RAINFALL FORECASTING MODEL USING ARTIFICIAL NEURAL NETWORK

Dr.P.Subhashini[1], John De Britto C[2], Dr.M.C.Babu[3], Muthuvelavan Pillai[4] N,Akhil Nair[5], A.Vidhya[6]

Email id: subhasaash@gmail.com[1], yjohnde@gmail.com[2], babumc@citchennai.net[3],muthuvikni@gmail.com[4],akhiln240@gmail.com[5],vidhyaraj1988@gmail.com[6]

Professor[1], Department of IT, Veltech Multitech Dr.Rangarajan Dr.Sagunthala Engineering College,Chennai-600 062 Tamilnadu, India.

Assistant Professor[2], Department of EEE, Saveetha Engineering College, Chennai-602 105 Tamilnadu, India.

Associate Professor[3], Department of Artificial Intelligence and Data Science, Chennai Institute of Technology, Chennai-600 069 Tamilnadu, India.

Professor[4], Department of CSE, J.N.N Institute of Engineering, Chennai-601102 Tamilnadu, India.

Assistant Professor[5], Department of CSE, J.N.N Institute of Engineering, Chennai-601102 Tamilnadu, India.

Assistant Professor[6], Department of CSE, Jeppiaar Engineering College, Chennai-600 119 Tamilnadu, India.

Abstract:

In the meteorological field the challenging task that is ever facing is forecasting rainfall and its level. The long term study and data helps to forecast the rainfall before arriving. If the prediction went wrong this may lead to disasters like floods. Inorder to circumvent the impact of disaster due to heavy rainfall the dynamic state prediction is done with the help of Artificial Neural Network (ANN) as a background learning model framework in an effective way. For the prediction, a recorded annual rainfall data set is used from the kaggle repository which consists of a huge volume of data about the subdivisions of India with several environmental factors such as temperature, pressure, humidity, wind direction, Evaporation and sunshine. The performance of ANN model is validated in two perceptions by deploying it with a change of two characteristics in the traditional architecture. One is using the back propagation algorithm and another one is using two hidden layers. The importance of accuracy is projected in the results along with the loss factor and attained the expected results since the Indian economy is mostly based on the agriculture sector.

Index Terms: Rainfall, Prediction, Artificial Neural Network, Accuracy, Agriculture

Number: 10.14704/nq.2022.20.7.NQ33363 **Neuro Quantology 2022; 20(7): 2842-2850**

I.Introduction

Water resource managers and hydrologists are growing more concerned about the changing rainfall pattern brought on by climate change. Heavy and erratic rainfall can have a variety of negative effects, including crop destruction and property damage. In order to improve agricultural farm management and provide early warning that can lessen risks to life and property, a better forecasting model is essential. There are a number of hardware tools for predicting rainfall that are based on weather factors including temperature, humidity, and pressure. The task is much more difficult because every decision made in meteorology must be made in the face of uncertainty. This difficult approach requires a wide variety of specialized skills. Due to the inefficiency of conventional methods, we can produce accurate results by employing machine learning algorithms.

This process will be simple thanks to the analysis of historical rainfall data and rainfall forecasting for upcoming seasons. Many random generators with outputs that closely resemble the weather data to which they have been fitted and used by scientists from all around the world to create stochastic weather models.

This is because the ANN (Artificial Neural Network) model is built on producing "predictions" by intelligently "analyzing" a huge historical set of data. The models are either mathematical or statistical in nature, aside from ANN. These models have been shown to be quite accurate in computations but less accurate in predictions because they cannot adapt to irregularly varying data patterns that cannot be described as functions or deduced from formulas. It has been demonstrated that "artificial neurons" that have the ability to learn from experience, such as through back-propagating errors in subsequent estimates, are better able to comprehend some real-life events. Precision may diminish, but our capacity to "understand the situation," "replicate it," or "draw inferences from it" will be increased.

The most significant meteorological event in recorded human history is rain. Its frequency and amount on various scales are crucial to human civilization. Numerous stochastic models have been developed for a specific geographic area to forecast rainfall, look at seasonal fluctuations, and calculate yearly and monthly rainfall. Creating an ANN model is the aim of most of the prediction applications. The monsoon season helps the Indian economy. The typical area's rainfall condition is forecasted using back propagation of ANN, and the fragrance of newness is further emphasized by the fact that different models are tested to see which one offers the best match.

