

Effective Features and Hybrid Classifier for Rainfall Prediction

KavithaRani B

*Associate Professor, Jyothishmathi Institute of Technology & Science,
Karimnagar, Andhra Pradesh, India
kavitharani0678@gmail.com*

A. Govardhan

*Professor in CSE, & Director, School of IT,
JNTUH, Hyderabad*

Received 29 April 2013; Accepted 14 June 2014

Abstract

Rainfall prediction has emerged as a challenging time-series prediction problem in recent years. In this paper, we propose a novel rainfall prediction technique using effective feature indicators and a hybrid technique. Our proposed model consists of three phases, namely, layer model simulation, training phase and testing phase. At the outset, the input rainfall dataset is preprocessed using the feature indicators. There are five feature indicators used in the preprocessing step namely, channel index (CI), ulcer index (UI), rate of change (ROC), relative strength index (RSI) and average directional movement index (ADX). Subsequently, feature matrices are formed based on the preprocessed rainfall data. Once the feature matrix is formed, the prediction is done based on the hybrid classifier. In the hybrid classifier, artificial bee colony algorithm is combined with the genetic algorithm for training the feed forward neural network. The performance of the algorithm is analyzed with the help of real datasets gathered from Rayalaseema, Aandhra and Telangana regions. Finally, from comparative analysis it is established that the proposed rainfall prediction yields better result (MAC=4.0672) when compared with Artificial Bee Colony with Neural Network.

Keywords: *rainfall prediction, hybrid classifier, feature indicator, ABC, genetic, FFNN*

1. Introduction with Challenges

A constantly altering climate conditions are observed in the present day world. Particularly in the agricultural segment of a country [1] climate changes lead to far reaching results. Rainfall forecast is very important for agriculture division and suitable for diverse features of climate change. In the third world countries like India, the whole agriculture depends upon rain and as a matter of rain is one of the amazing gifts of nature. It is normally believed that rainfall is changeable. Real-time water resources estimation is described as a quick estimation of the water resources produced in a rainfall incident or in a past era from an exact day of the year to the existing rainfall event [2]. The recognition of the situation of rainfall in advance can help in supervising and dealing with agricultural administration and failure prevention [1]. For food production plan, water resource management, and all action plans in nature, information regarding rainfall is significant. The incidence of

extended dry period or heavy rain at the decisive stages of the crop growth and enlargement could lead to significant decrease in crop yield [3].

The presentations of both stochastic and deterministic rainfall forecast models [4] are affected as rainfall is one of the most complicated components of the hydrological cycle to forecast, and is gigantically doubtful. In every place rainfall is not a usual event. It has a few seasonality results. Hence, the rainfall forecast problem is not similar to other usual atmospheric parameters like temperature, humidity, etc. Rainfall is furthermore a time series record like atmospheric pressure, temperature, vapor pressure, relative humidity, radiation, etc [5]. At local and national levels an extensive array of rainfall forecast methods are used in weather forecasting. There are basically two approaches to forecast rainfall, such as the empirical and dynamical techniques. Over various parts of the world the empirical approach is based on the study of historical records of the rainfall and its correlation to a range of

atmospheric and oceanic variables. Based on methods of equations in a dynamic approach, forecasts are produced by corporeal models that forecast the growth of the global climate method in reaction to initial atmospheric conditions [6] [7].

Particularly in rainfall prediction there has been more proof in literatures that Artificial Neural Networks (ANNs) are gradually and extensively used for hydrological modeling with numerous rainfall predicting techniques based on stochastic or deterministic techniques and computational approaches [8-13]. Quite a lot of reasons are there regarding why ANNs are important and suitable for use in such predicting systems. Devoid of the requirement for making priori assumptions which are understood in most classic statistical approaches, they are, first and foremost, data-driven methods which have the capacity to model both linear and non-linear systems. Secondly, they are proficient of simplification. Given to them during the training after educating the records, they can frequently correct the estimate of the hidden part of a population which is not a part of the teaching data [14]. Back propagation-based training algorithm has developed zooming interest in weather forecasting [10] among dissimilar types of ANNs. However, back propagation algorithm is not much successful in teaching in respect of the actual time problems. Accordingly, the successful designing of teaching algorithm can show the way for further improved results in ANN models.

For several decades, the data classification methods for depicting weather events related prediction models have gained greater achievements, but with the more multisource and multi-dimensional data, the conventional prediction methods based on statistical theory have exposed certain limitations. So, in recent years the artificial intelligence techniques have been applied to weather related predictions, and the main purpose is to reduce the undesired limitations of the conventional prediction methods. Artificial Bee Colony (ABC) [15] and Genetic Algorithm (GA) [16 and 17] as the two main algorithms in swarm intelligence have been applied successfully to optimization problems in many research areas. Here, GA and ABC are combined and used for neural network training. The hybridization of these two popular algorithms to select better connections to the network and the performance of the

optimized network are to be measured in predicting future rainfall in a historical dataset.

In this paper, we have developed a novel rainfall prediction technique using effective feature indicators and a hybrid technique. Here, the proposed rainfall prediction consists of three phases, namely, layer model simulation, training phase and testing phase. Firstly, the input rainfall dataset is preprocessed to make data suitable for extracting the feature indicators. Then, feature indicators are used to extract feature components from the preprocessed input data. There are five feature indicators used in the preprocessing step. Subsequently, feature matrices are formed based on the preprocessed rainfall data. Once the feature matrix is formed, the prediction is done based on the hybrid classifier. In the hybrid classifier, artificial bee colony algorithm is combined with genetic for training the feed forward neural network. The trained neural network by the hybrid ABC and genetic algorithm is used to predict the rainfall.

The organization of this paper is as follows: a brief review of some of the literature works in rainfall prediction models are presented in Section 2. Contribution of the paper is discussed in section 3. The proposed rainfall prediction model is detailed in Section 4. The experimental results and performance evaluation discussion are provided in Section 5. Finally, the conclusions are summed up in Section 6.

2. Literature Review

By means of neural network, Literature offers more than a few techniques for rainfall prediction. At this point, we analyze a few of the works offered. A model based on artificial neural networks (ANNs) and wavelet decomposition as a teaching tool to forecast successive daily rainfalls on reports of the previous events of rainfall data has been extended by Wassamon Phusakulkajorn *et al.* [13]. Wavelet coefficients are of two sets, for which one model signifies the detailed information of rainfall records and the other operates as a smoothing filter, which are removed for the ANNs. Here, for the education and knowledge eradication processes a back-propagation neural network is applied. Tha Sae district in Chumphon province, Kanchanadit district in Surat Thani province, Muang district in Nakhon Si Thammarat province, Muang district in Phatthalung province and Hatyai district in Songkhla province are from five locations in the south of Thailand

which have been experimented on rainfall records for methodology.

Using dissimilar datasets of yearly, biannually, quarterly and monthly-sampled values, Kyaw Kyaw Htike and Othman O. Khalifa [14] have planned and applied rainfall prediction models. By means of Focused Time-Delay Neural Networks (FTDNN) the major involvement of their research is in the plan, completion and relationship of rainfall predicting models. The most favorable parameters of the neural network architectures are acquired from tests whereas networks have been educated to execute one-step-ahead forecasts. The everyday rainfall dataset, attained from Malaysia Meteorological Department (MMD), is changed to monthly, biannually, quarterly and monthly datasets. Teaching and experimenting are executed on every of the datasets and consequent accuracies of the predictions are calculated by means of Mean Absolute Percent Error.

Predictive models in order to predict rain force (mm/day) in Athens, Greece, by Artificial Neural Networks (ANN) models have been erected by P.T. Nastos *et al.* [18]. The ANNs results worry the schemed mean, maximum and minimum monthly rain force for the next four successive months in Athens. For a 111-year period (1899–2009), the meteorological information applied to guess the rain intensity, are the monthly rain totals (mm) and the relevant rain days, which have been obtained from the National Observatory of Athens. A moderately consistent prediction of the rain force for the next four months is shown by the effects of the developed and pertained ANN models. Suitable statistical indices are taken into consideration for the assessment of the effects and the capability of the developed prognostic models. A group of Artificial Neural Networks (ANNs) based model by wavelet decomposition for forecast of monthly rainfall on accounts of the previous occasions of rainfall data has been offered by Nizar Ali Charaniya and Dr. Sanjay.V.Dudul [19]. For extortion of fairly accurate and detailed coefficient of the rainfall information series, wavelet transform is applied. All along with an ANN for teaching and knowledge extortion processes, those coefficients are utilized. For dissimilar geographical area of India and moreover for the whole country the replica has been experimented on rainfall information. The offered replica is competent of predicting monthly rainfall one month in progress.

Using Wavelet Support Vector Machine Regression (W-SVR) for rainfall prediction a hybrid Radial Basis Function Neural Network (RBF-NN) ensemble model has been developed by Lingzhi Wang and Jiansheng Wu [20]. The primary data set is separated into different training sets by using Bagging and Boosting technology at the initial stage in the course of assembly modeling. Based on variety principle, in the next stage, those teaching sets are keyed in to the dissimilar individual RBF-NN models, and subsequently different single RBF-NN forecasters are generated. To choose the suitable number of neural network assembly members the Partial Least Square (PLS) technology is employed in the third stage. In the finishing stage, W-SVR is utilized for collection of the RBF-NN for the calculation purposes. For experimentation purposes, this learning compares the presentation of the collection model with a few surviving neural network ensemble approaches in terms of monthly rainfall prediction on Guangxi, China. The calculations by means of the offered approach have every time been better than those gained by the other methods offered in this study in terms of the similar measurements revealed by experimental results.

A Prediction Model by means of Artificial Neural network has been applied by Kumar Abhishek *et al.* [21]. For the most excellent presentation, this document executes one of those functions by building teaching and experimenting data sets and discovering the number of concealed neurons in these layers. Through non-natural neural network models they have examined the opportunity of forecasting average rainfall over Udupi district of Karnataka. Three encrusted networks have been erected by preparing non-natural neural network based analytical models. The replicas under study have dissimilar in the number of unseen neurons.

Using radar reflectivity and TB data, Andrew Kusiak *et al.* [22] have built a rainfall forecast replica in a watershed basin. TB data has been accumulated at the three sites of Oxford, South Amana, and Iowa City and the radar reflectivity data has been acquired from the NEXRAD system. The information is processed to have similar occurrence and takes away outliers and negative values. The MLP has performed as the best among the five data-mining algorithms tested in this paper. For three models for all outlook time horizons, it has been chosen to forecast rainfall. With radar reflectivity data merely the baseline Model I have been erected. By means of radar and local TB data the offered Model II

has been made. With radar and close by TB data Model III has been made. The working out effects have pointed out that the three models have a related presentation in forecasting rainfall at present time, and the offered Model II accomplishes more than the other models in forecasting rainfall at expected time horizons.

Additionally, Wang *et al.* [23] have compared several artificial intelligence models, namely, Auto-Regressive Moving Average (ARMA), ANN, Adaptive Neural-Fuzzy Inference System (ANFIS), Genetic Programming (GP) and Support Vector Machine (SVM) to predict monthly discharge time series. The results indicate that the best performance is obtained by ANFIS, GP and SVM. Wu *et al.* [24] have compared performance of data driven model to forecast monthly stream flow. The results have shown that ARMA and K-Nearest-Neighbors (KNN) have performed prediction better than ANN and its variants when the correlation between input and output is low. Lohani [25] has compared ANN, FIS and linear transfer model for daily rainfall-runoff model under different input domains. The results illustrate that FIS has outperformed linear model and ANN. Nayak *et al.* [26] and Kermani *et al.* [27] have introduced ANFIS model to river flow time series. Jain and Kumar [28] have applied conventional preprocessing approaches to ANN for stream flow time series data. According to above discussions, the primary aim of this study is to develop an efficient model for rainfall time series prediction problem.

3. Contribution of the Research

The main contributions of proposed rainfall prediction technique are:

- ❖ Input rainfall dataset is firstly converted into suitable format for training purpose. Here, five feature indicators are taken such as, Channel index (CI), Ulcer index (UI), Rate of change (ROC), Relative Strength Index (RSI) and Average direction movement index (ADX). Based on these feature indicators, rainfall dataset that is required for analysis is prepared.
- ❖ Prediction of rainfall is done with the help of hybrid classifier. In the hybrid classifier, artificial bee colony algorithm is combined with genetic algorithm for training the feed forward neural network.
- ❖ Evaluation matrices parameters such as Mean Absolute Deviation, Root mean square error and Mean Absolute Percentage Error are used to predict the rainfall.

4. The Proposed Rainfall Predictive Model

The target of this research is to create a rainfall prediction model that can be used as a decision support tool to predict the future behavior by using five technical indicators. There are many technical indicators and theories in prediction analysis. Generally, the technical indicators are used in stock market models to predict the future stock market price. Most of the indicators are worked based on the time-series analysis. According to this, we assume that the technical indicators are suitable choices for time-series data analysis which are also the roots for our rainfall prediction model. This is the reason that we bring those indicators into our rainfall prediction model.

The overall block diagram of proposed rainfall prediction model is depicted in figure 1.

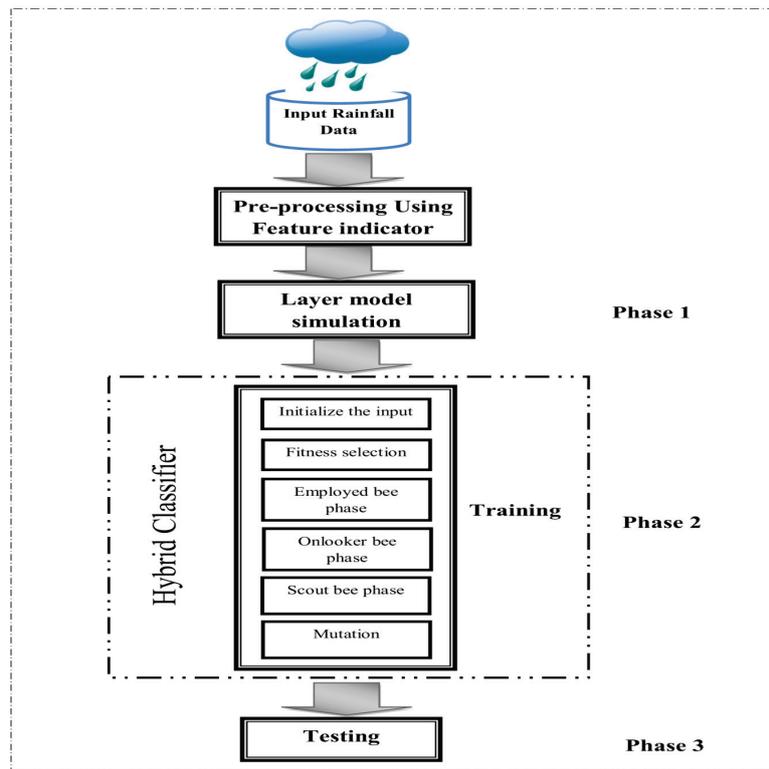


Fig.1. Overall block diagram of proposed rainfall prediction technique

As illustrated in figure 1, the proposed rainfall prediction model consists of the following phases, (1) Layer model simulation (2) Training phase and (3) Testing phase. The training phase includes the following two major steps

Step 1: For rainfall set RDS, it is firstly converted for suitable format to data preprocessing, i.e. preprocessing is performed based on the five feature indicators. Thereafter, the preprocessed dataset is initially divided into training set (TR) and testing set (TS).

Step 2: The training dataset is fully trained using hybrid algorithm and preprocessed data to predict the future rainfall easily and efficiently in testing time.

The two stages of proposed rainfall predictive model raise four important issues: (1) how to create preprocessed rainfall Data based on feature Indicators (2) how to classify the rainfall data using hybrid classifier with FFNN in training phase. These issues will be addressed by the following sections respectively.

4.1 Preprocessing using Feature Indicators

Preprocessing the rainfall data, the former issue of the proposed approach, is performed to make the data adaptable for further analysis. Our preprocessed data consists of five indicators [29 and 30] along with corresponding class values. Here, class value is defined for every year, which is prepared based on the next year average of rainfall value. For example in figure 2, class value of year 1884 is defined as the average value of year 1885. Figure 2 shows the example of rainfall dataset for 10 years and figure 3 shows the preprocessed dataset using five technical indicators. Here, the preprocessing is performed based on the four feature indicators like as Channel index (CI), Ulcer index (UI), Rate of change (ROC), Relative strength index (RSI) and Average directional movement index (ADX). For all the experimentation with the rainfall data, the window size is constantly fixed as 7 which are decided based on the trial and error method. For the selection of best window size, the window size is changed from 5 to 10 and the output is computed. Based on the performance, the window is fixed as 7 for all the experimentation. Feature indicator description is detailed in the section below.

4.1.1 Description of Feature Indicators

1. Channel Index (CI): Channel Index measures the current rainfall level relative to an average rainfall level over a given period of time. The number of CI periods is also used for the calculations of the simple moving average and Mean Deviation.

$$CI = \frac{(\text{Typical Rain fall value} - 7 \text{ period SMA of Typical Rain fall value})}{(0.015 \times \text{Mean Deviation})} \quad (1)$$

To find typical rainfall value

$$\text{Typical rainfall value} = \frac{(\text{High Rain fall} + \text{Low Rain fall} + \text{Closing Rain fall})}{3} \quad (2)$$

Where,

High rainfall is the rainfall which is the maximum value at particular period.

Low rainfall is the rainfall which is the minimum value at particular period.

Closing rainfall is the rainfall value at the end of the period. Closing rainfall plots provide a good perspective on trending direction of the future rainfall level.

To find Simple moving average (SMA)

Calculation of Simple moving average (SMA):

For example, yearly closing rainfall value: 11,12,13,14,15,16,17,18,19,20,21,22,23

First year of 7-year Simple moving average (SMA): $(11 + 12 + 13 + 14 + 15 + 16 + 17) / 7 = 14$

Second year of 7-year Simple moving average (SMA): $(12 + 13 + 14 + 15 + 16 + 17 + 18) / 7 = 15$

Seventh year of 7-year Simple moving average (SMA): $(17 + 18 + 19 + 20 + 21 + 22 + 23) / 7 = 20$

Constant = 0.015

To find Mean deviation

There are four steps for calculating the Mean Deviation:

- (i) First, subtract the most recent 7-period average of the typical rainfall from typical rainfall value of each period.
- (ii) Second, take the absolute values of these numbers.
- (iii) Third, sum the absolute values.
- (iv) Fourth, divide it by the total number of periods (7)

2. Ulcer Index (UI): Using a default setting of 7 periods, the Ulcer Index reflects the expected percentage drawdown over this period.

$$R_i = \frac{(\text{closing rainfall})_i - \text{Max closing rainfall}}{\text{Max closing rainfall}} \quad (3)$$

$$\text{Ulcer Index} = \sqrt{\frac{\sum_{i=1}^N R_i^2}{N}} \quad (4)$$

Where,

$N \rightarrow$ is the size of the time period or window (here, $N=7$)

$(\text{closing rainfall})_i \rightarrow$ Closing rainfall in a particular period

$R_i \rightarrow$ is the drawdown from the previous high, here, drawdown explains that the peak-to-trough during a specific period of a rainfall data.

3. Rate of Change (ROC): The Rate of Change (ROC) indicator [30] is simple, and it measures the percent change in rainfall from one period to the next. It is defined as the rainfall change in particular fixed time duration. The ROC calculation compares the current rainfall with the rainfall "n" periods ago.

$$\text{Rate of change} = \frac{(\text{Rainfall in last month of window} - \text{last 'n' periods ago})}{\text{last 'n' periods ago}} \times 100 \quad (5)$$

Where, $n=7$

4. Relative strength index (RSI): Relative strength index [29] basically consists of three components, RS, average positive grow rainfall and average negative rainfall.

$$\text{Relative strength index} = 100 - \frac{100}{1 + RS} \quad (6)$$

Where,

❖ Calculation of RS:

$$RS = \frac{\text{Average positive growrain fall}}{\text{Average negative growrain fall}} \quad (7)$$

The very first calculations for average positive grow rainfall and average negative grow rainfall are simple 7 period averages.

$$\text{First average positive grow rainfall} = \frac{\text{Sum of positive grow rainfall over the past 7 periods}}{7} \quad (8)$$

$$\text{First average negative grow rainfall} = \frac{\text{Sum of negative grow rainfall over the past 7 periods}}{7} \quad (9)$$

The second, and subsequent, calculations are based on the prior averages and the current positive negative rainfall grow.

$$\text{Average positive grow rainfall} = \frac{(\text{Previous Average positive growrain fall}) \times 6 + \text{current positive rain flow}}{7} \quad (10)$$

$$\text{Average negative grow rainfall} = \frac{(\text{Previous Averagenegative growrain fall}) \times 6 + \text{current negative rain flow}}{7} \quad (11)$$

5. Average directional movement index (ADX)

The ADX indicator is a combination of positive directional indicator or +DI and negative directional indicator or -DI. Both these are combined and the result is smoothed by an exponential moving average.

Initially, the absolute up movement or Plus Directional Movement (+DM) or absolute down movement or Minus Directional Movement (-DM) is determined.

If the rainfall value has moved up, the distance between the present high and previous high is taken as +DM. Similarly, if the rainfall value has moved down, the distance between the present low and previous low is taken as -DM.

Here please note that, “+” and “-” denote only upward and downward movements respectively. “-DM” is not a negative number. But, it is a positive number of a downward movement.

Then Directional Index is calculated by dividing Directional Movement by True Range.

$$\text{True range} = \max[\text{high} - \text{low}, \text{abs}(\text{high} - \text{close}_{\text{prev}}), \text{abs}(\text{low} - \text{close}_{\text{prev}})] \quad (12)$$

That is, Directional Indicator of day 1 = Directional Movement of day 1 / True Range of day 1

$$+DI_1 = (+DM_1 / TR_1) \times 100$$

$$-DI_1 = (-DM_1 / TR_1) \times 100$$

In this way DI is calculated for 7 years. On 7th year the formula will be:

$$+DI_7 = (+DM_7 / TR_7) \times 100$$

$$-DI_7 = (-DM_7 / TR_7) \times 100$$

4.1.2 Sample and Preprocessed Rainfall Dataset

In this section, in Figure 2, we have specified an example rainfall dataset which represents rainfall data for 10 years with class. Figure 3 represents the Converted sample rainfall dataset by technical indicators.

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	normalized next year average rainfall (Class)
1884	0.05	0.04	0.03	0.04	0.07	0.14	0.27	0.34	0.40	0.49	0.28	0.12	0.19
1885	0.03	0.03	0.08	0.03	0.11	0.23	0.27	0.15	0.39	0.31	0.39	0.25	0.27
1886	0.03	0.03	0.06	0.03	0.18	0.29	0.37	0.59	0.27	0.93	0.27	0.17	0.20
1887	0.03	0.03	0.03	0.03	0.10	0.31	0.25	0.34	0.23	0.61	0.44	0.06	0.15
1888	0.04	0.03	0.03	0.04	0.20	0.12	0.27	0.18	0.17	0.22	0.49	0.04	0.20
1889	0.03	0.03	0.03	0.04	0.06	0.30	0.29	0.27	0.52	0.63	0.15	0.09	0.16
1890	0.03	0.03	0.03	0.04	0.07	0.29	0.24	0.21	0.26	0.29	0.33	0.06	0.12
1891	0.03	0.06	0.05	0.03	0.10	0.12	0.23	0.25	0.27	0.15	0.05	0.06	0.23
1892	0.02	0.03	0.02	0.03	0.03	0.24	0.27	0.58	0.47	0.98	0.07	0.07	0.26
1893	0.03	0.06	0.10	0.08	0.07	0.27	0.39	0.22	0.45	0.45	0.90	0.03	0.21

Fig. 2. Example rainfall dataset model of proposed rainfall prediction technique

	ROC	RSI	CI	UI	ADI	Class
1884	0.54	51.11	-58.22	75.99	98.27	0.19
1885	-0.50	59.02	-29.10	75.99	98.18	0.27
1886	3.98	52.83	113.84	68.47	98.27	0.20
1887	-0.38	45.87	-32.58	67.71	98.35	0.15
1888	-0.19	44.56	-89.40	66.48	98.40	0.20
1889	-0.46	48.64	-6.80	66.11	98.48	0.16
1890	-0.12	46.03	-131.06	64.27	98.54	0.12
1891	-0.49	45.92	-109.71	64.20	98.59	0.23
1892	-0.73	46.75	103.64	68.13	98.72	0.26
1893	-0.81	42.77	88.72	71.36	98.84	0.21

Fig.3. Converted Sample rainfall dataset by five technical indicators

4.2 GABC Algorithm for Feed Forward Neural Network

After preparing the rainfall data using the discussed preprocessing techniques the feed forward neural network is going to be trained. For this training process, we have developed effective hybrid classifier. In this proposed hybrid classifier, we have used two optimization technique like artificial bee colony and genetic algorithm. The trained neural network by the hybrid artificial bee colony and genetic algorithm is used to predict the rainfall. In this section, we have classified the proposed technique into three phases. The detailed explanations of the three phases are described below.

PHASE 1: LAYER MODEL SIMULATION

Output of the i^{th} neuron is described by following expression:

$$y_i^{(1)} = f \sum_{i=1}^n w_{ij}^{(1)} x_i + \theta_i$$

$$y_i^{(2)} = f \sum_{i=1}^n w_{ij}^{(2)} y_i^{(1)} + \theta_i$$

$$\vdots$$

$$\vdots$$

$$y_i^{(h)} = f \sum_{i=1}^n w_{ij}^{(h)} y_i^{(h-1)}$$

Where, x_i is the input node, y_i is the output of the node x_i , w_{ij} is the connection weight between the node j and input x_i , θ_i is the threshold (or bias) of the node and f is the node transfer function. Usually, the node transfer function is a nonlinear function such as a heaviside function, a sigmoid function, a Gaussian function, etc.

PHASE 2: TRAINING PHASE

Normally, the neurons in the network architecture are interconnected between the layers. The depiction of the feed forward neural network is given in figure 4. The computation starts from input neurons where data input is received and then propagates to hidden neurons and further to output neurons. A neuron in the output layer produces the model output. If the number of neurons and layers are established, the only unknown parameter in the computation is the weights. The process of data training determines the weights. Data training is a process of feeding rainfall data to the input and output of the network model so that the network model simulates the sample data. In this paper, the data training process involves feed-forward and back propagation computation cycles. The back propagation computation is an adjustment of output and hidden neuron weights based on the gradient descent method. These weights are normally initialized with random values to speed up the data training process to a solution.

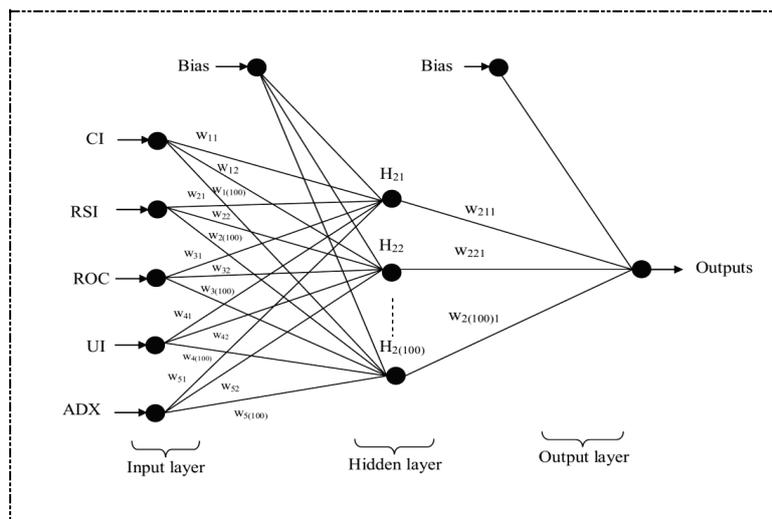


Fig.4. Neural network model of proposed approach

Step 1: Objective function

The process of data training determines the weights.

From this training process, we need w_{ij}^h

\forall $MSE = (\hat{y}_i - \hat{x}_i)$ Should be minimum

$\hat{y}_i \rightarrow$ Target output

$\hat{x}_i \rightarrow$ Predicted output

Initialize the solution: Firstly, we have randomly generated the N_s number of solution

$$W(n) = w_{ij}^{(h)}(1), w_{ij}^{(h)}(2), \dots, w_{ij}^{(h)}(p) \quad (13)$$

Where, p is the population.

Step 2: Fitness function

Fitness function selection is regarded as the error between the anticipated output and real output as individual fitness.

$$\forall \quad MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - \hat{x}_i)^2 \quad (14)$$

Where,

\hat{y}_i is the target output and \hat{x}_i is the predicted output

Step 3: Employed bee phase

The bee $w_{ij}(p)$ in step n search for a new location in the current position vector nearby, search formula is

$$v_{mi} = w_{ij} + \phi_{ij}(w_{ij} - y_{ij}) \quad (15)$$

Where, w_{ij} is a randomly selected food source, i is an input node and ϕ_{ij} is a random number within the range $[-a, a]$. After producing the new food source (w_{ij}), its fitness is calculated and the greedy selection is applied between w_{ij} and (y_{ij}).

Step 4: Onlooker bee phase

In this phase, onlooker bee chooses a food source depending on the probability values calculated using the fitness values provided by employed bees. For this purpose, a fitness based selection technique is used, such as the roulette wheel selection method. The probability value P_i with which w_{ij} is chosen by an onlooker bee can be calculated by using the expression given in the following equation

$$P_i = \frac{fit_i}{\sum_{k=1}^{SN} fit_n} \quad (16)$$

Where, fit_i is the fitness value of solution i which is proportional to the nectar amount of the food source in the solution i and SN is the number of food sources which is equal to the number of employed bees.

Step 5: Scout bee phase

The unemployed bees that choose their food sources randomly are called scouts. The scout discovers a new food source to be replaced with W_{ij} . This operation is defined in equation (1).

Step 6: Mutation

In this step, we have randomly selected particular fixed chromosomes for training purpose from the length of the chromosome.

PHASE 3: TESTING PHASE

In testing phase, select best solution $w_{ij}^{(h)}$ from the training phase, which is given to the input of the feed forward neural network. Subsequently, we obtain the final rainfall value from after completed testing phase. The flowchart of GABC algorithm is depicted in figure 5.

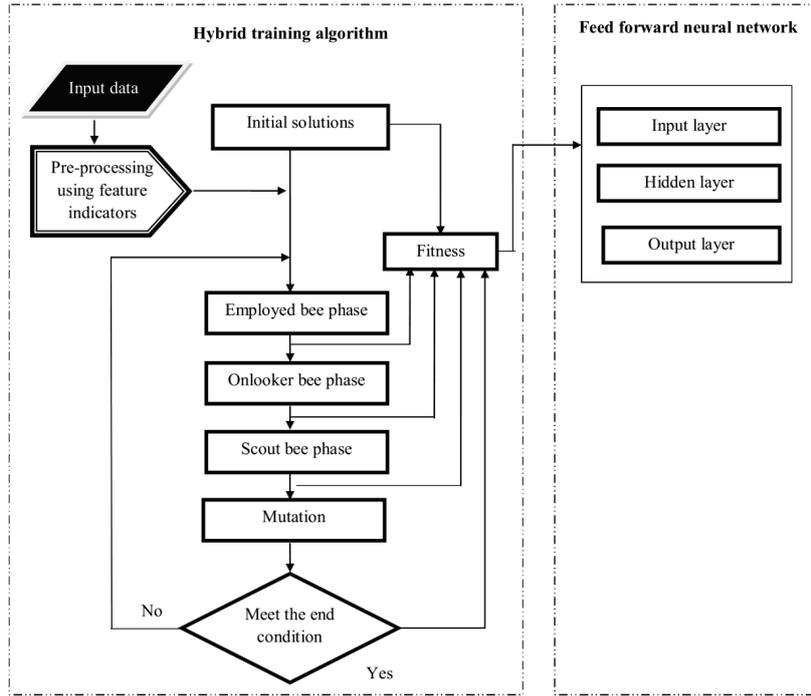


Fig.5. Flowchart of rainfall prediction technique

5. Simulation Results and Discussion

This section explains the experimental results of proposed rainfall prediction technique using real datasets. The proposed rainfall prediction is performed in a windows machine having configurations Intel (R) Core i5 processor, 3.20 GHz, 4 GB RAM, and the operation system platform is Microsoft Wnidow7 Professional. We have used mat lab version (7.12) for this proposed model.

5.1 Dataset Description

The performance of the algorithm is analyzed with the help of real datasets taken from Rayalaseema, Andhra pardesh and Telangana regions. A monthly rainfall data in the period of 1871-2011 from three regions in Rayalaseema, Andhra Pradesh and Telangana are collected from Indian Institute of Tropical Meteorology (IITM) via the <ftp://www.tropmet.res.in/pub/data/rain/iitm-regionrf.txt>. Data is arranged region/subdivision wise, with a header record for each region/subdivision followed by 135 data records each record containing data for one year (12 monthly values).

Preparation of Subdivisional /Regional rainfall series:

The monthly (January-December) area weighted rainfall series for each of the 30 meteorological subdivisions have been prepared by assigning the district area as the weight for each rain-gauge station in that subdivision. Similarly assigning the subdivision area as the weight to each of the subdivisions in the region, area weighted monthly rainfall series are prepared for Homogeneous regions of India.

5.2. Performance Measures

The evaluation of rainfall prediction in different real datasets is carried out using the following metrics as suggested by the following equations,

- ❖ Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_t - d_t)^2} \quad (17)$$

- ❖ Mean Absolute Deviation (MAD)

$$MAD = \frac{1}{N} \sum_{i=1}^N |p_t - d_t| \quad (18)$$

- ❖ Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{p_t - d_t}{p_t} \right| \quad (19)$$

Where,

p_t - Expected value for period t

d_t - Actual value for period t

N - Total number of test data

5.3. Experimental Results

Rainfall prediction is essential and of great interest because successful prediction of rainfall may promise

smart benefits. An efficient rainfall prediction technique is proposed to predict the future rainfall. Here, firstly the input rainfall dataset is converted into suitable format for prediction. The obtained experimental results such as, sample results after converting data with feature indicators (dataset 1 to dataset 3) are illustrated in table 1 to table 3. Table 4 shows neural network parameters and Table 5 shows hybrid algorithm parameter.

Table 1. Sample results after converting data with feature indicators (dataset1)

Year	RSI	ROC	CI	UI	ADX	Class
COAPR1884	0.883117	51.1964	-57.68	80.33623	98.40745	0.164707
COAPR1885	-0.52381	59.10658	-28.341	80.33623	98.33436	0.24573
COAPR 1886	22.48276	52.91664	114.6284	72.50257	98.42085	0.181611
COAPR1887	-0.50133	45.96372	-31.9567	72.37471	98.49499	0.127922
COAPR 1888	-0.32609	44.661	-89.0628	71.91118	98.55854	0.179161
COAPR 1889	-0.53556	48.75312	-5.92093	72.01534	98.63921	0.132682
COAPR 1890	-0.16505	46.15693	-129.998	70.81766	98.70835	0.093011
COAPR 1891	-0.61609	46.06025	-109.561	71.4349	98.76762	0.213198
COAPR 1892	-0.81481	46.91636	104.1763	76.26028	98.89069	0.234845
COAPR 1893	-0.9486	42.94072	89.24701	82.18155	98.99617	0.184184

Table 2. Sample results after converting data with feature indicators (dataset2)

Year	RSI	ROC	CI	UI	ADX	Class
TELNG1884	-0.26846	63.90256	103.9532	89.05574	100	0.1472
TELNG 1885	-0.89387	70.4181	-12.9567	80.63717	100	0.187537
TELNG 1886	-0.98512	48.35	63.75933	78.29309	100	0.123915
TELNG 1887	-0.94783	45.83281	-99.2866	82.09056	100	0.113978
TELNG 1888	-0.9906	47.73743	-70.1679	80.60417	100	0.164252
TELNG 1889	-0.99471	46.77796	-4.55684	79.73862	100	0.149109
TELNG 1890	-0.9765	51.78482	-44.4326	76.12948	100	0.106896
TELNG 1891	14.61774	45.3609	-54.754	84.9957	100	0.208244
TELNG 1892	-0.98155	46.01202	170.1155	91.78598	100	0.242168
TELNG 1893	66.19737	45.38109	97.41569	93.79344	100	0.166536

Table 3. Sample results after converting data with feature indicators (dataset3)

Year	RSI	ROC	CI	UI	ADX	Class
RLSMA1884	7.955882	74.22321	22.85923	62.12744	100	0.131447
RLSMA 1885	14.48	78.14069	3.412038	52.66941	100	0.154245
RLSMA 1886	-0.97092	46.12299	-51.1426	48.38013	100	0.148489
RLSMA 1887	1.842466	55.56087	31.59505	51.32285	100	0.128001
RLSMA 1888	1.588235	45.06687	2.81847	53.5199	100	0.164148
RLSMA 1889	3.538462	51.79145	77.5483	54.51422	100	0.145762
RLSMA 1890	-0.76611	45.5944	-113.244	63.73003	100	0.069776
RLSMA 1891	-0.71757	48.30513	-187.488	70.18343	100	0.176153
RLSMA 1892	-0.78553	48.07853	90.93209	73.75635	100	0.165512
RLSMA 1893	33.38281	41.75601	-26.7783	76.6383	100	0.137961

Table 4. Neural network parameters

	No of iteration	Error	Training algorithm	Number of hidden layers	Number of neuron in h1	Number of neuron in h2
NN1	1000	0.3896	Hybrid algorithm	2	500	100

Table 5. Hybrid algorithm parameters

	Length of the chromosome (D)	Colony size	Pa	Mutation rate	Cycle
ABC with genetic	701	20	0.5	20	1000

5.4 Performance Evaluation of Proposed Technique

The performance of proposed rainfall prediction technique is evaluated in terms of the evaluation metrics value. Here, with the aid of the input training and testing dataset, the values of MAPE, RMSE, MAD, and MSE

are estimated. By analyzing the results, proposed technique is found to yield better performance. Table 6 to 8 show MAPE, RMSE and MAD values of various hidden layer in the proposed rainfall technique for the three real datasets.

Table 6. MAPE, RMSE and MAD values of various hidden layers in proposed technique for dataset

Proposed algorithm	MAPE	RMSE	MAD
h1=250; h2=50	22.2674	0.5417	4.1581
h1=500; h2=100	22.3991	0.5338	4.1521
h1=1000; h2=200	22.0378	0.5302	4.0956

Table 7. MAPE, RMSE and MAD values of various hidden layers in proposed technique for dataset 2

Proposed algorithm	MAPE	RMSE	MAD
h1=250; h2=50	22.3389	0.5388	4.1591
h1=500; h2=100	22.0598	0.5276	4.0767
h1=1000; h2=200	22.2402	0.5371	4.1011

Table 8. MAPE, RMSE and MAD values of various hidden layers in proposed technique for dataset 3

Proposed algorithm	MAPE	RMSE	MAD
h1=250; h2=50	22.2197	0.5333	4.1295
h1=500; h2=100	22.6535	0.5506	4.2318
h1=1000; h2=200	20.8949	0.4705	3.6644

From the above tables, we have plotted following figures for three real datasets. In the figures 6 to 8, we have evaluated three measures for various hidden layers. In figure 8 and figure 8, hidden layer 50 is the better value when compared h=100 and h=200 for three measures. In figure 8, hidden value 100 is a better result when compared other two hidden values for three measures.

Fig.6. Performance evaluation graph of MAPE, RMSE and MAD values of various hidden layers in proposed technique for dataset 1

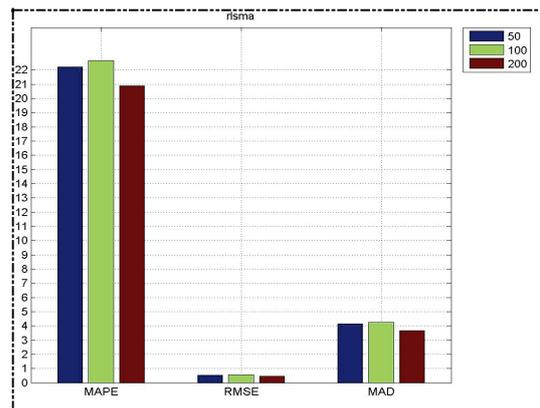
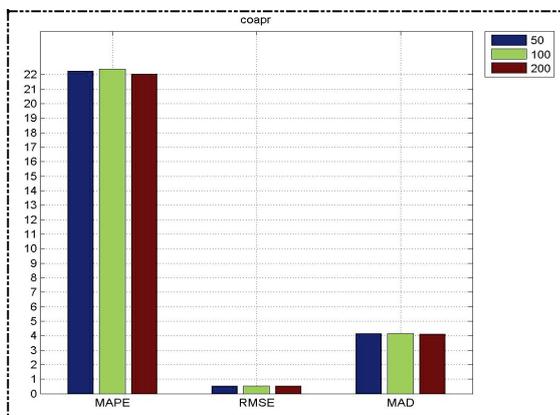


Fig.7. Performance evaluation graph of MAPE, RMSE and MAD values of various hidden layers in proposed technique for dataset 2



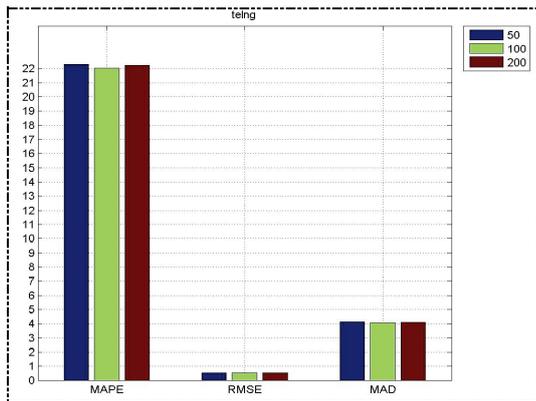


Fig.8. Performance evaluation graph of MAPE, RMSE and MAD values of various hidden layers in proposed technique for dataset 3

5.5. Comparative Analysis Evaluation of the Proposed Technique

In this section, we compare our proposed model against ABC+ Neural network. The performance analysis is made by plotting the graphs of evaluation metrics such as Mean Absolute Deviation, Root mean square error, Mean Absolute Percentage. By analyzing the plotted graph, the performance of the proposed rainfall prediction technique has been found to have significantly improved the rainfall prediction compared with ABC+ Neural Network (NN). The evaluation values of the Mean Absolute Deviation, Root mean square error, and Mean Absolute Percentage graph for three real rainfall datasets are shown in tables 9 to 11. From table 9, the proposed rainfall prediction model is MAPE =22.0576, RMSE=0.5148, MAD=4.0153 when compared ABC+NN (MAPE=22.1471, RMSE=0.5315, MAD=4.1162). From table 10, the proposed rainfall prediction technique is MAPE=22.1756, RMSE=0.5360, MAD=4.1035 when compared with ABC+NN (MAPE=22.3938, RMSE=0.5226, MAD=4.0751). From table 11, the proposed rainfall prediction technique is MAPE=21.9581, RMSE=0.5317, MAD=4.0672 when compared with ABC+NN (MAPE=22.6358, RMSE=0.5377, MAD=4.1878). Totally, the accuracy level has proved that the proposed technique is superior in rainfall prediction.

To ensure effectiveness and validate our rainfall prediction model, we have compared against one previously published paper [31]. Authors [31] have developed a method of short term wind power

prediction for a wind power plant by training neural networks based on historical data of wind speed and wind direction. When compared with their model (based on MAPE range of 9.901-29.192), our proposed model is a little better than MAPE performance of 21.9581, which proves that our proposed rainfall prediction model yields superb performance in the prediction of weather-related events.

Table 9. Comparative analysis of proposed rainfall prediction and NN for dataset 1

	MAPE	RMSE	MAD
Proposed algorithm	22.0576	0.5148	4.0153
ABC+NN	22.1471	0.5315	4.1162

Table 10, Comparative analysis of proposed rainfall prediction and NN for dataset 2

	MAPE	RMSE	MAD
Proposed algorithm	22.1756	0.5360	4.1035
ABC+NN	22.3938	0.5226	4.0751

Table 11. Comparative analysis of proposed rainfall prediction and NN for dataset 3

	MAPE	RMSE	MAD
Proposed algorithm	21.9581	0.5317	4.0672
ABC+NN	22.6358	0.5377	4.1878

5.6 K-fold cross-validation

This section presents the performance analysis of the proposed system using K-fold cross-validation method. According to this, the training data set is divided into k-subsets (k=10) and for every validation, a single subset is used as the testing data and the remaining subsets are utilized as training data. This procedure is repeated until all the subsets of data are utilized as testing data. Here, we have chosen k=10 so that, the training data is divided into '10' sub-samples to extensively analyze the proposed system. The comparative analysis of K-fold cross-validation plots are plotted in figures 9 to 11. Analyzing figures 9, 10 and 11, we observe that our proposed model ushers in better performance when compared with existing system.

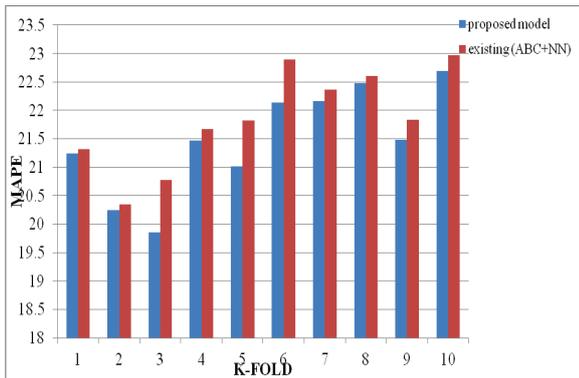


Fig.9. MAPE based Comparison graph of proposed against existing using k-fold cross validation method

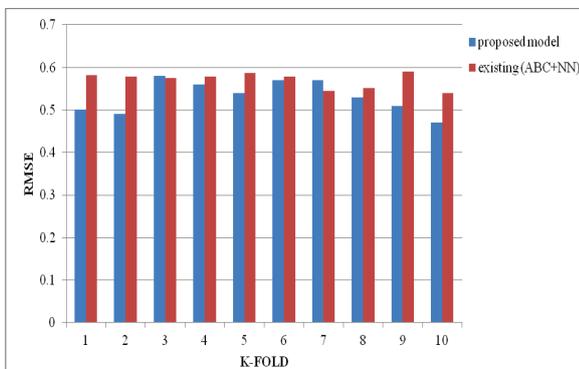


Fig.10. RMSE based Comparison graph of proposed against existing using k-fold cross validation method

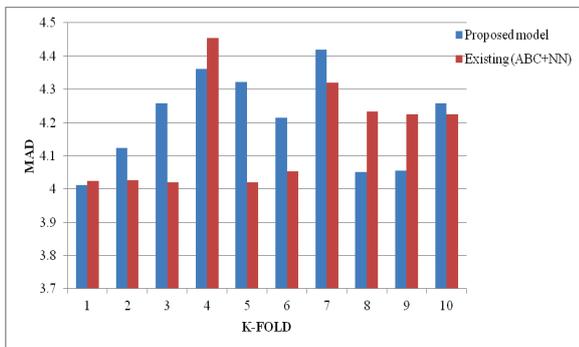


Fig.11. MAD based Comparison graph of proposed against existing using k-fold cross validation method

6. Conclusion

In this paper, we have presented a novel rainfall prediction model using effective feature indicators and a hybrid technique. Here, the proposed rainfall prediction consists of three phases, namely, layer model simulation, training phase and testing phase. Firstly, the input rainfall dataset is preprocessed using feature indicators. There are five feature indicators used in the

preprocessing step namely, channel index (CI), ulcer index (UI), rate of change (ROC), relative strength index (RSI) and average directional movement index (ADX). Subsequently, feature matrices are formed based on the preprocessed rainfall data. Once the feature matrix is formed, the prediction is done based on the hybrid classifier. In the hybrid classifier, artificial bee colony algorithm is combined with genetic for training the feed forward neural network. The performance of the algorithm is analyzed with the help of real datasets taken from rayalaseema, andhra and telangana regions. Finally, from comparative analysis, it is clear that the proposed technique yields better results when compared with ABC+NN. As a future work, we are planning to improve time series rainfall prediction, wavelet neural network with AR modeling and support vector regression (SVR) model will be effectively hybridized and used.

References

- [1] Khaing Win Mar and Thinn Thu Naing, "Optimum Neural Network Architecture for Precipitation Prediction of Myanmar", World Academy of Science, Engineering and Technology, Vol. 48, pp. 130134, 2008.
- [2] Pijush Samui, Venkata Ravibabu Mandla, Arun Krishna and Tarun Teja, "Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine", Journal of Earth Science India, Vol. 4, No. 4, pp. 188 - 200, October, 2011.
- [3] Wint Thida Zaw and Thinn Thu Naing, "Empirical Statistical Modeling of Rainfall Prediction over Myanmar", World Academy of Science, Engineering and Technology, Vol. 46, pp. 565-568, 2008.
- [4] Toth, Brath and Montanari, "Comparison of short-term rainfall prediction models for real-time flood forecasting", Journal of Hydrology, Vol. 239, pp. 132–147, 2000.
- [5] Bimal Dutta and Angshuman Ray, Srimanta Pal and Dipak Chandra Patranabis, "A Connectionist Model for Rainfall Prediction", Neural, Parallel, and Scientific Computations, Vol. 17, pp. 47-58, 2009.
- [6] Kannan, Prabhakaran and Ramachandran, "Rainfall Forecasting Using Data Mining Technique", International Journal of Engineering and Technology, Vol.2, No. 6, pp. 397-401, 2010
- [7] Sivakumar, "Climate prediction and agriculture: current status and future challenges", Climate Research, