An Approach for Rainfall Prediction using Soft Computing

K C Gouda[#1], Libujashree R[*],Priyanka Kumari[*],Manisha Sharma[*], Ambili D Nair[*]
[#]*Senior Scientist, CSIR Fourth Paradigm Institute,*
Wind Tunnel Road, Bangalore -37, India
[*]*Department of ECE, MVJ College of Engineering, Near ITPB, Whitefield, Bangalore-560 067*

Abstract

Rainfallis an important weather parameter and it has direct impacts on different sectors like agriculture, health, water management etc. and also controls the economy of the country. Future predictions of rainfall in water-scarce region are highly important for effective water resource management particularly over country like India. An accurate rainfall forecasting is also very much needed for agriculture dependent countries like India. Statistical techniques for rainfall forecasting cannot perform accurately. Rainfall being highly non-linear and complicated phenomena it requires to use the state-of-the-artSoft computing techniques likeArtificial Neural Network (ANN) for the accurate and advanced forecasting of rainfall. In this work, the back-propagation algorithm is used to predict the rainfall over Bangalore city in India using the ANN in the super computer platform. The back-propagation model developed is trained and validated against actual rainfall of the region. Results showed that ANNs could perform well for predicting long-term the rainfall with acceptable accuracy.

Keywords - *Rainfall prediction, ANN, BPNN, data mining.*

I. INTRODUCTION

Predicting rainfall is a challenging task in this era of highly unsettled climate.Generally,rainfall and climate are immensely non-linear occurence in nature showingwhich is called as "butterfly effect"[1]. Hence, in spite of many advances made in the field of weather forecast, accurate rainfall prediction is one of the greatest challenges in operation hydrology [2]. Many different techniques have been applied for the task of rainfall prediction worldwide. But the prediction accuracy obtained didn't have satisfactory result because of large uncertainties present in the climatic processes [3-4]. The concept of artificial neural networks was introduced back in 1943 [5]. But main research of applications of ANNs has begun only after the introduction of the backpropagation training algorithm for feed-forward ANNs in 1986[6]. Many soft computing techniques like Conventional Regression, Linear Regression, K Nearest Neighbour, and Back Propagation has been discussed. These real-life conditions have been found to be better elucidated by 'artificial neurons' which can learn from experience, i.e. by experiencing back propagation errors and so on . ANNs are observed assatisfactorilyrecentmechanism in the field of forecasting. Hence, in order to adopt a systematic approach in the development of ANN models, factors such as data pre-processing, determination of adequate model inputs and suitable network architecture, optimization [7] and model validation is important.

ANN has become popular because of their ability to tolerate noisy data and to predict or classify patterns they have not been trained [8]. The most common neural network is Backpropagation Neural Network.

There have been numerous data parameters affecting the rainfall, seven meteorological data parameters have been obtained from Indian Meteorological department [9]. Mined data parameters are trained for predicting rainfall of Bangalore, India. The parameters are rainfall, maximum temperature, precipitation, wet day frequency, mean temperature, relative humidity, total cloud amount and wind speed.

Scientists have discovered that Back propagation algorithm using multi-layered neural network performs better with acceptable accuracy when compared to other methods. Therefore, this paper shows a brief study of other methods and the implementation of Back propagation algorithm using multi-layered neural network for Bangalore region using 100 years data.

II. METHODS

A. Conventional Regression model and Artificial Neural Network

The Rajsamand damon Gomti river in Rajasthan was studied by Vyas et. al [10-12]. The monthly monsoonrainfall data for the period 1996 to 2015 had been used. The soft computing technique used wasconventional regression model and Artificial Neural Network model (ANN).

A functional relationship between a dependent variable and a number of independent variables is described by a Regression model. In case of run-off and rainfall this functional relationship can be a polynomial of any degree. In their research, regression equation of polynomial degree two type is shown in equation (1) has been used.

$Y = a_0 + a_1x + a_2x^2 + \varepsilon$ -------- (1)

Where, x → independent variable representing the weighted rainfall in mm,
Y→ effective runoff,
a_0→ intercept term,
a_1, a_2→ partial regression coefficients,
ε→ random error.

The coefficients are decided by minimizing the sum of squares of deviations ($\sum \varepsilon^2$).

In their study, a neural network with three layers has been used comprising four output neurons and four input neurons representing the rainfall and runoff in units of mm.The hidden layer uses tangent hyperbolic transfer function and output layer use linear transfer function.
The previous 20 years data were divided into four parts each having five years' data. The 15-year data was used to train and the remaining data for 5 years was used to validate. Using this approach 4 models namely 1V, 2V, 3V and 4V were prepared having the 4 parts of the data respectively.
Their study uses 3 different performance metrices namely, co-relation coefficient (R), root mean square error (RMSE) and mean absolute error (MAE).

B. *Linear regression and K-fold:*

Mohapatra et. al research work was based on rainfall prediction using 100 years of meteorological data [14-16]; where they have used thelinear regression method for data mining and K-fold technique for cross verification.The dependent variables like temperature, humidity, cloud pressure, etc.were focused upon to prepare the model.

→The data of Bangalore of last 100 years month wise, was divided into 3 seasons after which the data was pre-processed.

→ Regression model has been used, where the weights are actually computed from the training sets. The input data set needs to be sampled for getting the training and testing sets. They have used 80-20 fixed sampling, which means after data pre-processing, 80% will be for training and other 20% data will be used for validating the model.

→The second technique used is K-fold cross validation, where the sampling is done randomly into k equal size of samples. Out of these k sub samples, one sub sample is retained for testing the model and rest k-1 sub samples are used for training the model.

→Once all the iterations were completed for each sub samples, the result from individual iterations needs to be averaged to produce one single result.

→For validating and getting the computational results, the libraries of Python namely Pandas and SciKit learn has been used.

C. *K-Nearest Neighbour*

To overcome the limitations of ANN, recently a novel ensemble forecasting model, i.e. artificial neural network ensemble (ANNE), has been developed.
It is an effective approach to the develop a high-performance forecasting system [17- 20].

In this model, original data set had been partitioned into fewnon-identical training subgroupsthrough Bagging technology. Then the other various ANN algorithms and network architecture generate different individual network ensemble by training subsets. The incomplete least square regression is adopted to extract ensemble members. Finally, the K-nn non-parametric regression was used for ensemble model.

A neural network ensemble has beenformulated in two steps, i.e. training a variety of neural network elements and integrating the component predictions.

The author, Jiansheng Wu's building process and the design of the functional model is discussed below.

- *The Building Process of the Artificial Neural Network Ensemble Model:*

The basic ANNE configuration for rainfall forecasting includes a triple layer Feed-forward Neural Network.

Firstly, many individual neural predictors are generated. Then an appropriate number of neural predictors are chosen from the substantial number of candidate predictors. Ultimately, chosen neural predictors are consolidated into a collectionof neural predictor in a nonlinear method.

Function Model

→Initially, original dataset are divided into a fewdistinct training subgroups TR1, TR2, ⋯, TRn through Bagging algorithm which is a Machine Learning ensemble meta- algorithm designed to improve the stability and accuracy used in statistical classification and regression.
→Secondly, the discrete ANN models with similar training data are therefore generated M ensemble individuals.
→Thirdly, ensemble members are drawn out by PLS technique. Finally, K–nn is used to merge the selected individual forecasting results into an ensemble model. A schematic of same is presented in figure 1.

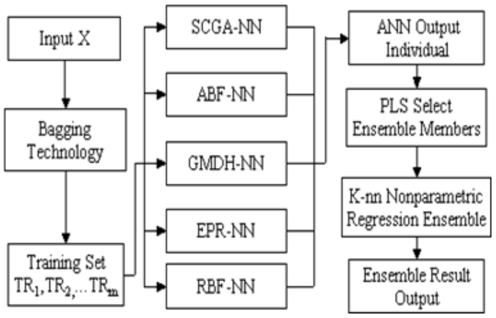

Fig 1: A flow chart of the proposed K-nn ensemble rainfall forecasting model (courtesy Jiansheng Wu [14]).

Where,
SCGA-NN→ scaled conjugate gradient algorithm neural network
ABF-NN→ adaptive basis function neural network
GMDH-NN→group method of data handling neural network
EPR-NN→ Elman partial recurrent neural network
RBF-NN→ radial basis function neural network

D. Wavelet Decomposition

Charaniya et. al have used wavelet decomposition for predicting rainfall. The transformation of time series from temporal domain to wavelet domain is regarded as wavelet analysis [21-23].

Wavelet analysis is the breaking up of a signal into the original (or mother) wavelet's shifted and scaled versions.Using a fully scalable modulated window solves the signal-cutting problem of signal cutting. The window is moved along the signal and the spectrum is calculated for each position. This process is then repeated many times for each new cycle with a slightly shorter (or longer) window. The result will ultimately be a collection of the signal's time-frequency representations, all with different resolutions. We can talk about multiresolution analysis because of this collection of representations. By breaking down a time series into time-frequency-space, one is able to determine both the dominant modes of variability and how those modes change with time. Wavelets are demonstrated to be a powerful tool for the analysis and synthesis of data from long memory processes. Wavelets are strongly connected to such processes where the same shapes repeat at different orders of magnitude. The ability of the wavelets to concurrently localize a process in time and scale domain results depicting many dense matrices in a sparse form.

Two categories of wavelet transform are:

- discrete wavelet transforms and
- continuous wavelet transforms

The output of discrete wavelet transforms (DWT) on a given set of discrete signal provides the corresponding approximation coefficients and detail coefficients of the input data vector.

- ***Artificial neural network model***

The two coefficients obtained from wavelet decomposition are approximate coefficients and detail coefficients which are applied to two different neural network models for prediction.
A second single hidden layer of time lag recurrent neural network with Gamma memory and momentum back propagation is used for predicting detailed coefficients.
This neural is fed as input to the normalized detail coefficients obtained from the decomposition. The model is presented schematically in figure 2.

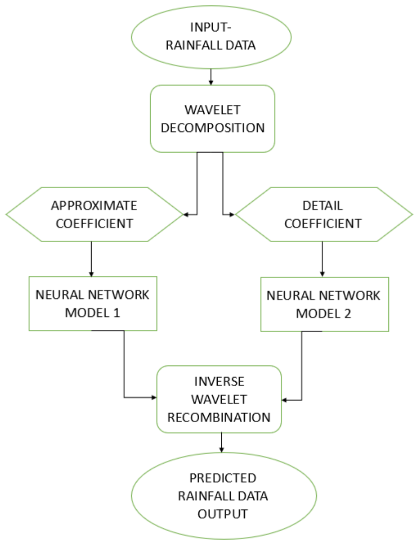

Fig 2: Neural Network Model

E. Back propagation algorithm using multi-layered neural network

The objective of this paper istodevelopanANNmodel using soft computing techniquetoforecast averagemonthlyrainfall o f BangalorecityinKarnataka state in India. The experimental set-up included the following consecutive steps:

- Choosingtheoutputandtheinputdatasetfor the supervised Back propagationtraining [24] of 100-year rainfall data taken from meteorological website.
- Input andthe output data is normalized
- Back Propagation algorithm is used to train the normalized data.
- Testingthe model fitness.
- Comparison ofthe predicted outputwith thedesired output.

→Present studyexploresthedataof allthemonthsfrom1901to2000.1200samplesare present intheinputandtheoutputfiles.

→The output and the input data obtained has been normalized. Initially the average of the rainfall dataset for each year were taken discretely.
Let the mean for each year be,

$$M = \frac{sum\ of all\ samples}{number\ of\ samples}$$

Thestandarddeviation SD,ofeach year werecalculated separately.

$$SD = \sqrt{\frac{\sum(x - M)^2}{N}}$$

Once the mean and SD are computed the dataset is normalized.

$$Normalized\ data = \frac{(x - M)}{SD}$$

→In order to train their normalized data,back propagation algorithm has been used. The algorithm takes 70% of the input data for training (i.e., 840 samples outof1200 samples). For each trial of training the data samples, the algorithm processchooses the training setirregularly from the entire dataset .Therefore, whenever we train the data, we obtaindissimilar values of mean square error (MSE) which is controlled by the 70% of the input dataset is selected for learning. Out of the left over 30% dataset, 15% (i.e., 180 samples) is utilized for validation and the other 15% (i.e., 180 samples) for testing.

→ Aplot has been generated (figure 3) between the targetand forecasted pointsand the high correlation between them clearly shows the ANN model is found to be precise in prediction of rainfall.

The plot after training the network has been shown below.

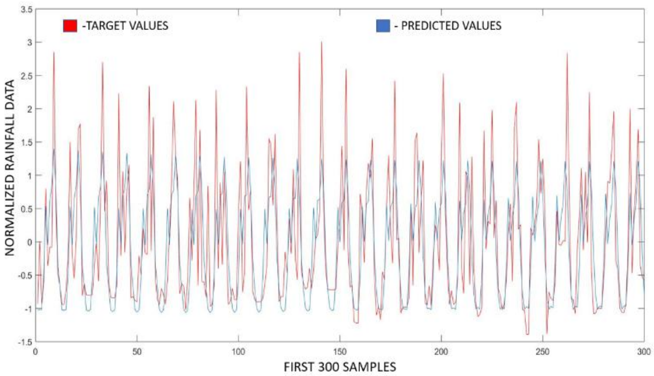

Fig. 3- Comparison between the target values and predicted values.

Table 1represents the observed RMSE and R values for Monsoon season (June-Sept) with one hidden layer having 40 neurons. The network had been trained for the normalized rainfall data.

Table 1: Observed RMSE and R values for Monsoon

	OBSERVED VALUE	TRAINING			
		JUNE	JULY	AUG	SEPT
1ST TRIAL	RMSE	0.36	0.61	0.67	0.1
	R	0.497	0.45	0.42	0.43
2ND TRIAL	RMSE	0.3	0.48	0.7	0.16
	R	0.4	0.42	0.4	0.41

In this work mainly 2 tools are used for executingthealgorithms in MATLAB.

- nntool (Neural Network tool) – allows different algorithms to be performed with multiple hidden layers.
- nftool (Neural network fitting tool) – Allows solely back propagation algorithm to be executed with single hidden layer.

III. RESULTS AND DISCUSSIONS

A. Conventional Regression model and Artificial Neural Network

Table1 represents the validation and training performance for the months July, August, September of Monsoon period of conventional regression model. It can be observed that thebestcombinationofperformancemetricsforthe monthofJulyisexhibitedbymodel4V [10-12].
Thetwo methods can be utilized to predict runoff from the rainfall data for a specific area of study. Although,simpletoformulate,theregressionmodelsdo notproducetherequiredpredictabilityandreliabilitywh enthe interlinkageamongvariablesareeitherunspecifiesordif ficultto construct.Furthermore,foreverymonthdissimilarrainf all run-off regressionequationsarein

demandtobedeveloped,making theirformationcomplicated.Onthecontrary,itisexhibited thatasingleneuralnetworkmodelisskilledtopreciselyacquire thetenuousRainfallRun-offinterplayforallcases,in contrast todiscreteregressionequationsschemedforevery case. The modelevolvedapplyingANNislesscomplexto be devisedandprovidesrun-offincloserelationshipwithdefinitevalues.

Acomparisonoftheapproaches has beenanalysedinthe researchwhich signifiesthattheneuralnetworkshas the power toproduce betterforecastprecision.

B. Linear regression and K-fold

This model accuracy has been validated by calculating Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Explained variance [14-16]. K-fold cross validation technique had been applied for each season and it had been observed that, the errors had decreased and variance score had increased.

Monsoon Season:

The RMSE value was found to be 15.8 and MAE equal to 12.4 for monsoon.

Summer Season:

The RMSE value was found to be 6.44 and MAE equal to 5.04 for Summer.

Winter season:

The RMSE value was found to be 10.10 and MAE equal to 5.01 for Winter.

It is observed that the ensemble technique provides better results as compared to fixed sampling.
Their future work includes validating and comparing the results from other regions apart from Bangalore.

C. K-Nearest Neighbour

The experimental results indicate the forecasting performance of Knn– ANN model has accurate results in terms of forecasting accuracy indices.

Since the Knn–ANN model has better nonlinear mapping capabilities it easily capture rainfall data patterns than the ANN models [17- 20].

Using the nonparametric ensemble model, empirical results obtained better results after the prediction.

Non parametric Knn estimation method has the potential to remove inapposite variables and provides a powerful statistical tool for investigating the underlying structure in a rainfall data set with a vector of smoothing parameters.

D. Wavelet Decomposition

Some studies by Nizar Ali and Dudul shows that the prediction of the rainfall for a given month based on the rainfall data for the previous five months using wavelet decomposition method is possible. 5-8-1 is the network architecture for predicting approximate coefficient and 0.4 is the memory coefficient with gamma and number of taps are equal to 3 [21-23].

A committee of two different ANN configurations (approximate coefficient and detail coefficient) is initiated for authentic rainfall prediction. The accuracy of the prediction for the suggested model is reasonable as per the results.
Data without wavelet decomposition consists of wide range of frequency components hence if rainfall data is directly used, then it is very difficult to design a neural network model with decent accuracy.

Thereby to obtain more accuracy wavelet decomposition is used to separate high and low frequency components will help in designing a neural network model.

E. Back propagation algorithm using multi-layered neural network

The following conclusions are made in case of the back-propagation algorithmusing multi-layered neural network:

- Different networks give different results with same training functions and adaptive learning functions having same number of neurons.
- BPA gives a correlation factor of 0.745.
- Increase in number of neurons shows decrease in MSE.
- Large data set gives less RMSE value.
- LEARNGDM is the optimaltrainingfunction tolearn the databut at the same time it is time consuming.
- TRAINLM is the besttrainingfunction.
- Multi layered network is better than single layered network in terms of performance.
- Both NNTOOLS and NFTOOLS provides predicted values of the same range. Figure 4 presents the schematic of neural network used in this study.

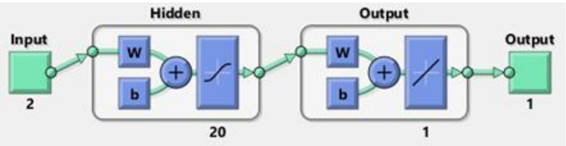

Fig 4: Neural Network Diagram

The RMSE and correlation value observed values after training the network using BP algorithm are 0.446 and 0.74 respectively.

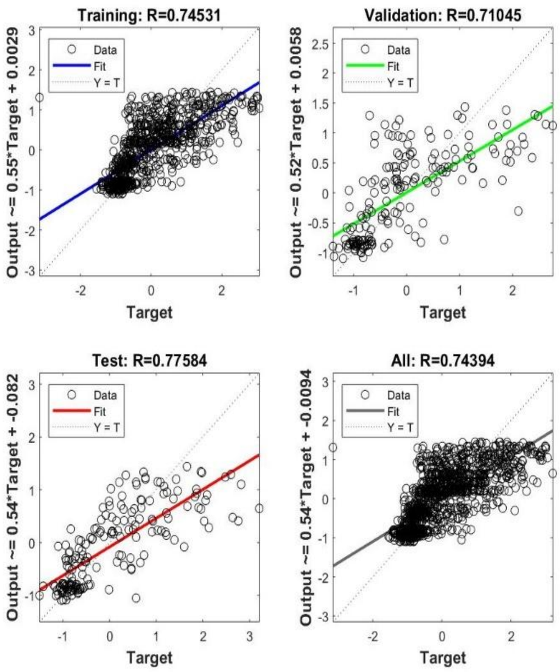

Fig 5: Regression lines for training, validation and testing respectively.

