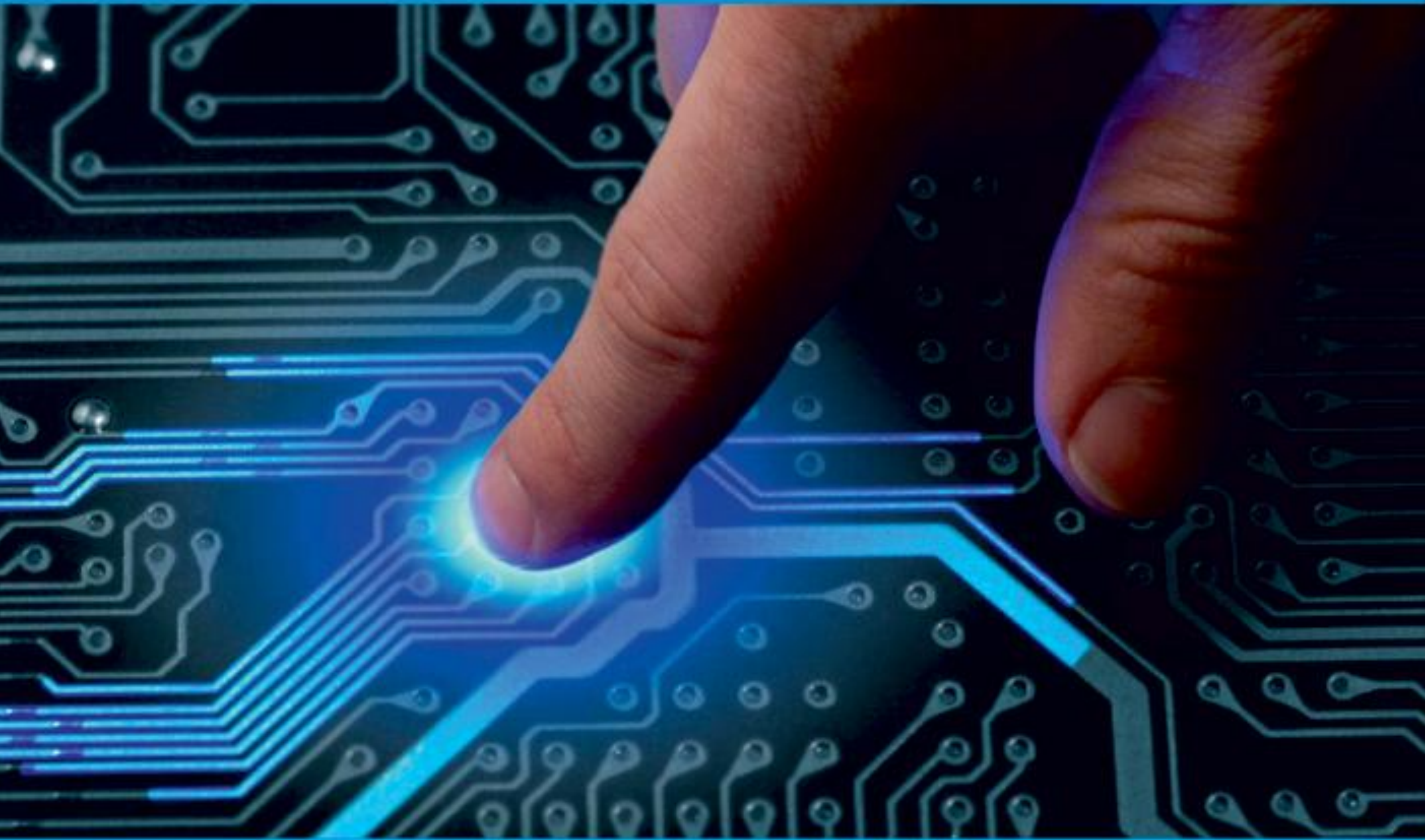
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

