

# Comparative Analysis for Prediction of Rainfall using Data Mining Techniques with Artificial Neural Network

R. Sukanya<sup>1\*</sup>, K. Prabha<sup>2</sup>

<sup>1\*</sup> Department of Computer Science, Periyar University PG Extension Centre, Dharmapuri, India  
<sup>2</sup> Department of Computer Science, Periyar University PG Extension Centre, Dharmapuri, Tamil Nadu, India

\*Corresponding author: [sukanyar.research@gmail.com](mailto:sukanyar.research@gmail.com) Tel.: +91-9488570051

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Received: 29/May/2017, Revised: 10/Jun/2017, Accepted: 24/Jun/2017, Published: 30/Jun/2017

**Abstract-** Rainfall Prediction is essential for countries which are based on agricultural economy like India. There are several factors are used to predict the rainfall such as temperature, pressure, wind speed, humidity, Mean sea-level etc. The accurate Rainfall Prediction is one of the most challenging problems in the atmospheric research. This paper discuss about Data mining technique which is suitable to predict the rainfall. This was carried out using several Classification algorithms such as Decision tree and Artificial Neural Network. ANN is a non-linear data modelling tool which is used to enhance the capability of Data Mining. It provides high accuracy, flexibility, good robustness, distributed storage and parallel processing. In this paper Back propagation Neural Network, Support Vector Machine is used for rainfall prediction. ANN improves the efficiency of Rainfall prediction by analysing the historical and current facts to make accurate predictions about future. For rainfall prediction, several Data mining techniques are used with ANN and comparison has been done by many researches are discussed.

**Keywords:** Rainfall Prediction, Data Mining, Classification algorithms, Artificial Neural Networks, Back Propagation.

## I. INTRODUCTION

Data Mining techniques used on weather data to find the hidden patterns inside the large data set so as to transfer the retrieved information into usable knowledge for classification and prediction of Rainfall. Data Mining is classified into two types. They are Descriptive Data mining and Predictive Data mining. Descriptive data mining describes concepts or task-relevant datasets in concise, summarized, informative and discriminative form. Predictive data mining is based on data and analysis, constructs models for the database and predicts the trends and properties of unknown data. Recently, there are several data mining techniques such as Classification, Clustering, Prediction, Association, Outlier Detection and Regression are used for Rainfall Prediction.

Classification is the main process of data mining which is used to classify the data according to different classes. Back Propagation is a Neural Network algorithm used for classification or numerical prediction. Basically, the Rainfall prediction has been done by two approaches. They are Empirical and Dynamical approach. Empirical approach is based on analysing the past historical data of weather and it is collected from Meteorological Department. Dynamic approach is based on mathematical equations and forward simulations of the atmosphere and it is developed as a computer model. It is suitable for large-scale weather data

and it is not suitable for short-term prediction. In this paper Section I contains the introduction of Data mining techniques and different approaches used for Rainfall prediction, Section II provides a literature survey on rainfall prediction using different data mining algorithms and neural networks, Section III explains the classification algorithm for rainfall prediction, Section IV gives the comparative analysis of various Classification algorithms and Neural network, Section V concludes that how to predict rainfall accurately using Decision Tree and neural networks.

## II. LITERATURE SURVEY

This section deals with related studies done by the researchers on different classification algorithms for Rainfall Prediction.

Ankita Sharma et al., 2015 [1] developed a Rainfall Prediction model using Back Propagation Neural Network and the result is compared with Cascaded Back Propagation, Layer Recurrent Network.

DingSheng Wan et al., 2012 [2] has developed the model for the annual average extreme rainfall prediction based on Back Propagation Neural Network combined with stepwise discriminate method and they have used Bayesian Classification to improve the accuracy of rainfall prediction.

Folorunsho Olaiya et al., 2012 [4] presented a data model for rainfall prediction using Decision Tree and Artificial Neural Network. The weather data was collected for 9 years at the city of Ibadan, Nigeria. The decision tree classification algorithm C5 was used to generate the decision trees and it is compared with the results of Artificial Neural Network.

Geetha.A. et al., 2014 [5] developed the Rainfall Prediction model using the Data mining technique such as Decision Tree. The training data was taken over the Trivandrum station for the year 2013 and the test data was taken for the year 2014. The original data consists of 20 parameters out of 12 parameters were taken for testing. Decision Tree is one of the effective methods for prediction and the accuracy rate will be increased in future when integrated with Artificial Neural Network, Fuzzy logic and Genetic Algorithms.

Jinghao Niu et al., 2015 [6] presented a rainfall prediction model and the data was taken from China Meteorological Administration (CMA) and the different data mining algorithms such as Naive Bayes, Support Vector Machine and Back Propagation Neural Network were used. In this model only seven attributes are taken as input and the comparison has been made between overall-data-rate (RO) and rainfall-data-rate (RR).

Jyothis Joseph et al., 2013 [8] developed the rainfall prediction model using Artificial Neural Network integrated with Data mining techniques such as classification and clustering. The weather data was obtained from the official meteorological website of National Oceanic and Atmospheric Administration and they could predict the rainfall accurately.

Kumar Abhishek et al., 2012 [9] develops an ANN model to predict the average monthly rainfall. The data was collected for Udipi, Karnataka for eight months and for fifty years. The data is normalized by finding mean and standard deviation of each attribute. This model used a three layered ANN with back propagation and shows the high accuracy in prediction.

Naik.A.R et al., 2013[10] presented a modified Back Propagation Artificial Neural Network model for Classification and Prediction of Indian Monsoon Rainfall and the weather data was collected from meteorological Department and out of the total data 80% were used for training purpose and 20% for validating purpose and the prediction accuracy achieved is in between 80-90%, hence neural network is suitable for predicting the rainfall.

Nirmala.M, 2015 [11] used traditional data processing technique Moving Average with Artificial Neural Network and the data was collected from Indian Institute of Tropical Meteorology, Pune, India for 136 years. From this 100 years weather data was taken as training data and 36 years as test

data. They have concluded that hybrid models improve the accuracy of prediction.

Nitin Mishra et al., 2016 [12] presented different methods used for rainfall prediction such as Support Vector Machine, Fuzzy logic, Back Propagation Neural Network and got significant results.

Sangari R.S et al., 2014 [13] presented data mining algorithms such as Naive-Bayes, K-Nearest Neighbour, Decision Tree, Neural Network and Fuzzy logic used for rainfall prediction and have proved that Neural Network gives better accuracy.

Siddharth S.Bhatkande et al., 2016 [14] presented weather prediction based on decision tree. The weather data for 3 years were collected for different cities and they have concluded that decision tree is suitable for multi variable analyses and for weather prediction.

Soo-Yeon Ji et al., 2012 [15] predicted the hourly rainfall in time efficiently using CART and C4.5 decision tree algorithms. Among these algorithms CART gives slightly better performance than C4.5.

### III.CLASSIFICATION ALGORITHMS FOR RAINFALL PREDICTION

Rainfall Prediction is done by making observations of the atmosphere such as Temperature, Pressure, Wind direction, Wind speed, Relative humidity, cloud cover, precipitation etc. Meteorological data are voluminous, dynamic, complex and high dimensional. Hence, decision tree algorithms are suitable for accurate prediction of rainfall.

Decision Tree is a tree-like structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test and tree leaves represent classes or class distributions. Decision Tree supports both quantitative and qualitative data. Large volume of meteorological data can be analyzed in a reasonable amount of time. Specific decision tree methods include Classification and Regression Trees (CART), Alternating Decision Tree, LogitBoost Alternating Decision Tree (LAD), C4.5 (Successor of ID3), ID3 (Iterative Dichotomiser 3), CHAID (CHI-Squared Automatic Interaction Detector) and MARS[5].

#### A) CART:

Classification and Regression Trees is used to characterize outcomes as a function of many predictors. It is very simple and more powerful algorithm. In this algorithm, the tree can be visualized more easily in high dimension. This is highly similar to regression.

#### B) C4.5:

C4.5 is a classification algorithm used to generate decision trees for the given data set. Decision tree is constructed by selecting the best possible parameter that will be able to split set of samples in most effective manner [3]. The attribute having the highest entropy difference or normalized information gain is selected as the splitting criteria for that particular node.

### C) Artificial Neural Network

Artificial Neural Network is one of the classification technique used in data mining for rainfall prediction. ANN is a set of connected input or output units in which each connection has a weight associated with it. The weights are adjusted during the learning phase to help the network to predict the accurate result. Neural networks are simplified model of biological neuron system. A neural network is a massively parallel distributed processor made up of simple processing units called neurons. There are three basic elements of neuron they are synapses connecting links obtained weight. Input signals weighted by synapses of neuron. An activation function for limiting the amplitude of output neuron. It is a non-linear predictive model which is suitable for rainfall prediction.

ANN can be classified into three types. Supervised Learning, Unsupervised Learning and Reinforced Learning. Back Propagation is a supervised Learning algorithm which is commonly used in Rainfall Prediction and it is Feed-Forward network used mainly for weather and financial prediction, face detection and character recognition [13].

### D) Back Propagation Neural Network

Back propagation neural network uses a feed forward network and it is a supervised learning algorithm. It is more powerful and expensive and it is widely used for Rainfall prediction. There are two phases in this algorithm. In the first phase, the input patterns are propagate to the network and the second phase is to adjust the weights in the network during output. This algorithm is mainly used to reduce the error until the ANN learns the training data. Back propagation neural network approach is suitable for rainfall prediction and it yields good results for all types of weather prediction [7].

### E) Support Vector Machine

Support Vector Machine is one of the supervised learning classification methods for both linear and non-linear data. SVM is most robust and accurate classification technique. SVM is typically used for classification, regression and ranking functions. SVM creates one or a set of hyper planes in a high or infinite dimensional space. The data points in the variable space are mapped into the hyper plane by selecting kernel functions. SVM gives better performance than other neural network models. It is a significant technique to solve many classification problems in the recent years. It is applicable for non-vectorial data such as graphs and maps. SVM provides unique and optimal solution [3]. The best separating hyper plane can be written as,

$$W.X + b = 0$$

Where  $w$  is a weight vector,  $x$  refers to the values of attributes and  $b$  is scalar often referred as bias.

## IV. COMPARISON OF DATA MINING TECHNIQUES WITH ARTIFICIAL NEURAL NETWORK

According to the previous research done by several researchers in the literature review, a comparative analysis can be done. Various Data mining techniques are used to predict rainfall using different Parameters, Algorithms, Attributes, Applications, Techniques, Accuracy [17]. The work done by different researchers and their comparison is noted down in Table-1. It concludes that Back propagation Neural Network and Decision Tree algorithms are best suitable for predicting the rainfall and it gives accurate results [18].

TABLE-1

COMPARISON OF DIFFERENT CLASSIFICATION ALGORITHMS FOR RAINFALL PREDICTION

Authors	Applications	Techniques	Algorithms	Attributes	Accuracy	Features	Pitfalls
Soo-Yeon, Sharad, Dong, Byunggu	Hourly Rainfall Prediction	Decision Tree	CART, C4.5	Wind speed, Wind direction, wind gust, outdoor humidity, Outdoor temperature, evaporation, solar radiation, wind chill, dew point, pressure altitude, cloud base, air density, vapour pressure	93%, 99%	It is easy to implement It Can be used with both discrete and continuous value	It does not work well on a small training data set. Small variation in data can lead to different decision trees.

JinghaoNiu,Wei Zhang	Rainfall Prediction	Classification	Naive Bayes, Support Vector Machine, Back Propagation Neural network	Pressure, Temperature, Evaporation, Humidity, Wind speed, Sunshine, Surface temperature	90%	It gives high prediction accuracy.	SVM is highly complex and it needs extensive memory requirements for classification
Gaurav,Sawale, Sunil R.Gupta	Weather Prediction	Artificial Neural Network	BPNN	Humidity, Temperature, Wind Speed	85%	It is easy to use with few parameters to adjust. It is easy to implement. It is mainly used with real life problems	BPNN requires high processing time if neural network is large. Learning can be slow.
Z.Jan	Climate Prediction	Lazy Learning	KNN	Temperature, Wind Speed, Dew point, Fog, Gust, Sea Level, Snow Depth	96%	Zero cost of the learning process. Long term accurate results with large set of attributes	KNN is sensitive to noisy or irrelevant attributes. Performance of algorithm depends on the number of dimensions.
Valmik B.Nikam, B.B.Meshram	Rainfall Prediction	Classification	Naive Bayes	Mean Sea Level Pressure, Relative Humidity, Station Level Pressure, temperature, vapour pressure, wind speed	80%	It is simple to implement. Classification efficiency is high. It predicts accurate results for most of the classification and prediction problem	It needs large number of records to obtain good results. The precision of algorithm decreases if the amount of data is less.

## V. CONCLUSION

Data mining techniques used for rainfall prediction is data intensive and in this paper analysis of various data mining algorithms suitable for Rainfall prediction is discussed. Since Rainfall is an accidental event and the cause of its occurrence is very typical. Even under the same climate condition no one can predict that it will rain or not. Hence the performance of the model is improved by using neural network. It is used to solve data mining problems because of its characteristics like high degree of fault tolerance, good robustness, self-organizing adaptive, parallel processing. Combining two or more prediction algorithms can improve the accuracy of rainfall prediction. There are several Decision Tree algorithms are available. Such as CART, C4.5, ID3 and Artificial Neural network algorithms such as Back propagation and Support Vector Machine. Nowadays most of the researchers use hybrid method for weather prediction. It gives the prediction result more accurate for any number of weather parameters.

## REFERENCES

- [1] Stuti Mishra, "Genetic study of advance breeding lines of soybean under Excess Monsoon Rainfall condition in Kymore Plateau zone of Madhya Pradesh", International Journal of Scientific Research in Biological Sciences, Vol.2, Issue.5, pp.4-9, 2015.
- [2] S.V.J.S.S. Rajesh, B.S.P. Rao, K. Niranjan, "Remote Sensing Study on Interlinking of rivers From Pennar to Cauvery", International Journal of Computer Sciences and Engineering, Vol.5, Issue.2, pp.100-111, 2017.
- [3] Fahad Sheikh, Karthick.S, Malathi. D, Sudarsan.J.S, Arun.C, "Analysis of Data mining Techniques for Weather Prediction", Indian Journal of Science and Technology, Vol.9, 2016.
- [4] Folorunsho Olaiya, Adesesan Barnabas Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies", International Journal of Information Engineering and Electronic Business, Vol.1, pp. 51-59, 2012.
- [5] A.Geetha, Dr.G.M.Nasira, "Data Mining for Meteorological Applications: Decision Trees for Modelling Rainfall Prediction", In the Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research.
- [6] Jinghao Niu, Wei Zhang, "Comparative Analysis of Statistical Models in Rainfall Prediction", In the Proceedings of the 2015 IEEE International Conference on Information and Automation, Lijiang, China, 2015.

- [7] Jyosthna Devi.Ch, Syam Prasad Reddy.B, Vaghan Kumar, Musala Reddy.B, Raja Nayak.N, "ANN Approach for Weather Prediction using Back Propagation", International Journal of Engineering Trends and Technology, Vol.3, Issue.1,2012.
- [8] Jyothis Joseph, Ratheesh T.K, "Rainfall Prediction using Data Mining Techniques", International Journal of Computer Applications, Vol-83, No.8, 2013.
- [9] Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar, "A Rainfall Prediction Model using Artificial Neural Network", IEEE Control and System Graduate Research Colloquium (ICSGRC 2012), pp 82-87.
- [10] A.R. Naik, Pathan S.K, "Indian Monsoon Rainfall Classification and Prediction Using Robust Back Propagation Artificial Neural Network", International Journal of Emerging Technology and Advanced Engineering, Vol.3, 2013.
- [11] M.Nirmala, "Integrated Soft Computing Approach for modelling Rainfall Prediction in Tamilnadu", IEEE Sponsored 9<sup>th</sup> International Conference on Intelligent Systems and Control, 2015.
- [12] Nitin Mishra, Dhawal Hirani, "A Survey on Rainfall Prediction Techniques", International Journal of Computer Application, Vol. 6-No.2, 2016.
- [13] Sanjesh Ghore, "Data Mining use of Neural Networks Approach", International Journal of Innovative Science, Engineering and Technology, Vol. 1 Issue.6, 2014.
- [14] R.V. Patil, S.S. Sannakki, V.S. Rajpurohit, "A Survey on Classification of Liver Diseases using Image Processing and Data Mining Techniques", International Journal of Computer Sciences and Engineering, Vol.5, Issue.3, pp.29-34, 2017.
- [15] Soo-Yeon Ji, Sharad Sharma, Byunggu Yu, Dong Hyun Jeong, "Designing a Rule-Based Hourly Rainfall Prediction Model", IEEE, 2012.
- [16] Valmik B Nikam, B.B. Meshram, "Modelling Rainfall Prediction Using Data Mining Method – A Bayesian Approach", Fifth International Conference on Computational Intelligence, Modelling and Simulation, Issue.2166-8531, pp.132-136, 2013.
- [17] K.V.S.R.P Varma, Enireddy Vamsidhar, Sankara Rao, Ravikanth Satapati, "Prediction of Rainfall Using Backpropagation Neural Network Model", International Journal of Computer Science and Engineering, Vol.2, No.4, pp.1119-1121, 2010.
- [18] S. Umar, A. Praveen, S. Gouse, N. Deepthi, "Imminent accession of Artificial Intelligence based Forensic Exploratory with Data Mining Analysis", International Journal of Computer Sciences and Engineering, Vol.5, Issue.3, pp.92-95, 2017.

## AUTHORS PROFILE

Mrs. R.Sukanya received M.Sc degree in Computer Science from Madurai Kamaraj University and M.Phil in Computer Science from Mother Teresa Womens University. Currently she is a Ph.D Research Scholar of Computer Science at Peiyar University P.G Extension Centre, Dharmapuri under the Guidance of Dr.K.Prabha, Assistant Professor of Computer Science. Her research area is Data Mining and Neural Network.



Dr.K.Prabha graduated in 2002 with a degree in Computer Science. She completed her post graduate M.Sc in Bharathiar University and M.Phil degree in Periyar University. Finally she awarded Doctoral degree in Computer Science from Bharathiar University. At present she is working as Assistant Professor of Computer science in Periyar University P.G. Extension Center, Dharmapuri. She has 9 years of teaching experience. Her interesting area of research is Network security and Data mining. Besides, she has published 16 referred journals and acted as a potential reviewer to her credit. She has delivered invited lectures in the various conferences and seminars.

