision tree, random forest and naive bayes algorithms are used for training each data in dataset. Using each algorithm we calculate the accuracy, precision, f1-score, and recall-score these values are calculated. Plotting is done for each feature in the dataset and the accuracy plot is plotted for all the algorithms.

In the dataset we have many fields in the dataset. To predict the rainfall rate in the country we use the fields to calculate the rainfall rate in the state. Dataset as the fields that are: year, month, day, temp avg, DP avg, humidity avg, SLP avg, visibility avg, Rainfall, wind avg [19–26].

RESULTS

In testing we taken the naive bayes to predict the rainfall rate of the given inputs. The Naive bayes algorithm has got a higher accuracy score compared to other algorithms so the Naïve bayes algorithm is used for prediction of rainfall rate and displayed in the Graphic User Interface form. Inputs can be given by the user and random inputs can be given predicted the rainfall rate. Figure 1. shows the block diagram of NB algorithm.

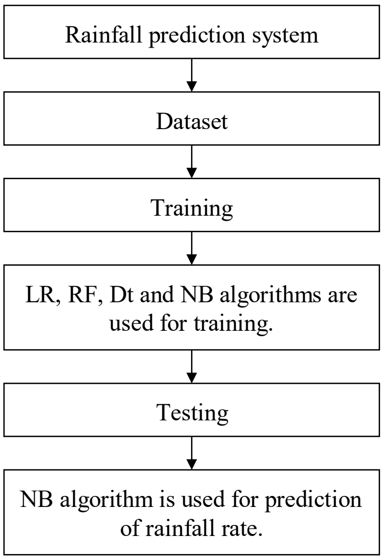

Figure 1. Block diagram of the NB algorithm.

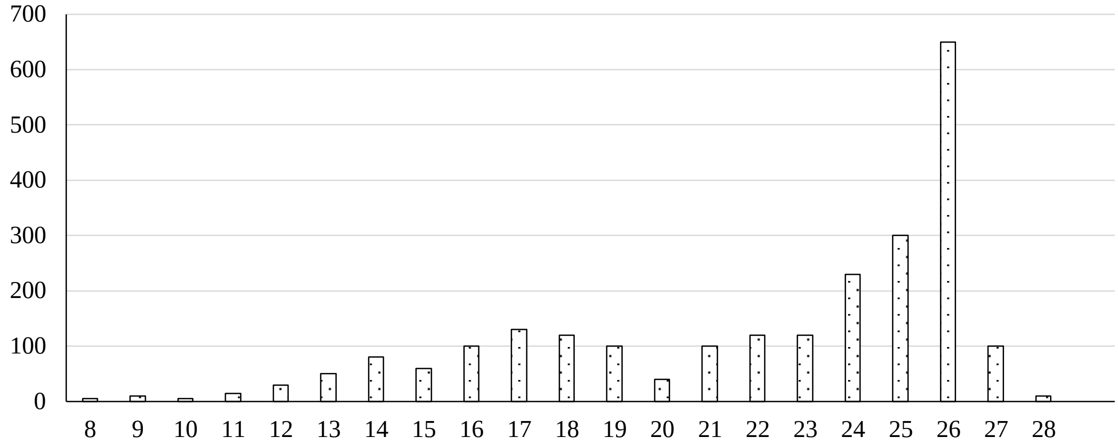

Figure 2. DPavg values in the dataset.