II.Literature Survey

Initial forecasts are based on human study of the location and already existing data sources. The Pettitt test, the Mann-Kendal test, and the Standard Normal Homogeneity test, among many other tools and statistical test models, were also used (SNHT). For fault recovery and precise prediction, the investigation then focused on supervised and unsupervised learning models. Many have constructed an ANN model for weather forecasting, which is considered as the first application of ANN in this field[1]. The last several decades have seen an increase in the usage of artificial neural networks (ANNs) for environmental event prediction. [2]This study contrasts multilinear regression's efficiency with that of artificial neural networks in order to estimate missing rainfall data. Using the forward and back extension principles, researchers developed a model that can generate precipitation in regions where the time series has stopped or where the archive has just begun.

Data that is in question can be confirmed by comparing it to records from nearby stations. It is novel to contrast the application of neural networks with the traditional multiple linear regression strategy. In a study, the Evolving Fuzzy Neural Network (EfuNN) and the ANN with Scaled Conjugate Gradient Algorithm (ANN-SCGA) are used to forecast rainfall time series. The model was trained using monthly precipitation data as the input. It has been shown that neurofuzzy networks outperform neural networks due to their more precise functioning. [4] An ANN-based technique for predicting monsoon rainfall was created in this work. With the use of backpropagation learning, the model was created. Its execution is made by the use of input data on summer monsoon rainfall and output data on summer rainfall average from the same year. After training and testing, it was shown that a network with more hidden layer nodes was more successful in forecasting rainfall.

The model with eleven hidden nodes and three layers ultimately defeated the asymptotic regression technique. [5] This work developed an ANN model to forecast rainfall in Bangkok, Thailand, with a lead time of 1-6 hours. An ANN model was created using hourly data accumulated over four years from 75 local rain gauge sites. Finally, they used the model to predict rain and prevent floods before they happened.

The three techniques investigated were the Cascaded Back Propagation (CBP), Layer Recurrent Network, and Back Propagation Algorithm (BPA). BPA provided the best accuracy and MSE values out of the three networks that were examined, according to the researchers. They concluded that BPA is the most effective algorithm among the three networks.

The forecast model was trained and validated using meteorological data collected in 2012 at Gangneung, in the Gangwon-do area of Korea. Temperature, wind speed, humidity, and sea surface pressure were among the climatic factors included in the weather dataset. In addition, the difference in the amount of rain that fell in the most recent and previous hours was considered. The LSTM-Networks model was tested in comparison to an Artificial Neural Network (ANN) model. The study's results showed that for the RMSE assessment measure, the LSTM-Networks model produces superior results. One of the components of the second stage of the experiment was the measuring of water vapor. However, the values of the RMSE over the training epochs showed an over-fitting behavior for the training batch of data.

[9]proposed a method that uses a ConvNet model and an LSTM-Networks model to predict monthly rainfall. [10] evaluated two LSTM- and RNN-based models for predicting monthly precipitation. The efficiency of the algorithms ARIMA, Extreme Learning Machine (ELM), and K Nearest Neighbor (KNN) has been evaluated using measures such as MAE, RMSE, and MASE.

Since different approaches offer varying degrees of accuracy, it is crucial to choose the proper algorithm and model it in accordance with the requirements. Therefore, the current study's objectives are to identify, look at, analyze, categorize, and estimate the amount of rainfall based on the knowledge gaps in the past literature.

III.Proposed Work

3.1 Process of ANN

Due to the fact that ANN learning is based on the design and operation of biological neural networks, it has an impact on the synaptic connections that exist in the neuronal brain centers. ANNs are built with a large number of input and output nodes, hidden layers, epochs, weights, and learning algorithms. Region selection for input data and parameters is required while developing an ANN model for weather forecasting. For analyzing nonlinear data, hidden layers are required. Each neuron in the neural network has activation functions attached to it so that it can recognise its output given a certain input. The computation and processes of Artificial neural network (ANN) are described in detail as follows:

(i) An artificial neural network offers the capability of signal transmission between layers via a threshold-based data aggregation technique.

(ii) Because the diversified field of concentration will not work, the learning model's effectiveness is increased by adding two more hidden layers, which makes it easier to see the true results.

(iii) A back propagation mechanism is employed for multilayer perceptrons and utilized to account for the erroneous values.

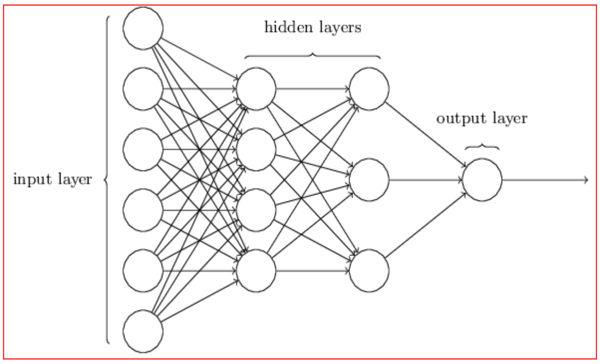

Fig.1 Artificial Neural Network

The pre-processing and cleaning of the input data follow.

3.2 Data preprocessing

(i) Eliminate null values

When a value is missing, it is designated as a null value. The learning model's performance and prediction accuracy suffer significantly during the data gathering and summing procedure. To avoid the prediction returning inaccurate results, the unknown data is thus eliminated from the dataset.

(ii) Encoding

The input is transformed into integers in order to supply the processing layer of the ANN model with the clear category mean data.

(iii) Filtering

The computations performed by the system may be incorrect as a result of the dataset's duplicated data, and the sensitive data may end up getting smashed into a wide range. The entire data set is filtered in order to ensure that the learning model functions properly.

(iv) Increasing hidden layer

Values at the input and output layers must be approximated for function. With more hidden layers, accuracy will rise. As the quantity grows, the quality gets better.

(v) Employ backpropagation

The weights of a neural set must be modified based on the error rate to increase dependability. It is referred to as reverse propagation or backpropagation.

The complete rainfall dataset extracted from kaggle were split into training and testing datasets for this investigation. The input data considers a meteorological region and a wide range of input factors, such as temperature, relative humidity, air pressure, wind direction and speed, the amount and height of clouds, rainfall, etc. The ANN model developed on its own using the anticipated results. Target outputs are offered for each item of input data. The neural network model analyses the input data first utilizing random weight values and an appropriate activation function with two hidden layers in between, in order to achieve the desired outcome. The same input dataset's target and anticipated outputs are then compared. Goal output is deducted from projected production to determine error. Based on this error, the weights are changed, and the process is repeated until the error is negligible or falls within acceptable bounds[13].

IV.Results and Discussion

In machine learning, a dataset is a group of data points that a computer may examine and forecast as a single entity (Fig 2). To a machine that doesn't interpret data the same way that people do, this means that the data collected should be standard and intelligible. 14000 chunks of data have been trained altogether for the dataset. After the data has finished being trained and the error has fallen below the tolerance criteria, testing is conducted with the validation ratio of 7:3 (train:validate).

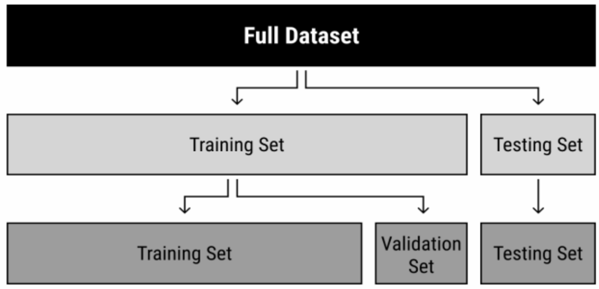

Fig.2 Data Set

The input data must originate from the same domain as the training and testing sets for the model to deliver accurate results. The neural network receives training using 70% of the input data. The remaining 30% of the input data are utilized for testing after the model has been trained using these observed data to anticipate the weather. Then, it is compared between the testing result and the desired output using the loss factor and accuracy of the model.The recent and historical climatic changes are facilitated by trend analysis, but future forecasting is not. This strategy produces more logical and trustworthy outcomes. The results are shown in figure 3 with the dataset information.

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	...	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow
0	2008-12-01	Albury	13.4	22.9	0.6	NaN	NaN	W	44.0	W	...	71.0	22.0	1007.7	1007.1	8.0	NaN	16.9	21.8	No	No
1	2008-12-02	Albury	7.4	25.1	0.0	NaN	NaN	WNW	44.0	NNW	...	44.0	25.0	1010.6	1007.8	NaN	NaN	17.2	24.3	No	No
2	2008-12-03	Albury	12.9	25.7	0.0	NaN	NaN	WSW	46.0	W	...	38.0	30.0	1007.6	1008.7	NaN	2.0	21.0	23.2	No	No
3	2008-12-04	Albury	9.2	28.0	0.0	NaN	NaN	NE	24.0	SE	...	45.0	16.0	1017.6	1012.8	NaN	NaN	18.1	26.5	No	No
4	2008-12-05	Albury	17.5	32.3	1.0	NaN	NaN	W	41.0	ENE	...	82.0	33.0	1010.8	1006.0	7.0	8.0	17.8	29.7	No	No
5	2008-12-06	Albury	14.6	29.7	0.2	NaN	NaN	WNW	56.0	W	...	55.0	23.0	1009.2	1005.4	NaN	NaN	20.6	28.9	No	No
6	2008-12-07	Albury	14.3	25.0	0.0	NaN	NaN	W	50.0	SW	...	49.0	19.0	1009.6	1008.2	1.0	NaN	18.1	24.6	No	No
7	2008-12-08	Albury	7.7	26.7	0.0	NaN	NaN	W	35.0	SSE	...	48.0	19.0	1013.4	1010.1	NaN	NaN	16.3	25.5	No	No

Fig.3 Dataset Information

A view on detailed parameters used for the rainfall prediction.The data fetched from the kaggle is shown in Fig 4 at the input layer stage. In the preprocessing stage the data is encoded and null value is removed.

#	Column	Non-Null Count	Dtype
0	Date	145460 non-null	object
1	Location	145460 non-null	object
2	MinTemp	143975 non-null	float64
3	MaxTemp	144199 non-null	float64
4	Rainfall	142199 non-null	float64
5	Evaporation	82670 non-null	float64
6	Sunshine	75625 non-null	float64
7	WindGustDir	135134 non-null	object
8	WindGustSpeed	135197 non-null	float64
9	WindDir9am	134894 non-null	object
10	WindDir3pm	141232 non-null	object
11	WindSpeed9am	143693 non-null	float64
12	WindSpeed3pm	142398 non-null	float64
13	Humidity9am	142806 non-null	float64
14	Humidity3pm	140953 non-null	float64
15	Pressure9am	130395 non-null	float64
16	Pressure3pm	130432 non-null	float64
17	Cloud9am	89572 non-null	float64
18	Cloud3pm	86102 non-null	float64
19	Temp9am	143693 non-null	float64
20	Temp3pm	141851 non-null	float64
21	RainToday	142199 non-null	object
22	RainTomorrow	142193 non-null	object

#	Column	Non-Null Count	Dtype
0	Date	145460 non-null	datetime64[ns]
1	Location	145460 non-null	int64
2	MinTemp	145460 non-null	float64
3	MaxTemp	145460 non-null	float64
4	Rainfall	145460 non-null	float64
5	Evaporation	145460 non-null	float64
6	Sunshine	145460 non-null	float64
7	WindGustDir	145460 non-null	int64
8	WindGustSpeed	145460 non-null	float64
9	WindDir9am	145460 non-null	int64
10	WindDir3pm	145460 non-null	int64
11	WindSpeed9am	145460 non-null	float64
12	WindSpeed3pm	145460 non-null	float64
13	Humidity9am	145460 non-null	float64
14	Humidity3pm	145460 non-null	float64
15	Pressure9am	145460 non-null	float64
16	Pressure3pm	145460 non-null	float64
17	Cloud9am	145460 non-null	float64
18	Cloud3pm	145460 non-null	float64
19	Temp9am	145460 non-null	float64
20	Temp3pm	145460 non-null	float64
21	RainToday	145460 non-null	int64
22	RainTomorrow	145460 non-null	int64

Fig.4 Preprocessed Data at the input layer

The empirical relationships of the normal distribution phenomenon allow us to calculate the mean and standard deviation of a given sample and to establish the threshold for categorizing outliers as being more than three standard deviations from the mean. Outliers are cases that differ from the predetermined lower and higher boundaries. Trimming and extraction are finished procedures. In the end, we evaluated the outlier elimination step's prediction performance using the ANN. An example is also shown in Figure 5.

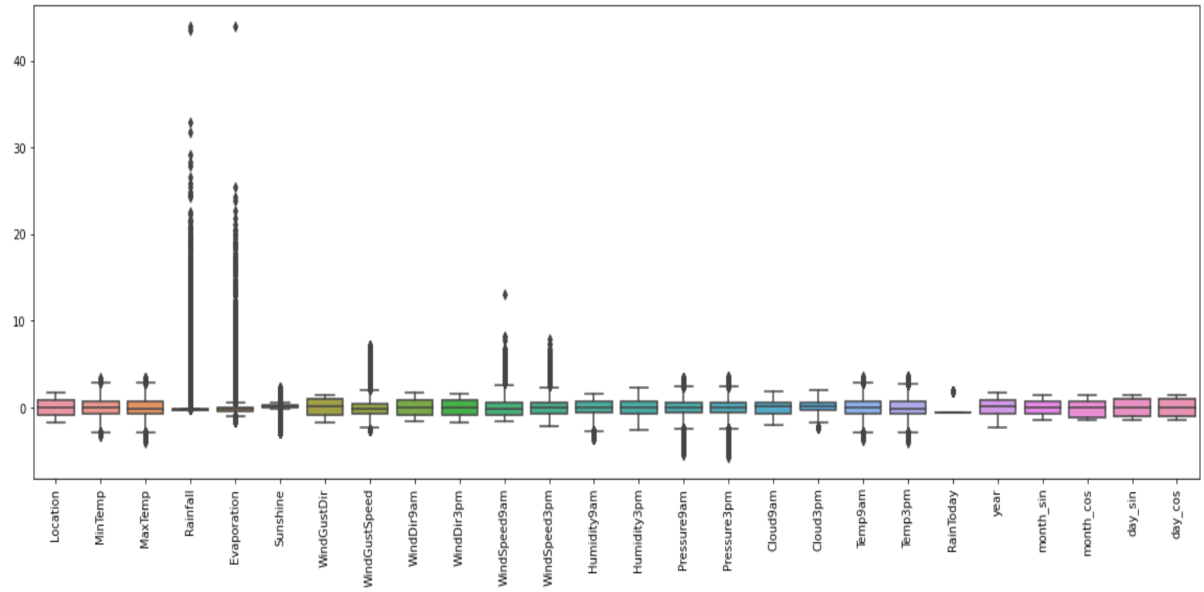

Fig.5 Outlier removal phase

Then the capping is done with the boundary constraints for the prediction score and it is shown in below figure 6.

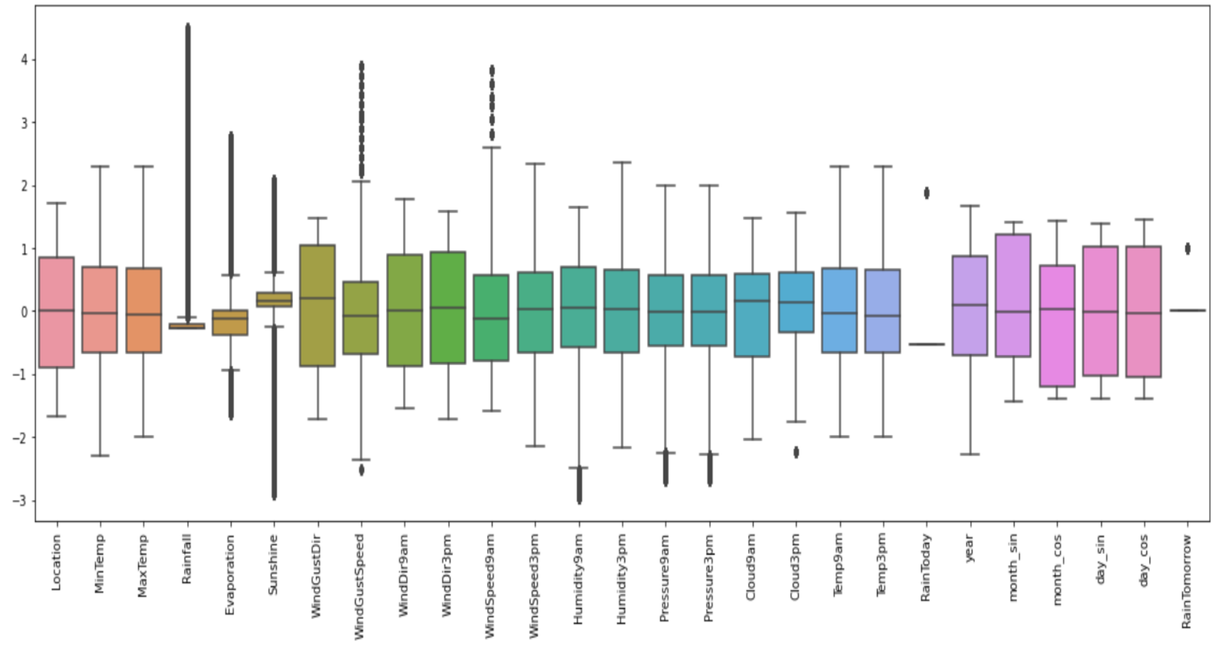

Fig.6 Prediction Score

Utilizing the training data correctly in both forward and backward passes from the input layer to the output layer and the input layer to the output layer is the first step in creating a neural network model. The utilized ANN algorithm's number of epochs is considered as its hyperparameter. In figure 7, which is below, the rainfall prediction is depicted together with the degree of accuracy for the number of epochs.

The number of epochs is crucial for generating findings with a high degree of agreement between projected and actual output data. If there are several iterations, the test data is confirmed and demonstrated to have a high accuracy level.If the number of epochs is high while using an ANN model, the attempt to minimize the loss is also accomplished simultaneously. Figure 8 depicts it.

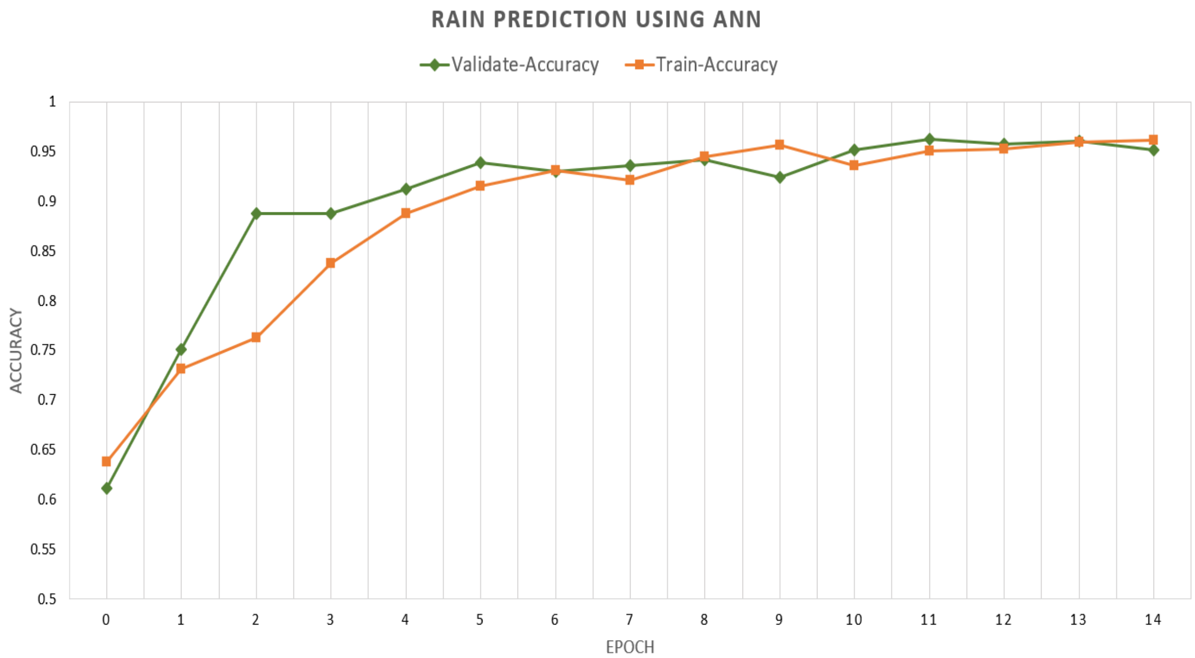

Fig.7 Accuracy validation for number of epochs

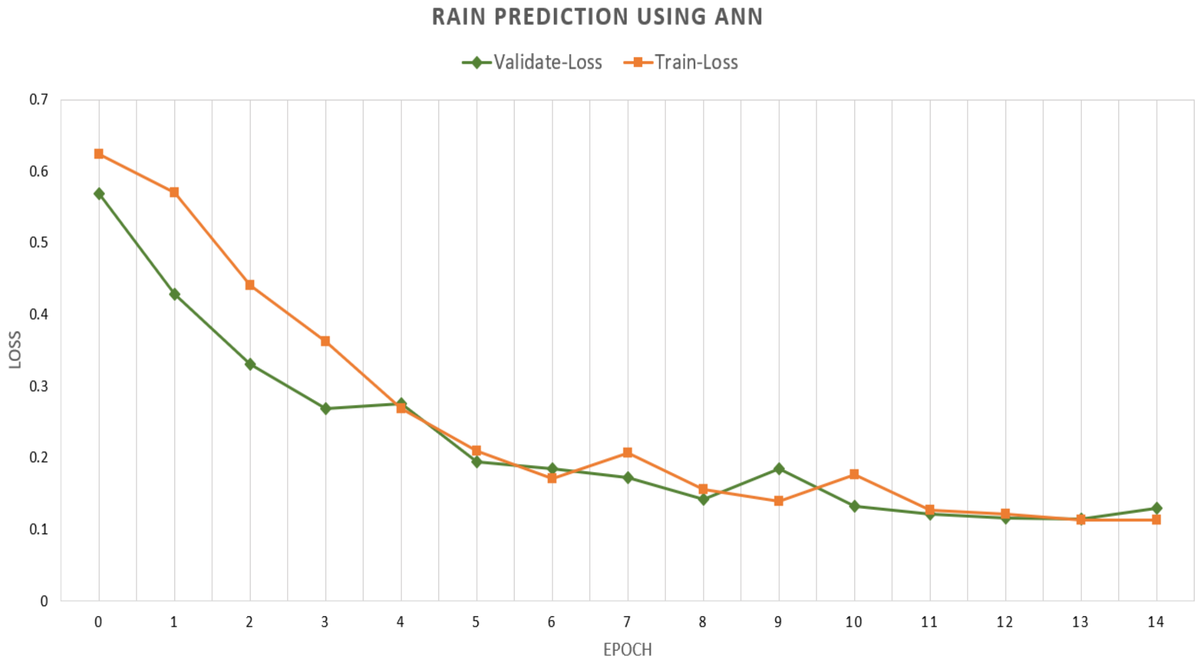

Fig.8 Loss factor for number of epochs

V.Conclusion

The responsibility of forecasting the weather is one of the most crucial and difficult ones for meteorological agencies worldwide. So the rainfall prediction based on a learning model weighs a lot of scope. In this research the results are accomplished using the Artificial Neural Network by considering it as the predominant algorithm. It is considered as more efficient than many hardware based flood monitoring systems which reports and analyzes the data in a static format. The rainfall dataset extracted from the kaggle repository is effectively utilized by categorizing it to 70% as training data and 30% as testing data. The regularization is done by the proper data preprocessing techniques like encoding and elimination of null values. Inorder to improve the performance of the ANN algorithm and its prediction results the quantity of neurons is increased with the incremental number of hidden layers in ANN architecture. The results were validated by the distribution properties like increased number of epochs which improved the efficiency of loss factor minimization. Using a deployed trained ANN model, it is proven that it further increased the level of rainfall forecast accuracy even higher.

REFERENCES

[1] M.J.C., Hu, Application of ADALINE system to weather forecasting, Technical Report, Stanford Electron, 1964

[2] Michaelides, S. C., Neocleous, C. C. & Schizas, C. N. "Artificial neural networks and multiple linear regression in estimating missing rainfall data." In: Proceedings of the DSP95 International Conference on Digital Signal Processing, Limassol, Cyprus. pp 668–6731995.

[3] Kalogirou, S. A., Neocleous, C., Constantinos, C. N., Michaelides, S. C. & Schizas, C. N.,"A time series construction of precipitation records using artificial neural networks. In: Proceedings of EUFIT '97 Conference, 8–11 September, Aachen, Germany. pp 2409–2413 1997.

[4] Lee, S., Cho, S.& Wong, P.M.,"Rainfall prediction using artificial neural network.",J. Geog. Inf. Decision Anal. 2, 233–242 1998.

[5] Wong, K. W., Wong, P. M., Gedeon, T. D. & Fung, C. C. , "Rainfall Prediction Using Neural Fuzzy Technique." 1999

[6] E.Toth, A.Brath, A.Montanari," Comparison of short-term rainfall prediction models for

real-time flood forecasting", Journal of Hydrology 239 (2000) 132–147

[7] Koizumi, K.: "An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network", Weather Forecast., 14, 109–118, 1999.

[8] N. Q. Hung, M. S. Babel, S. Weesakul, and N. K. Tripathi "An Artificial Neural network Model for rainfall Forecasting In Bangkok,Thailand", Hydrol. Earth Syst. Sci., 13, 1413–1425, 2009

[9] Kyaw Kyaw Htike and Othman O. Khalifa, "Research paper on ANN model using focused time delay learning", International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur.

[10] Ajith Abraham, Dan Steinberg and Ninan Sajeeth Philip,"Rainfall Forecasting Using Soft Computing Models and Multivariate Adaptive Regression Splines", 2001.

[11] Rainfall data from http://www.Indiastat.com/karnataka/rainfall

[12] Ben Krose and Patrick van der Smagt , "An introduction to neural networks", Eighth edition, November 1996.

[13] Dr S. Santosh Baboo and I. Khadar Shareef, "An efficient Weather Forecasting Model using Artificial Neural Network", International Journal of Environmental Science and Development, Vol. 1, No. 4, October 2010.

[14] Enireddy Vamsidhar et. al.,"Prediction of rainfall Using Backpropagation Neural Network Model", International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010, 1119-1121

[15] Paras, Sanjay Mathur, Avinash Kumar, and Mahesh Chandra, "A feature based on weather prediction using ANN"World Academy of Science, Engineering and Technology 34 2007

[16] Yu, C.; Li, C.; Xin, Q.; Chen, H.; Zhang, J.; Zhang, F.; Li, X.; Clinton, N.; Huang, X.; Yue, Y.; et al. Dynamic assessment of the impact of drought on agricultural yield and scale-dependent return periods over large geographic regions. Environ. Model. Softw. 2014, 62, 454–464.

[17] Hayes, M.J.; Svoboda, M.D.; Wardlow, B.D.; Anderson, M.C.; Kogan, F. Drought monitoring: Historical and current perspectives. In Remote Sensing of Drought: Innovative Monitoring Approaches; CRC Press: Boca Raton, FL, USA, 2012; pp. 1–19. ISBN 9781439835609.

[18] Yusuf, A.A.; Francisco, H. Climate change vulnerability mapping for Southeast Asia vulnerability mapping for Southeast Asia. East 2009, 181, 1–19.

[19] Mishra, S.S.; Nagarajan, R. Forecasting drought in Tel River Basin using feedforward recursive neural network. Int. Conf. Environ. Biomed. Biotechnol. 2012, 41, 122–126.

[20] Pulwarty, R.S.; Sivakumar, M.V. Information systems in a changing climate: Early warnings and drought risk management. Weather Clim. Extrem. 2014, 3, 14–21.

[21] Khashei, M.; Bijari, M. An artificial neural network (p, d, q) model for timeseries forecasting. Expert Syst. Appl. 2009, 37, 479–489.

[22] Barua, S.; Ng, A.W.M.; Perera, B.J.C. Artificial Neural Network–Based drought forecasting using a nonlinear aggregated drought index. J. Hydrol. Eng. 2012, 17, 1408–1413.

[23] Hard Winarto, S.; Aipassa, M. Rainfall monthly prediction based on Artificial Neural Network: A case study in Tenggarong station, East Kalimantan–Indonesia. Procedia Comput. Sci. 2015, 59, 142–151.

[24] Morid, S.; Smakhtin, V.; Bagherzadeh, K. Drought forecasting using artificial neural networks and time series of drought indices. Int. J. Climatol. 2007, 27, 2103–2111.

[25] Liou, Y.A.; Liu, S.F.; Wang, W.J. Retrieving soil moisture from simulated brightness temperatures by a neural network. IEEE Trans. Geosci. Remote Sens. 2001, 39, 1662–1672.

[26] Belayneh, A.; Adamowski, J.; Khalil, B. Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods.

Sustain. Water Resour. Manag. 2016, 2, 87–101. [CrossRef]

[27] Deo, R.C.; ¸Sahin, M. Application of the Artificial Neural Network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. Atmos. Res. 2015, 161–162, 65–81.

[28] Le, M.H.; Perez, G.C.; Solomatine, D.; Nguyen, L.B. Meteorological drought forecasting based on climate signals using Artificial Neural Network—A case study in Khanhhoa Province Vietnam. Procedia Eng. 2016, 154, 1169–1175.

[29] Schubert, S.D.; Stewart, R.E.; Wang, H.; Barlow, M.; Berbery, E.H.; Cai, W.; Hoerling, M.P.; Kanikicharla, K.K.; Koster, R.D.; Lyon, B.; et al. Global meteorological drought: A synthesis of current understanding with a focus on SST drivers of precipitation deficits. J. Clim. 2016, 29, 3989–4019.

[30] Roundy, J.K.; Wood, E.F.; Roundy, J.K.; Wood, E.F. The attribution of land–Atmosphere interactions on the seasonal predictability of drought. J. Hydrometeorol. 2015, 16, 793–810.

[31] Lyon, B. Seasonal drought in the greater horn of Africa and its recent increase during the March–May long rains. J. Clim. 2014, 27, 7953–7975.

[32] http://neuralnetworksanddeeplearning.com/chap1.html

eISSN 1303-5150 www.neuroquantology.com