RAINFALL PREDICTION FOR SOCIO-ECONOMIC BENEFITS USING MACHINE LEARNING OVER GIRD REGION OF INDIA

[1]Sudipti Banerjee, [2]Amit Kumar Manjhvar

[1]Research Scholar, [2]Asst. Professor.

[1,2] Department of Computer Science Engineering & Information Technology, Madhav Institute of Technology and Science, Gwalior, India

Abstract : Rainfall prediction aims to determine rainfall over a particular location. It is important for agriculture and other industry as well as the economy of the country. Accuracy in prediction could decrease the number of fatalities. Correct and in time information about rainfall can be helpful to take a decision on an on-going construction project, transportation business, flight navigation, etc. In the current work, a Hybrid Neural Network (HNN) model is proposed for the prediction of rainfall. HNN model gains the accuracy of 93.74%. The dataset used in this work is of Gwalior region over the period from 1990 to 2018.

Index Terms **: Rainfall Prediction, Artificial Neural Network (ANN), Clustering, Classification, Expert Models, Levenberg-Marquardt Backpropagation, sequentialfs.**

I. INTRODUCTION

One of the vital things of nature is Rainfall, which acts as bane & boon both for mankind.

Prediction of rainfall would be of great help. Many related types of research have been done. Natural Calamities like drought, a flood is all affected by rainfall. Thus, there is a need to predict rainfall accurately.

There are several parameters affecting rainfall, some are:- Temperature, Wind direction, Cloud Cover, etc.

Broad Classification of Rainfall can be done in 2 categories:-

First by observing physical relation between the parameters responsible for rainfall in a particular area. This approach is temporal and spatial as well as not computationally feasible as it involves lengthy mathematical calculations.

The second approach is uncovering unseen patterns of different parameters that affect rainfall, how they are literally related to physical rainfall. This approach is much more feasible.

Many research has validated that a knowledgeable system based on Neural Network (NNs) are remarkably accurate and robust in solving real-life problems.

Prediction of the class label of a hidden sample can be very well done by a trained ANN.

The most widely applied method for rainfall prediction is ANN.

They can catch and signify the complex relations of the data being modeled, as they are powerful and flexible. Many comparison works have been done of models based on ARIMA, MLP-FFN. These models are suitable and reliable.

In ANN different learning algorithm can be applied to train the model. The learning algorithm tries to find out the best set of weights for the NN connections to reach the goal/target set. In the Training phase, an error function is usually reduced.

In the current work, an HNN model has been proposed to forecast rainfall based on parameters like Temperature, Cloud Cover, Humidity, etc.

The hourly data of Gwalior region (India) is collected over a period from 1990 to 2018. In the training phase, first, the dataset is divided into 2 clusters via k-means. Succeeding, each cluster is trained using distinct NN. The experimental results established the lead of the proposed model over the well-known perceptron neural model.

The remaining paper is as tails, section 2 illustrate the proposed HNN model in detail, including the training process. Next section 3, the experimental method is discussed and in the last section, the results of the experiment have been testified for the conclusion.

II. PROPOSED METHODOLOGY

In this work, an HNN model has been proposed in which different NN is used for each cluster to train them to identify the patterns of that cluster only.

Steps are as follows:-

1. In the first step, the greedy feature selection algorithm (sequential feature selector) is applied to extract out the best promising features from the raw dataset.

2. NN based classification-
   The ANN is one amongst the foremost used approaches. It attains correct cataloging even with the terribly tiny dataset. It will deal with the general relationships throughout its training stage. The ANN structure is consisting of interrelated computational neurons, which concerned in the mathematical mapping through the learning method, that plan to regulate the weight value. Primarily, the training section is started by a locality of the dataset to classify its inputs alongside its class label to form the classification model. Subsequently, the validation section is performed to verify the efficiency of the trained model exploitation another dataset. Finally, the analysis section is employed to check the cataloging model accuracy exploiting another set of test data.

3. K-means clustering
   The clustering can be stated as a segregating problem where a set of points are to be allocated into clusters such that the squared error among the cluster centers and the rest of the cluster is reduced. k-means (k=2) clustering algorithm is used. After that separate NN is used to train each cluster.

4. Hybrid Neural Network model
   In the proposed HNN model first clustering is done on data using the k-means algorithm. After that separate NN is employed on each cluster and trained to build the model.