Rainfall Prediction using Machine Learning Techniques

Y.Bhagya Lakshmi, G.Deepika, D. Mahalakshami, G.Hrushikesh, G.Revanth

Assistant Professor, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

ABSTRACT: Heavy rainfall prediction is a major problem for meteorological department as it is closely associated with the economy and life of human. It is a cause for natural disasters like flood and drought which are encountered by people across the globe every year. Accuracy of rainfall forecasting has great importance for countries like India whose economy is largely dependent on agriculture. Due to dynamic nature of atmosphere, Statistical techniques fail to provide good accuracy for rainfall forecasting. Nonlinearity of rainfall data makes Artificial Neural Network a better technique. Review work and comparison of different approaches and algorithms used by researchers for rainfall prediction is shown in a tabular form. Intention of this paper is to give non-experts easy access to the techniques and approaches used in the field of rainfall prediction.
Keywords: Machine Learning, Statistical Techniques

I. INTRODUCTION

Rainfall prediction is helpful to avoid flood which save lives and properties of humans. Moreover, it helps in managing resources of water. Information of rainfall in prior helps farmers to manage their crops better which result in growth of country's economy. Fluctuation in rainfall timing and its quantity makes rainfall prediction a challenging task for meteorological scientists. In all the services provided by meteorological department, Weather forecasting stands out on top for all the countries across the globe. The task is very complex as it requires numbers of specialized and also all calls are made without any certainty. Section 2 discusses the different methods used for rainfall prediction for weather forecasting with their limitations. Various neural networks algorithm which are used for prediction are discussed with their steps in detail. Section 3 categorizes various approaches and algorithms used for rainfall prediction by various researchers in today's era. Finally, section 4 presents conclusion of paper.

II. BACKGROUND THEORY

Two widely used methods for rainfall forecasting are: Statistical methods and Numerical Weather Prediction (NWP) model [16]. Nature of rainfall data is non-linear. Frequency, intensity and amount are main characteristics for time series rainfall. These values can be varied from one position on earth to other position of earth and from one time to other time. Every statistical model has some drawbacks. Combination of AR and MA together forms a general and useful class of the time series model known as ARMA model. ARMA model is only useful for stationary time-series data and forecasting of short term rainfall. The statistical approaches do not have the ability to identify nonlinear patterns and irregular trend in the time series [16].

A. ARIMA Model

Box and Jenkins [29] proposed a methodology that consists of four steps:

Step 1. In the identification stage, IDENTIFY statement is used to specify the response of series and identification of candidate. The IDENTIFY statement considers time series that will be used in later statements, possibly for

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005072|

differentiating them, and calculates autocorrelations, inverse autocorrelations, cross correlations and partial autocorrelations.

Step 2 and 3. In the estimation and diagnostic checking stage, ESTIMATE statement helps ARIMA model to fit to the variable taken in the previous IDENTIFY statement, and to estimate the parameters of that model.

Step 4. In the forecasting stage, FORECAST statement is used to forecast future values of the time series. Confidence intervals are generated for these forecasts from the ARIMA model .

Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are important analytical tools used with the time series analysis and forecasting [30]. Main uses of these models are to measure the statistical relationships between observations in a single data series. ACF has big advantage of measuring the amount of linear dependence between results of a time series which will be separated by a lag k. In order to identify the model (step 1), ACF and PACF have to be estimated. They are used to guess the form of the model and to obtain approximate estimates of the parameters as well [30].

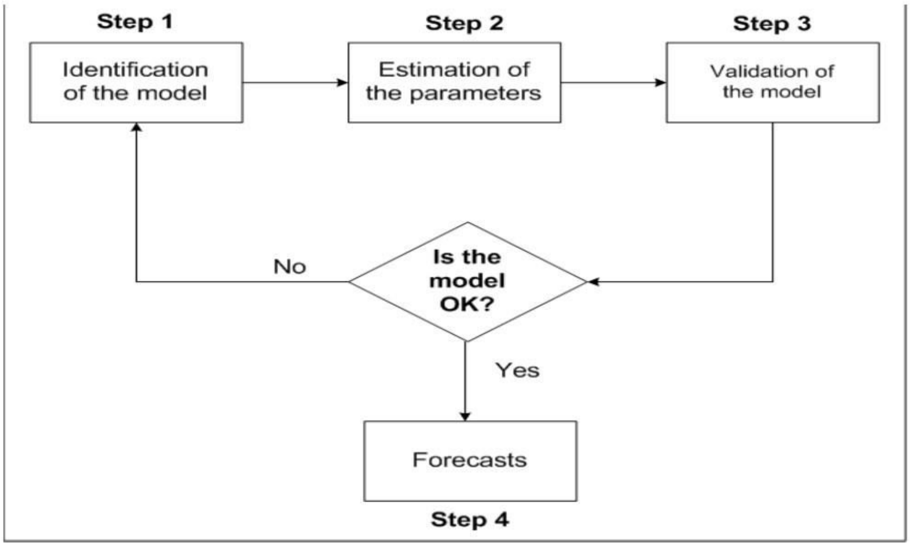


Fig. 1 Outline of Box-Jenkins Methodology

B. Artificial Neural Network

ANN is a computational model that is inspired by the human brain [31]. ANN contains a big number of interconnected neurons, which mostly operate in parallel, and are well structured. Categories of neural networks are either single layer or multi-layer. Layer between input layer and output layer is called as hidden layer. A single-layer feed forward (SLFF) neural network consists one input layer whose nodes have weights assigned and one output layer. A multilayer feed-forward (MLFF) neural network architecture can be developed by adding hidden layers in SLFF neural network.

1) Back-Propagation Neural Network:

BPNN is made of MLFF neural network which contains one input layer, hidden layers and one output layer. BPNN architecture with one hidden layer is shown in figure 2. The ultimate goal of BPNN is to decrease the calculated error obtained from the difference between the calculated output and desired output of the neural network by adjusting the weights after each iteration. So in BPNN, each information is propagated in backward direction until the calculated error is very small or zero. There are three phases of BPNN training: (a) using FFNN for training process of input. Adjustment of weights and nodes are made in this phase, (b) to calculate the error, and (c) modification of weights.

ANN model has great ability to learn by doing proper adjustment of these parameters for achieving the desired

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |
|| Volume 10, Issue 5, May 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1005072|

output. During the training process, this output may fit to the data very well, but it may provide poor results during the testing process. This suggests that the neural network may not generalize well. This might be because of overfitting or overtraining of data [32], which can be controlled by analyzing the error during training process and stoping the process when the error reaches a minimum threshold with respect to the testing set [33]. Alternate option to make the neural network generalize enough is by doing small changes in the number of layers and neurons in the inputs, without changing the output components. However, best neural network architecture selection is a heuristic approach. Solution is to keep the architecture of neural network relatively simple and small [34], because complex architectures are much more prone to overfitting [35].