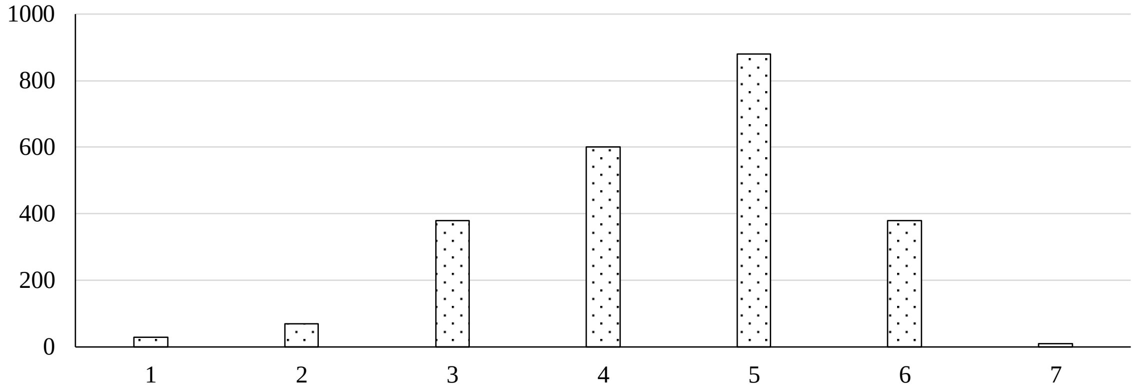

Figure 3. Visibility avg values in the dataset.

In Figure 2, relative humidity—the amount of moisture in the air in relation to the maximum amount that can be held at a given temperature—may be calculated using the DP temperature. Relative humidity is high when the DP temperature is near to the real air temperature, and low when the DP temperature is significantly lower than the air temperature.

As a result, the DP average is a measure of the average Dew Point temperature for a certain time, and it can provide information about the quantity of moisture in the atmosphere and how it may impact meteorological conditions such as precipitation, humidity, and fog.

The average distance at which an item may be plainly seen in the atmosphere is referred to as visibility average with respect to weather. Several factors affect visibility, including air pollution, precipitation, fog, and haze. Visibility is diminished when there is significant humidity or pollution in the air. Similarly, when there is fog or precipitation, vision can be drastically limited, making it difficult to see and maneuver. The visibility average is shown in Figure 3.

Humidity is a measurement of the quantity of water vapor in the air and is used to anticipate weather and understand atmospheric conditions. When the air is overly humid, it can feel sticky and unpleasant, while insufficient humidity can cause dry skin and respiratory difficulties. High humidity also influences how the human body cools itself through sweat evaporation. Figure 4 shows the Humidity values for every 20 cm of rainfall and its max is at the 9th, 10th and 11th months respectively. Weather stations and meteorological agencies use hygrometers to measure the quantity of moisture in the air. The readings obtained are used to compute the average humidity for a certain time.

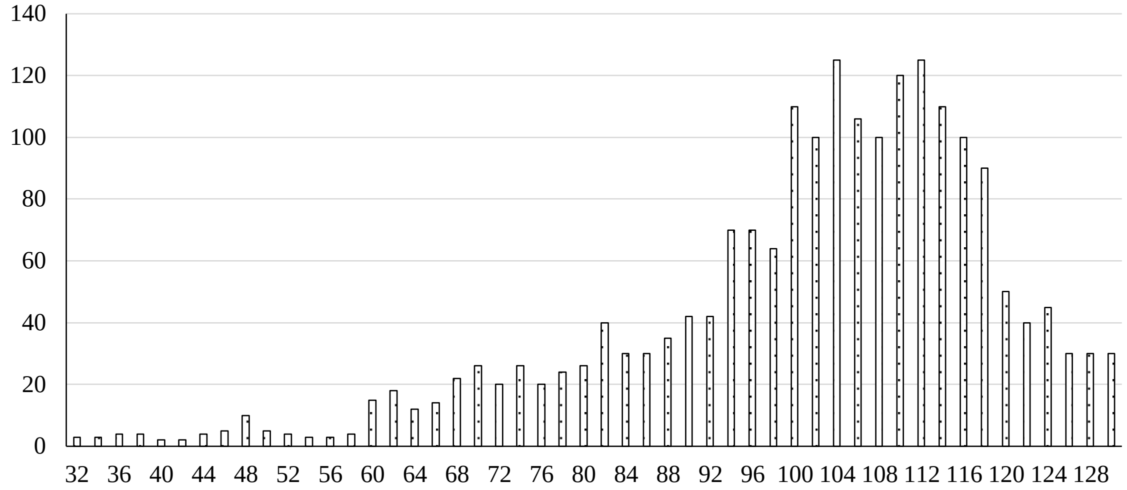

Figure 4. Humidity avg. values in the dataset.

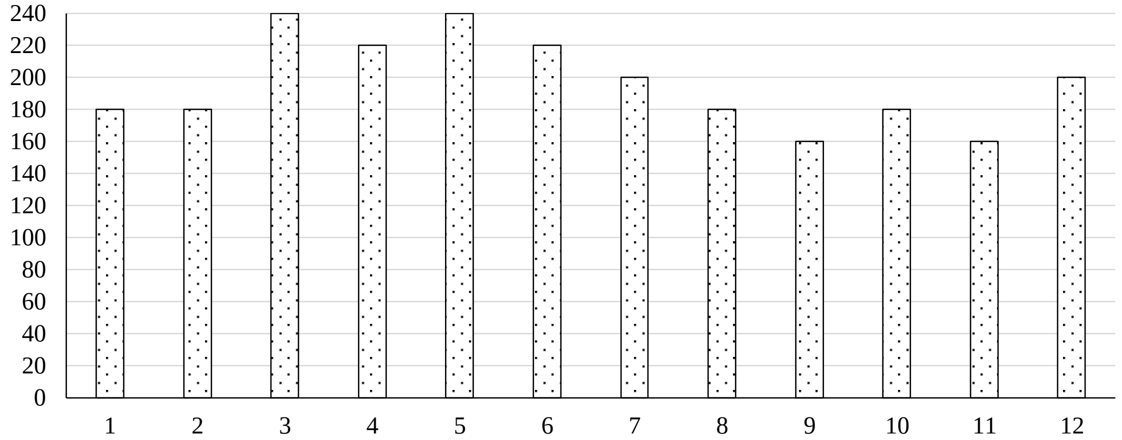

Figure 5. Displays the monthly rainfall totals in centimeters.

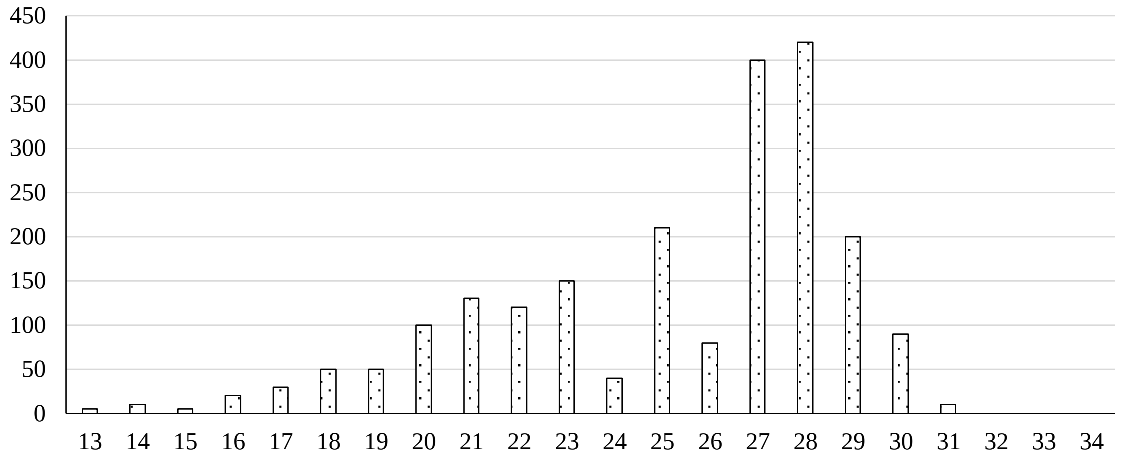

Figure 6. Temperature avg values in dataset.

Figure 5 shows the rainfall occurrence in cm for each month. Hence, we can deduce that more rainfall occurs in months 3, 4, 5 when compared to other months which is an absurd amount of rainfall during the summer season while month 12 has the highest rainfall during the winter season while other months are having moderate rainfall.

In terms of weather, average temperature refers to the average air temperature throughout a specific time, such as a day, a week, a month, or a year. Temperature is an important factor in forecasting the weather and comprehending atmospheric conditions, and it can have negative effects on the environment and human health. Figure 6 shows the temperature values for every 100 cm of rainfall and that the temperature is at its highest in the 10th month slowly reduces in the subsequent months.

Figure 7 represents the amount of rainfall count in cm every year. We can see that every year after 2011 has more than 350 cm of rainfall which is a staggering amount.

Wind average is the average speed and direction of wind for a given period, such as a day, week, month, or year. Wind is an essential aspect in weather forecasting and understanding atmospheric conditions, as well as in predicting rainfall. Wind may impact the movement of clouds and the distribution of precipitation, as well as the pace at which moisture evaporates from the soil and bodies of water. Wind may move moisture from one point to another, influencing the volume and distribution of rainfall in a specific area. Understanding the wind average may assist meteorologists in forecasting the movement of meteorological systems, such as storm fronts, as well as the likelihood and severity of rainfall in a specific location. Furthermore, wind patterns can help identify areas that are likely to experience drought conditions because persistent winds can transport moisture away from a given area.

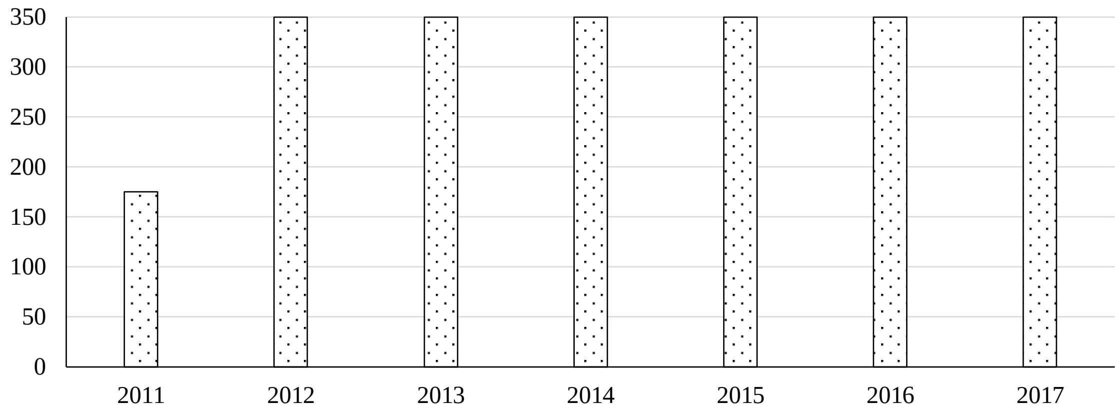

Figure 7. The year in the dataset.

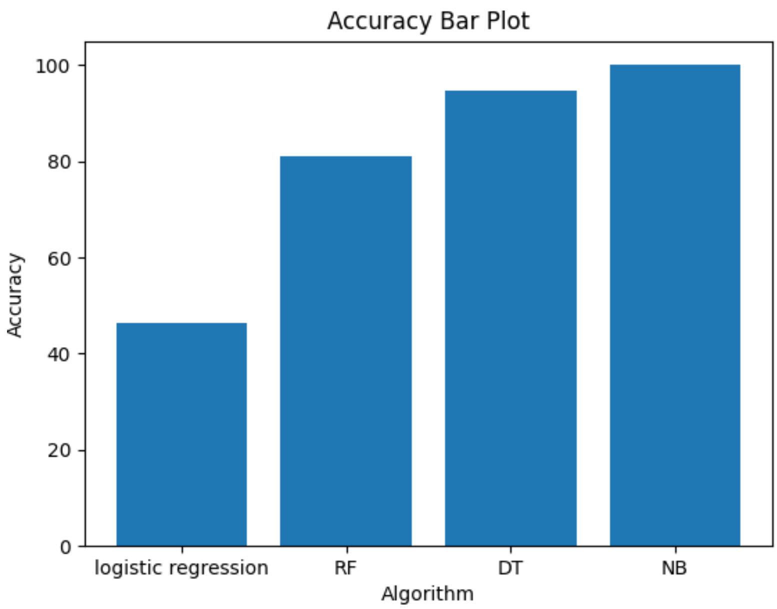

Figure 8. Accuracy v/s algorithm.

Data for the experiments has been divided into a training set and a testing section. Training data is considered as 75% and testing data 25% of all datasets. For conduction all the relevant experiment and model design prediction have been conducted in python IDE 3.6.2 and following libraries such as Mat plot for plotting, Pandas-for data analysis and manipulation. From skit learn machine learning algorithm such as logistic regression, Random Forest regression have been used. In the training phase, each of the individual models are trained with extensive parameter optimization.
Figure 8 indicates the comparison which shows accuracy vs algorithm.

CONCLUSION

Any algorithm would find it difficult to predict rain. This paper investigates the use of several machine learning methods and particularly suggests using the Random Forest method coupled with model design and rainfall forecasting in state. The rainfall series includes mainly rainy period (June to September) in state and comparatively analysis have been done. In this study it has been observed that the maximum rain fall 10.3mm occurred in the year 1961 and minimum 7.2mm rain fall occurred in year 1974. It has been found that maximum accuracy for Naïve Bayes of 98% and for Logistics regression accuracy was 46.28%, Decision Tree accuracy was 95%, Random Forest is 84.3% and from the present study it has been observed that Naive Bayes is a viable model to be used in the field of rain fall prediction. They are more accurate, can manage vast amounts of data, and can handle inconsistencies. One of the true spearheads in the field of weather forecasting is Nave Bayes and Decision Tree.

REFERENCES

1. S.M. Sumi, M.F. Zaman and H. Hirose a rainfall forecasting method using machine learning models and its application to Fukuoka city case". Article in International Journal of Applied Mathematics and Computer Science December 2012(,1-14), amcs.
2. 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, https://www.semanticscholar.org/paper/A-Framework-for-Medical-Image-Retrieval-Using-and-Rahman-Bhattacharya/cdcf84a6bb5581ae7fc25680be1083344871029e
3. M. Morales, L. Tapia, R. Pearce, S. Rodriguez & M. M. Amat o. (2004), "A machine learning approach for feature- sensitive motion planning. In Algorithmic Foundations of Robotics "VI, star-17, (pp. 361-376). Springer, Berlin, Heidelberg, https://link.springer.com/chapter/10.1007/10991541_25.
4. 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, Volume: 7, Issue-(3), 3372-3399.
5. 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, Volume-23. Issue-(4), 725-749.
6. N. Chantasut, C. Charoenjit, and C. Tanprasert, "Predictive mining of rainfall predictions using artificial neural networks for the Chao Phraya River," 4th Int Conf. of the Asian Federation of Inform. Technology in Agriculture and the 2nd World Congr. on Comput. in Agriculture and Natural Resources, pp. 117-122, Bangkok, Thailand, 2004.
7. 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. Cheriyadat, A. M. (2014). Unsupervised feature learning for aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing, Volume-52, issue-(1),439-451.
8. Bauer, S., Nolte, LP., Reyes, M. (2011). Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization. In: Fichtinger, G., Martel, A., Peters, T. (eds) Medical Image

Computing and Computer-Assisted Intervention – MICCAI 2011. MICCAI 2011. Lecture Notes in Computer Science, vol 6893. Springer, Berlin, Heidelberg.
9. S. Chattopadhyay and M. Chattopadhyay, "A Soft Computing Technique in Rainfall Forecasting," International Conference on Information Technology, HIT, pp. 19-21, 2007, https://arxiv.org/ftp/nlin/papers/0703/0703042.pdf
10. S. Chattopadhyay and G. Chattopadhyay, "Comparative study between different neural net learning algorithms applied to rainfall time series," Meteorological Applications, vol. 15, no. 2, pp. 273-280, 2008.
11. P. Guhathakurta, "Long-lead monsoon rainfall prediction using a deterministic Artificial Neural Network model," Meteorology and Atmospheric Physics 101, pp. 93-108, 2008.
12. C. L. Wu, K. W. Chau, and C. Fan, "Prediction of rainfall time series using Modular Artificial Neural Networks coupled with data preprocessing techniques," Journal of Hydrology, Volume-389, issue- 1, pp. 146-167, 2010, https://www.sciencedirect.com/science/article/abs/pii/S0022169410003252.
13. K. K. Htike and O. O. Khalifa, "Rainfall Forecasting Models Using Focused Time-Delay Neural Networks," Int. Conf. on IEEE, Comput. and Commun. Eng. (ICCCE), 2010, IEEE Xplore.
14. Subimal Ghosh and S. Kannan, "Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output," Springer-Verlag, July 2010.
15. M.Kannan, S.Prabhakaran, and P.Ramachandran, "Rainfall Forecasting Using Data Mining Technique," International Journal of Engineering and Technology, Vol.2, issue- (6), pp. 397-401, 2010.
16. G. Geetha and R. S. Selvaraj, "Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model," International Journal of Engineering Sciences and Technology, vol. 3, issue- 1, pp. 211-213, 2011.
17. M. A. Sharma and J. B. Singh, "Comparative Study of Rainfall Forecasting Models," New York Science Journal, vol. 115, no. 1, 2011, pp. 115-120.
18. J. Abbot and J. Marohasy, "Application of Artificial Neural Networks to Rainfall Forecasting in Queensland, Australia," Advances in Atmospheric Sciences, vol. 29, no. 4, 2012, pp. 717-730.
19. A. Kumar, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," Control and Syst. Graduate Research Colloq. (ICSGRC), 2012, pp. 82-87.
20. R. R. Deshpande, "On the Prediction of Rainfall Time Series Using Multilayer Perceptron Artificial Neural Network," International Journal of Emerging Technology and Advanced Engineering, vol. 2, no. 1, pp. 148-153, 2012.
21. "Designing a Rule-Based Hourly Rainfall Prediction Model," Soo-Yeon Ji, Sharad Sharma, Byunggu Yu, and Dong Hyun Jeong, IEEE IRI 2012, August – 2012, https://ieeexplore.ieee.org/document/6303024.
22. G. Shrivastava, S. Karmakar, and M. K. Kowar, "BPN model for long-term monsoon rainfall forecasting over a very small geographical region and its validation for 2012," Geofizika, vol. 30, issue- 2, pp. 155-172, 2013.
23. C. L. Wu and K. W. Chau, "Prediction of rainfall time series using modular soft computing methods," Engineering Applications of Artificial Intelligence, vol. 26, issue-3, pp. 997-1007, 2013.
24. "Prediction Of Rainfall Using Data Mining Technique Over Assam," Pinky Saikia Dutta and Hitesh Tahbilder, Indian Journal of Computer Science and Engineering (IJCSE), Vol. 5 issue-2 Apr May 2014.
25. A. Pankratz, Concepts and Cases for Forecasting with Univariate Box-Jenkins Models. 414 pages, John Wiley & Sons, Inc., New York.
26. M. T. Hagan, H. B. Demuth, and M. Beale, Neural Network Design, Thomson Asia Pte. Ltd, Singapore, 2002, https://hagan.okstate.edu/NNDesign.pdf