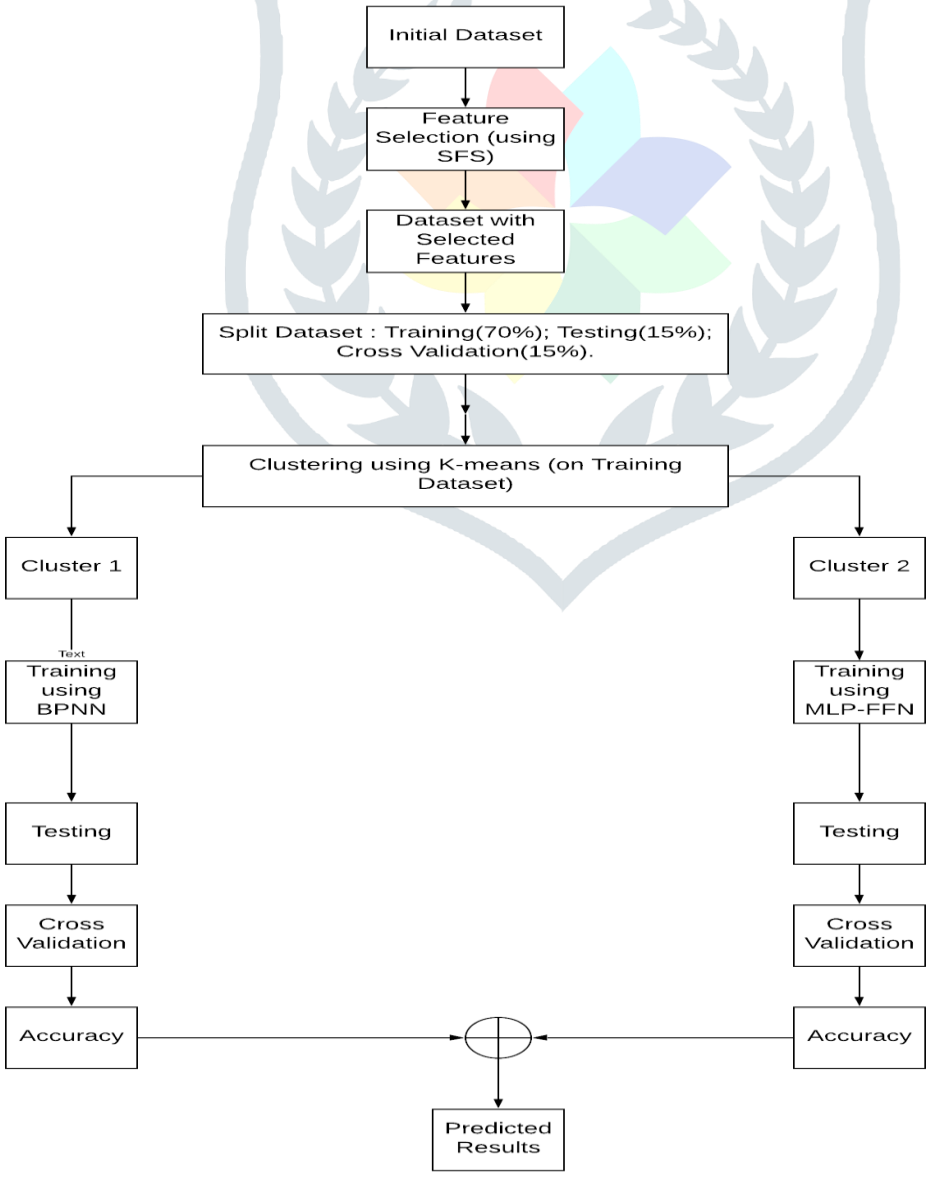

Fig. 1. Method Flow Diagram of Current Study including feature selection.

A comparison work is also carried to prove that feature selection approach gives better performance. Below is a flow diagram which states the working of the model without feature selection.

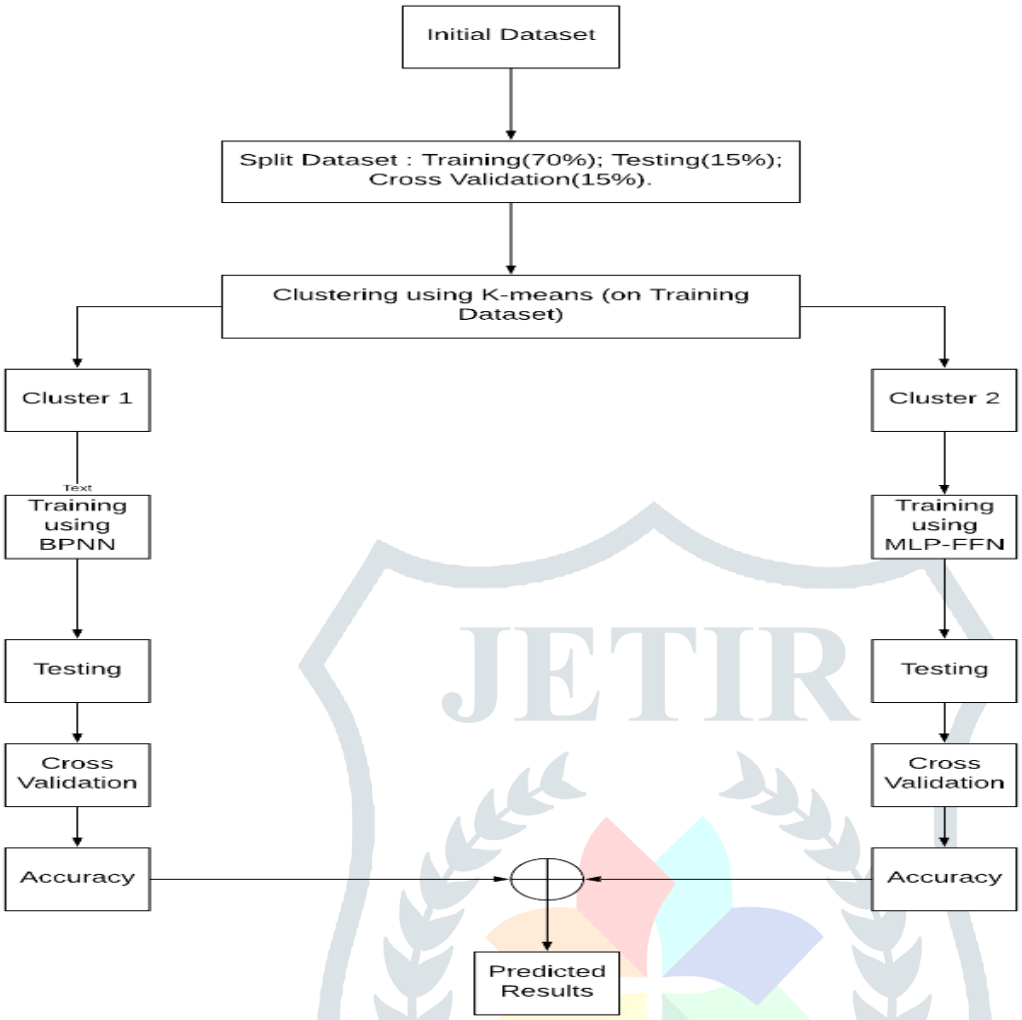

Fig. 2. Method Flow Diagram of Current Study without feature selection.

III. EXPERIMENTAL METHODOLOGY

Hourly dataset of rainfall over Gwalior region having attributes as humidity, cloud cover, precipitation, temperature, etc, is gathered from the year 1990 to the year 2018. Features of the dataset are described in Table 1.

Attributes	Description
Year	Duration of Data (in Years)
Month	Duration of Data (in Months)
Day	Duration of Data (in Days)
Hour	Duration of Data (in Hour)
Minute	Duration of Data (in Minute)
Temperature	Average Temperature of that region
Humidity	Relative Humidity of that region
Sea Level Pressure	Mean Sea Level Pressure of that region
Cloud Cover	Total Cloud Cover over that region
Wind Speed	Average Speed of Wind in that region
Wind Direction	Direction of Wind
Precipitation	Total Rainfall received in that region

Table. 1. Attribute Set of Raw Dataset.

Here attribute Precipitation is set as Target, and among the rest attributes, feature selection is done to determine the most promising features.

First Sequential Forward Selection (SFS) algorithm, a member of the greedy forward selection method, is applied on the dataset to extract best promising features for predicting rainfall, as an improper/ill-suited set of features have an adverse effect on the

performance of ANNs. SFS first consider each feature as a single set of features then it keeps on adding new features based on a greedy improvement approach.

Table 2 shows the features selected in a tabular manner.

Attributes	Description
Temperature	Average Temperature of that region
Humidity	Relative Humidity of that region
Sea Level Pressure	Mean Sea Level Pressure of that region
Wind Speed	Average Speed of Wind in that region
Wind Direction	Direction of Wind

Table. 2. Attribute Set after performing Feature Selection.

Now, these top 5 selected attributes (Humidity, Sea Level Pressure, Temperature, Wind Speed & Wind Direction) are considered as Inputs to the Neural Network.

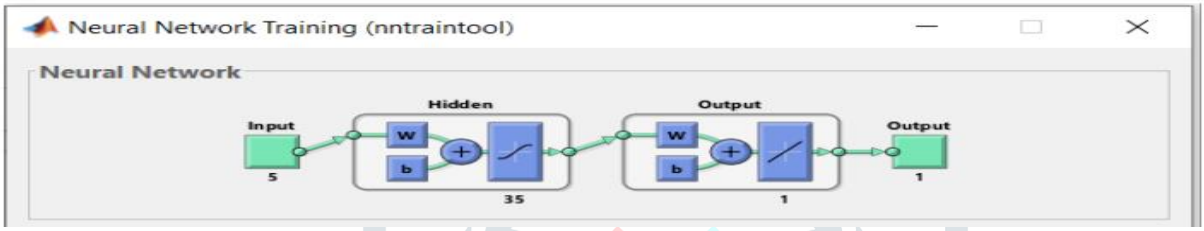

Fig. 3. Shows Neural Network Structure.

After that the dataset is grouped into 2 clusters using the K-means algorithm (let k=2), Next, the data of each cluster is trained using a separate neural network. Here the Learning Function used is Levenberg-Marquardt which has a property of reducing error function while transferring data in between the hidden layers. This acts as Local error minimization and optimized results are obtained.

Data is divided into a training dataset (70% of feature selected dataset), testing dataset (15% of feature selected dataset) and dataset for cross-validation (15% of feature selected dataset).

Measuring parameters are described as Accuracy, Recall, Precision & F-Measure.

Accuracy=(trp+trn)/(trp+trn+flp+fln)

Recall=trp/(trp+fln)

Precision=trp/(trp+flp)

F-Measure=2*((Recall*Precision)/(Recall+Precision))

Where trp: true positive

trn: true negative

flp: false positive

fln: false negative, generally the constituents of confusion matrix.

IV. EXPERIMENTAL RESULTS

The experiment is executed by following the steps mentioned in section III.

Table 3 states the Performance Measure of HNN Models with and without Feature Selection. Before feature selection, the HNN Model gains 86.26% Accuracy, 88.34% Precision, 94.24% Recall, and 91.19 F-Measure. However, after performing feature selection, better results were obtained 93.74% Accuracy, 88.48% Precision, 94.49% Recall and 91.39% F-Measure.

Figure 4 shows Best Validation Performance is 0.14468 at epoch 46.

Figure 5 shows the regression values at different stages. Regression value at Training Stage is R=0.557 with Output equation=0.31*Target+0.2. Regression value at Testing Stage is R=0.566 with Output equation=0.32*Target+0.2. Regression value at Validation Stage is R=0.546 with Output equation=0.3*Target+0.21. Overall Regression value is R=0.557 with Output equation=0.31*Target+0.2

Performance Measure	HNN (without feature selection)	HNN (with feature selection)
Accuracy	86.26	93.74
Precision	88.34	88.48
Recall	94.24	94.49
F-Measure	91.19	91.39

Table. 3. Performance Measures

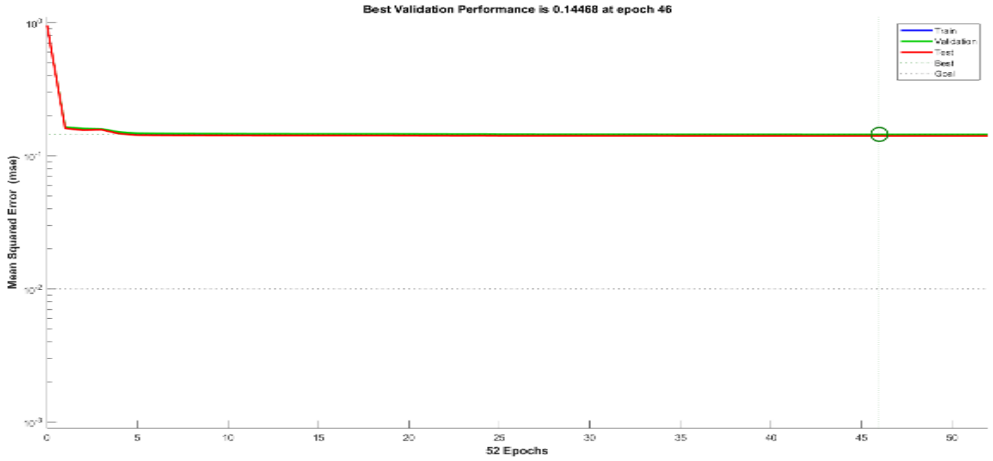

Figure. 4. A plot of MSE vs. epochs

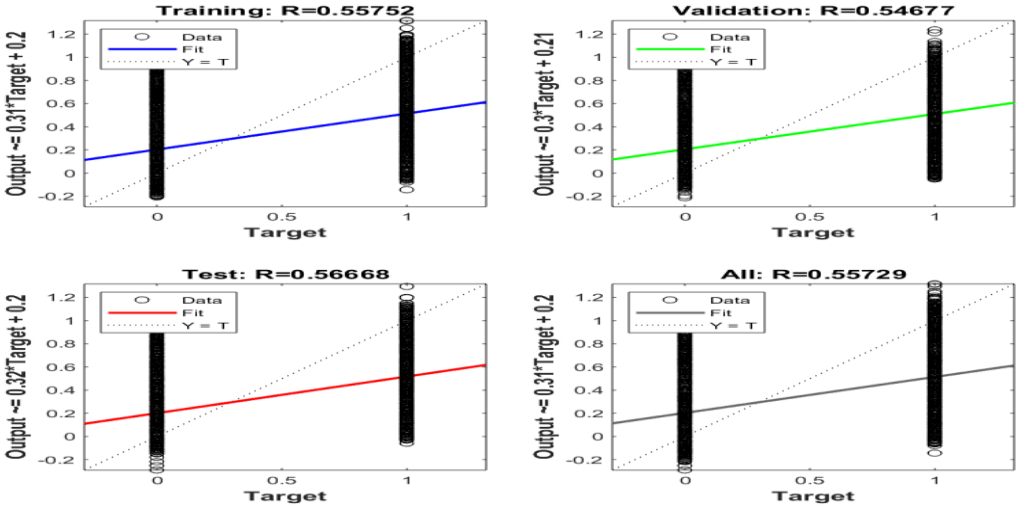

Figure. 5.

Regression plot of different stages.

V. CONCLUSION

Rainfall plays an important part in determining the weather condition of a local region as well as natural disasters such as Drought, Flood, etc. Advance knowledge about weather condition will benefit Agricultural sectors and take precautionary steps if needed. The national economy is also benefitted directly by this. There is a need for an accurate & robust model to predict rainfall. In the present work HNN, based prediction model for predicting rainfall has been proposed. Experimental results have recommended that while predicting rainfall, feature selection can sensibly progress the performance. In the present work, high accuracy of 93.74% is achieved while predicting rainfall.

VI. REFERENCES

[1] Nayak, D. R., Mahapatra, A., & Mishra, P. (2013). A survey on rainfall prediction using an artificial neural network. International Journal of Computer Applications, 72(16).

[2] Dubey, A. D. (2015). Artificial neural network models for rainfall prediction in Pondicherry. International Journal of Computer Applications, 120(3).

[3] Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A. S., &Balas, V. E. (2016). Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. Neural Computing and Applications, 1-12.

[4] Chatterjee, S., Ghosh, S., Dawn, S., Hore, S., &Dey, N. (2016). Forest Type Classification: A Hybrid NN-GA Model-Based Approach. In Information Systems Design and Intelligent Applications (pp. 227-236). Springer India.

[5] Knezevic, M., Cvetkovska, M., &Trombeva-Gavriloska, A. (2014). APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN CIVIL ENGINEERING. Tehnickivjesnik/Technical Gazette, 21(6).

[6] Socha, K., & Blum, C. (2007). An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training. Neural Computing and Applications, 16(3), 235-247.

[7] Dehuri S, Cho SB (2010) A hybrid genetic-based functional link artificial neural network with a statistical comparison of classifiers over multiple datasets. Neural Comput Appl 19(2):317–328

[8] MacIntyre J (2013) Applications of neural computing in the twenty-first century and 21 years of neural computing and applications. Neural Comput Appl 23(3–4):657–665

[9] Azar AT, El-Said SA, Balas VE, Olariu T (2013) Linguistic hedges fuzzy feature selection for differential diagnosis of Erythematous-Squamous diseases. Soft Comput Appl AISC 195:487–500

[10] Awan SM, Aslam M, Khan ZA, Saeed H (2014) An efficient model based on artificial bee colony optimization algorithm with neural networks for electric load forecasting. Neural Comput Appl 25(7–8):1967–1978

[11] Siddiquee MSA, Hossain MMA (2015) Development of a sequential artificial neural network for predicting river water levels based on Brahmaputra and Gange's water levels. NeuralComputAppl 26(8):1979– 1990

[12] Cao Z, Cheng L, Zhou C, Gu N, Wang X, Tan M (2015) Spiking neural network-based target tracking control for autonomous mobile robots. Neural Comput Appl 26(8):1839–1847

[13] Gao S, Ning B, Dong H (2015) Adaptive neural control with intercepted adaptation for time-delay saturated nonlinear systems.NeuralComputAppl 26(8):1849–1857

[14] Ciancio C, Ambrogio G, Gagliardi F, Musmanno R (2015)Heuristic techniques to optimize neural network architecture in manufacturing applications. Neural Comput Appl. doi:10.1007/s00521-015-1994-9

[15] Mirjalili SZ, Saremi S, Mirjalili SM (2015) Designing evolutionary feedforward neural networks using social spider optimization algorithm. Neural Comput Appl 26(8):1919–1928

[16] Han J, Kamber M (2005) Data mining: concepts and techniques, 2nd edn. Morgan and Kaufmann, San Francisco, pp 285–378

[17] Nanda, S. K., Tripathy, D. P., Nayak, S. K., &Mohapatra, S. (2013). Prediction of rainfall in India using Artificial Neural Network (ANN) models. International Journal of Intelligent Systems and Applications, 5(12), 1.

[18] Hung, N. Q., Babel, M. S., Weesakul, S., &Tripathi, N. K. (2009). An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrology and Earth System Sciences, 13(8), 1413-1425.

[19] Maren, A. J., Harston, C. T., & Pap, R. M. (2014). Handbook of neural computing applications. Academic Press.

[20] Baughman DR, Liu YA (2014) Neural networks in bioprocessing and chemical engineering. Academic Press.

[21] Rojas, R. (2013). Neural networks: a systematic introduction. Springer Science & Business Media.

[22] Dash, R. N., Panigrahy, C., &Subudhi, B. (2015, January). A comparison between RNN and RBF NN techniques for the detection of stator inter-turn fault of an induction motor. In Electrical, Electronics, Signals, Communication and Optimization (EESCO), 2015 International Conference on (pp. 1-7). IEEE.

[23] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100-108.

[24] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666.

[25] Zhang T (2009) On the consistency of feature selection using greedy least squares regression. J Mach Learn Res 10:555–568.

[26] Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182.

[27] Karayiannis, N., &Venetsanopoulos, A. N. (2013). Artificial neural networks: learning algorithms, performance evaluation, and applications (Vol. 209). Springer Science & Business Media.

[28] Dutta, B., Ray, A., Pal, S., Patranabis, D.C. (2009) A connectionist model for rainfall prediction. Neural, Parallel and Scientific Computations, vol. 17, pp. 47-58.

[29] Chatterjee, S., Dey, N., Shi, F., Ashour, A. S., Fong, S. J., & Sen, S. (2017). Clinical application of modified bag-of-features coupled with hybrid neural-based classifier in dengue fever classification using gene expression data. Medical & Biological Engineering & Computing, 1-1.

[30] Ieeexplore.ieee.org. (2019). *Rainfall prediction using hybrid neural network approach - IEEE Conference Publication*. [online] Available at: https://ieeexplore.ieee.org/abstract/document/8325807 [Accessed 9 Jun. 2019].

[31] Sudipti Banerjee, Amit Kumar Manjhvar, , , , , , ."*Comparative Analysis of Machine Learning Techniques for Rain Prediction"*, Volume 7, Issue II, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: , ISSN : 2321-9653,www.ijraset.com

[32] Sharma, Bhavna & K. Venugopalan, Prof. (2014). Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images. IOSR Journal of Computer Engineering. 16. 31-35. 10.9790/0661-16123135.