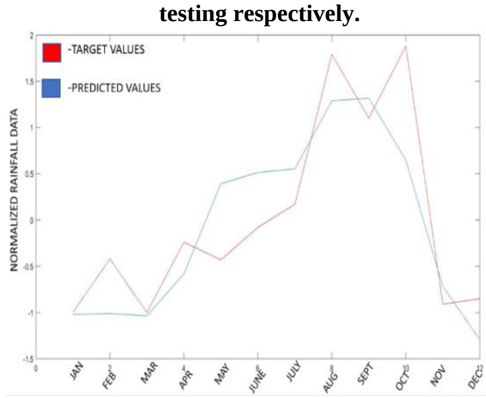

Fig 6: Comparison of target values and predictedvalues of monthly rainfall for the year 2000

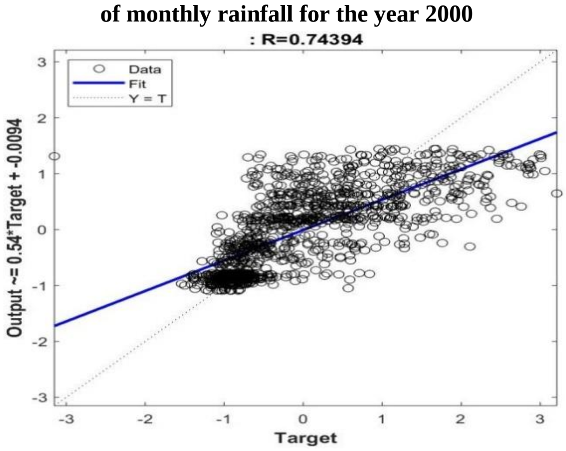

Fig 7: Regression line between target and output data

The regression lines for training, validation and testing are presented in figure 5 which shows the correlation is very strong in all the cases and also the comparison of target and predicted rainfall at monthly scale for the year 2000 (figure 6) shows high predictability capability of the neyuaral network model configuration.

IV. CONCLUSION

This paper outlines a comprehensive survey on rainfall predictions using five different neural network architectures. It was found from the survey that most of the researchers used back propagation algorithm and had substantial results.
Therefore, we have trained a neural network using back propagation algorithm for Bangalore city in Karnataka.

The output of the neural network trained using back propagation algorithm gives less error when compared to other methods.
Besides, the survey concludes about the forecasting techniques that use Linear regression, K-fold, K-Nearest Neighbour, and Wavelet Decomposition are suitable to predict rainfall. However, some limitations of those methods have been found. Various developments of ANN research given in this paper would be of great help to ANN researchers to accurately predict rainfall in the mere future.

REFERENCES

[1] “Weather Patterns are influenced by the Butterfly Effect”http://www.unexplainable.net/artman/publish/article_1037.sht ml.

[2] Gwangseob, K. and Ana, P. B.: Quantitative flood forecasting using multisensor data and neural networks, Journal of Hydrology, 246, 45– 62, 2001

[3] D.R. Nayak, A. Mahapatra and P. Mishra, “A survey on rainfall prediction using artificial neural network,” Int J Computer Appl., vol. 72, no. 16, 2013.

[4] S. Haykin, 1999:Neural Networks-A Comprehensive Foundation.Addison Wesley Longman.

[5] McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas imminent in nervous activity. Bulletin and Mathematical Biophysics 5, 115–133

[6] Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986a. Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing. MIT Press, Cambridge.

[7] Maier, H.R., Dandy, G.C. 1999b. Neural network-based modelling of environmental variables: A systematic approach. Mathematical and Computer Modelling, in press.

[8] Han and M. Kamber, “Data mining: concepts and techniques,” Edition: 2, Morgan Kaufmann, 2006.

[9] Meteorological data source for Bangalore (online), http://www.indiawaterportal.org/met_data/

[10] Sunil Kumar Vyas, Gunwant Sharma, Yogesh Prakash Mathur, Vinay Chandwani , “Rainfall- run off modelling: Conventional Regression an Artificial Neuron Networks Approach”; IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2016), December 23-25, 2016, Jaipur, India.

[11] M.P. Rajurkar, U.C. Kothyari, U.C. Chaube, "Artificial neural networks for daily rainfall-runoff modelling", *Hydrological Sciences Journal*, vol. 47, no. 6, pp. 865-877, 2002.

[12] S. Riad, J. Mania, L. Bouchaou, Y. Najjar, "Rainfall Runoff Model using an Artificial Neural Network Approach", *Mathematical and Computer Modelling*, vol. 40, pp. 839-846, 2004.

[13] Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar, "A Rainfall Prediction Model using Artificial Neural Network"; 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC 2012)

[14] Sandeep Kumar Mohapatra, Anamika Upadhyay, Channabasava Gola, "Rainfall Prediction based on 100 years of Meteorological Data".

a. D'Souza, E.C. Barrett, C.H. Power, "Satellite rainfall estimation techniques using visible and infrared imagery", *Remote Sensing Reviews*, vol. 4, no. 2, pp. 379-414, 1990.

[15] S. Parthasarathy, C.C. Aggarwal, "On the Use of Conceptual Reconstruction for Mining Massively Incomplete Data Sets", *IEEE transactions on Knowledge and Data Engineering*, vol. 15, no. 6, Nov/Dec 2003.

[16] Jiansheng Wu, "A Novel Artificial Neural Network Ensemble Model Based on K–Nearest

[17] Neighbour Nonparametric Estimation of Regression Function and Its Application for Rainfall Forecasting"; Department of Mathematics and Computer Liuzhou Teachers College, Guangxi, Liuzhou, China; Wjsh2002168@163.com.

[18] Fan. J, Truong. Y. Nonparametric regression with errors in variables Annals of Statistics, 21: 1900-1925, 1993.

[19] Hall.P, Q. Li and J. Racine. Nonparametric Estimation of Regression Functions in the Presence of Irrelevant Variables forthcoming in Review of Economics and Statistics, 89: 784-789, 2007.

[20] G.F.Lin, L.H.Chen. A nonlinear rainfall-runoff model using radial function network. Journal of Hydrology, 289:1-8, 2004.

[21] Nizar Ali Charaniya, Lecturer Department of Electronics, B.N.College of Engg, Pusad, India; and Dr.Sanjay.V.Dudul Prof and Head Applied Electronics Department S.G.B. Amravati University Amravati, India ; "Committee of Artificial Neural Networks for Monthly Rainfall Prediction using Wavelet transform".

[22] Kumar K Krishna, MK Soman, Kumar K Rupa, "Seasonal forecasting of Indian summer monsoon rainfall", *Weather*, vol. 50, pp. 449-467, 1995.

[23] M Rajeevan, P Guhathakurta, V Thapliyal, "New models for long range forecasts of summer monsoon rainfall over Northwest and Peninsular India", *Meteorol Atmos. Phys.*, vol. 73, pp. 211-225, 2000.

[24] Ch.Jyosthna Devi, B.Syam Prasad Reddy, K.Vagdhan Kumar, B.Musala Reddy, .Raja Nayak, "International Journal of Engineering Trends and Technology", Volume3 Issue1, pp. 19-23, 2012.