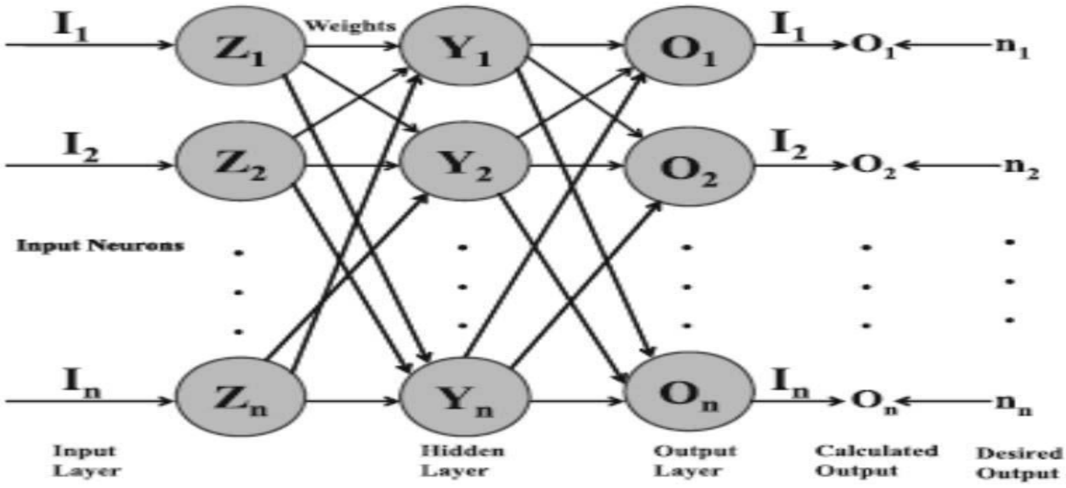


Fig. 2 A BPNN architecture with one hidden layer

2) Cascade Forward Back Propagation Network:

The CFBP network shown in Figure 2.3 is one of the artificial neural network types, which is used for the prediction of new output data. All the layers in networks are not only connected with its previous layer but also connected with input. Inputs are provided to each layer in network.

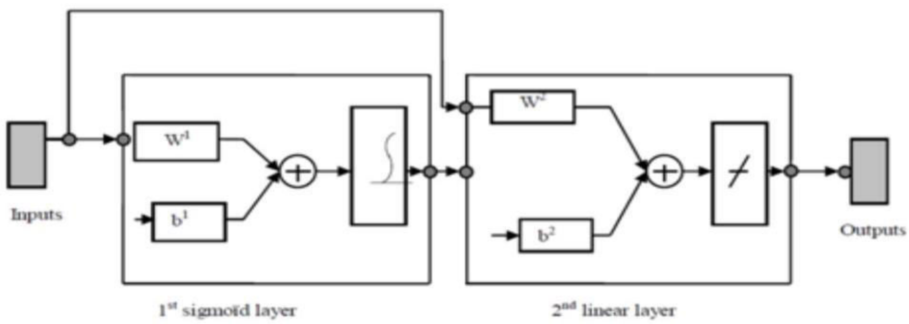


Fig. 3 Cascade Forward Back Propagation Network [40]

3) Layer Recurrent Network:

In this type of neural network, connections between units create a directed cycle. Unlike other feedforward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs [36]. RNN are neural networks with a feedback loop. The previous processes of hidden layer and functional outputs are fed back into the network as part of the input to the next hidden layer processes.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |
|| Volume 10, Issue 5, May 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1005072|

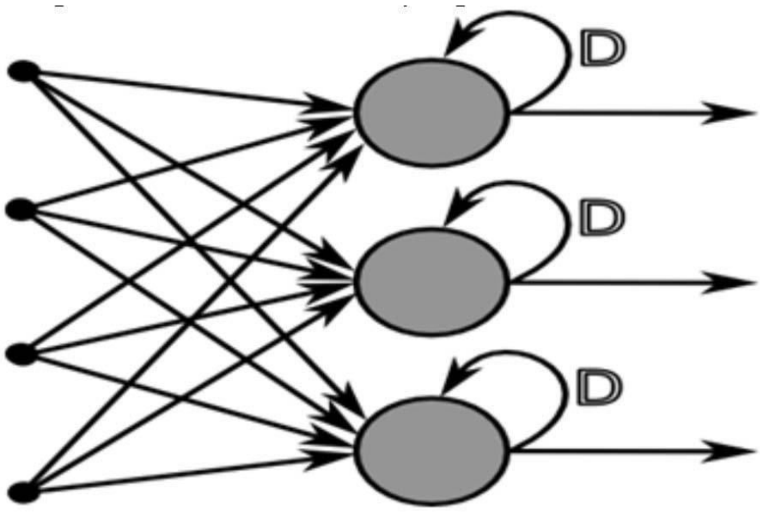


Fig. 4 Layer Recurrent neural Network

C. Support Vector Machine (SVM)

Support Vector Machine is also one type of feed forward network. Support vector machines are applicable for tasks like pattern classification, nonlinear regression etc. Support Vector Machines were created by Vapnik and his co-workers which has been used for supervised learning due to – (i) Better generalization ability than other NN models (ii) SVM solution is identical, optimal and absent from local minima (iii) Applicable to non-vectorial data (Strings and Graphs) and (iv) Very few parameters are needed for tuning the learning m/c. Very few scientists have applied this technique for rainfall prediction and results were acceptable.

D. Self Organizing Map (SOM)

SOM is also a part of artificial neural network. It is based on competitive learning. Neurons in the network compete between themselves to get activated. One neuron only can get activated at one time. Only winning neuron's weight is updated. This scheme is called as "winner takes all" scheme. Base of SOM is unsupervised learning, in which human interpretation is not required during the learning process. Structure of neurons is in one or two dimensional lattice. Self Organizing Map is data visualization technique invented by Prof. Teuvo Kohonen which reduces the dimensions of data through self-organizing neural networks [37]. SOM reduces dimensions & display similarities ([37], [38]).

III. LITERATURE REVIEW

P. Goswami and Srividya [1] have combined RNN and TDNN features and conclusion of their work was that composite models gives better accuracy than the single model. C.Venkatesan et al. [2] used Multilayer Feed Forward Neural Networks (MLFNN) for predicting Indian summer monsoon rainfall. Error Back Propagation (EBP) algorithm is trained and applied to predict the rainfall. Three network models with two, three and ten input parameters have analyzed. They also compared the output result with the statistical models.

A.Sahai et al. [3] used error back propagation algorithm for Summer Monsoon Rainfall prediction of India on monthly and seasonal time series. They used data of previous five years of monthly and seasonal mean rainfall values for rainfall prediction. N.Philip and K.Josheph [5] used ABF neural network for yearly rainfall forecasting Kerala region. Their work suggests that ABFNN performs better than the Fourier analysis.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005072|

V. Somvanshi et al. [7] predictied rainfall of Hyderabad, INDIA region using ANN model. They also compared ANN with ARIMA technique. They used past four months rainfall data as inputs to neural network model. S. Chattopadhyay and M. Chattopadhyay [9] have used two parameters minimum temperature and maximum temperature for rainfall forecasting.

S. Chattopadhyaya and G. Chattopadhyaya [10] used Conjugate Gradient Decent (CGD) and Levenberg–Marquardt (LM) learning algorithm for training. Performances of both algorithms were same in prediction task. C. Wu et al. [12] predicted the rainfall of India and China. They applied Modular Artificial Neural Network (MANN). MANN's performance was compared with LR, K-NN and ANN.

K. Htike and O. Khalifa [13] used yearly, biannually, quarterly and monthly rainfall data for rainfall prediction. They trained four different Focused Time Delay Neural Networks (FTDNN) for rainfall forecasting. Highest prediction accuracy was provided by the FTDNN model when yearly rainfall data is taken for training. S. Kannan and S. Ghosh [14] contributed towards developing K- mean clustering technique combined with decision tree algorithm, CART, is used for rainfall states generation from large scale atmospheric variables in a river basin. Rainfall state on daily basis is derived from the historical daily multi-site rainfall data using K- mean clustering. M. Kannan et al. [15] predicted short term rainfall. Empirical method technique is used for prediction task. Data of three specific months for five years is analyzed for particular region. Clustering is used for grouping the elements.

G. Geetha and R. Selvaraj [16] used ANN model for predicting monthly rainfall of Chennai region. M. Sharma and J. Singh [17] considered parameters such as rainfall, maximum and minimum temperature, and relative humidity. They predicted weekly rainfall over Pantnagar region. ANN obtained higher prediction accuracy than multiple linear regression model. J. Abbot and J. Marohasy [18] used Time Delay Recurrent Neural Network (TDRNN) for monthly rainfall prediction over Australia region. A. Kumar et al. [19] predicted average rainfall over Udipi district of Karnataka. They used ANN models for prediction task of rainfall. They concluded that Back Propagation Algorithm (BPA) was better than the layer recurrent and cascaded back propagation. Soo-Yeon Ji et al. [21] predicted the hourly rainfall. CART and C4.5 are used for prediction, which may provide hidden important patterns with their reasons. There were 18 variable used from weather station. 10 fold cross validation method is performed for validation purpose. CART performed better than C4.5.

S. Nanda et al. [24] predicted rainfall using a complex statistical model ARIMA and three ANNs models which are MLP, LPE (Legendre Polynomial Equation) and FLANN (Functional- Link Artificial Neural Network). In Comparision, FLANN gives better prediction accuracy compared to the ARIMA model. A. Naik and S. Pathan [25] used the ANN model for rainfall prediction. They modified back propagation algorithm which was more robust than the simple back propagation algorithm. Pinky Saikia Dutta and Hitesh Tahbilder [28] predicted monthly Rainfall of Assam by traditional statistical technique -Multiple Linear Regression. Parameters selected for the model are min-max temperature, mean sea level pressure, wind speed and rainfall. Acceptable accuracy is given by prediction model based on multiple linear regression.

Table presents categorization of different approaches of rainfall prediction. The categorization is based on following features: authors, region, dataset time period, techniques, accuracy measure and rainfall predicting variables.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005072|

TABLE I. CATEGORIZATION OF DIFFERENT APPROACHES OF RAINFALL PREDICTION

Authors	Region	Dataset Time Period	Techniques	Accuracy Measure	Rainfall Predicting Attribute
P. Goswami, Srividya (1996) [1]	Global (all over India)	Yearly (135 years)	Artificial Neural Network (EBP)	Relative percentage error	Mean rainfall
C. Venkatesan et al.	Global (all over	Monthly (1939-1994)	Artificial Neural	RMSE	Min-Max temperature
(1997) [2]	India)		Network (EBP)		
A. Sahai et al. (2000)	Global (all over	Monthly (1876-1994)	Artificial Neural	RMSE,	Min-Max temperature
[3]	India)		Network (EBP)	correlation coefficient	
N. Philip et al. (2001)	Local (Kerala)	Monthly (1893-1933)	ABFNN	RMSE	Wind, temperature,
[4]					latitude-longitude
					pr es
N. Philip, K. Joseph (2002) [5]	Local (Kerala)	Yearly (1893-1933)	ABFNN	RMSE	Wind, temperature, precipitation,
N. Chantasut et al.	Local (Chao Pharya River)	Monthly (1941-	Artificial Neural	MSE	Temperature
(2004) [6]			Network (EBP)		
V. Somvanshi et al.	Local	Yearly (103 years)	ANN, ARIMA	RMSE, MAE	Humidity,
(2006) [7]					
S. Chattopadhyay	Global (all over	Monthly	Artificial Neural	MSE	Temperature,
(2007) [8]	India)		Network (EBP)		
S. Chattopadhyay,	Global (all over	Monthly	Multilayer	MSE	Min-Max
Chattopadhyay	India)				
[9]					
S. Chattopadhyay,	Global (all over	Monthly	Artificial Neural	MSE	Min-Max
Chattopadhyay	India)		Network (EBP)		
[10]					
P. Guhathakurta	Global (all over	Yearly (1941-2005)	Artificial Neural	RMSE	Min-Max
[11]	India)		Network (EBP)		
C. Wu et al. (2010)	Global (India,	Daily, Monthly	Modular	RMSE	Min-Max
[12]	China)		Neural Network		
K. Htike, O. Khalifa	Global (India)	Yearly, Monthly,	Focused Time	MAPE	Temperature,
(2010) [13]		Biannually,	Neural Network		evaporation
S. Kannan, S. Ghosh	Local (River)	Daily (50 years)	Decision tree,	MSE	Temperature,
(2010) [14]			K-mean		rai
M. Kannan at el.	Global (India)	Quarterly (5 years)	Regression	MSE	Min-Max

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005072|

(2010) [15]						direction,
G. Geeta, R.		Local (Chennai)	Monthly (1978-	Multilayer Back	RMSE	Wind speed,
(2011) [16]				Propagation		relative
				Network		
M. Sharma, J. Singh		Local (Pantnagar,	Weakly (39 Years)	Multiple	Absolute mean	Min-Max
(2011) [17]		India)		Model,	difference	humidity, pan
J. Abbot, J.		Local (Australia)	Monthly (1900-	TDRNN	RMSE, Pearson	Rainfall,
(2012) [18]					correlation coefficient	atmospheric temperature, solar data
A. Kumar et al.		Local (Udipi)	Monthly (1960-	EBPNN,	MSE	Average
[19]						sp
R. Deshpande		Local	Monthly	Elman neural	MSE	Ra
[20]						
S. Yeon et al.	(2012)	Local	Hourly (3 years)	Decision tree,	High prediction	Temperature,
[21]					accuracy	humidity,
G. Shrivastava et al.		Local	Yearly (1951-2011)	Artificial Neural	MSE	Humidity,
(2013) [22]				Network (EBP)		
C. Wu, K. Chau		Global (India,	Daily, Monthly	Moving	RMSE	Min-Max.
[23]		China)				
S. Nanda et al.		Global (India)	Yearly (1990-2012)	ARIMA model,	MSE	Min-Max.
[24]						
A. Naik et al.	(2013)	Global (India)	Monthly	Artificial Neural	RMSE	Wind speed,
[25]				Network (EBP)		

IV. CONCLUSION

The estimation of rainfall is of great importance in terms of water resources management, human life and their environment. It can be met with the incorrect or incomplete estimation problems because rainfall estimation is affected from the geographical and regional changes and properties. This paper presented review of different methods used for rainfall prediction and problems one might encounter while applying different approaches for rainfall forecasting. Due to nonlinear relationships in rainfall data and ability of learning from the past makes Artificial Neural Network a preferable approach from all available approaches.

REFERENCES

1. P. Goswami and Srividya, "A novel Neural Network design for long range prediction of rainfall pattern," Current Sci.(Bangalore), vol. 70, no. 6, pp. 447-457, 1996.
2. C. Venkatesanet, S. D. Raskar , S. S. Tambe , B. D. Kulkarni , and R. N. Keshavamurty , "Prediction of all India summer monsoon rainfall using Error-Back- Propagation Neural Networks," Meteorology and Atmospheric Physics, pp. 225-240, 1997.
3. K. Sahai, M. K. Soman, and V. Satyan, "All India summer monsoon rainfall prediction using an Artificial Neural Network," Climate dynamics, vol. 16, no. 4, pp. 291-302, 2000.
4. N. S. Philip and K. B. Joseph, "On the predictability of rainfall in Kerala-An application of ABF neural network," Computational Science- ICCS, Springer Berlin Heidelberg, pp. 1-12, 2001.
5. N. Chantasut, C. Charoenjit, and C. Tanprasert, "Predictive mining of rainfall predictions using artificial neural networks for 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, Bangkok, Thailand, pp. 117-122, 2004.
6. V. K. Somvanshi, O. P. Pandey, P. K. Agrawal, N.V.Kalanker1, M.Ravi Prakash, and Ramesh Chand, "Modeling and prediction of rainfall using Artificial Neural Network and ARIMA techniques," J. Ind. Geophys. Union, vol. 10, no. 2, pp. 141-151, 2006.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING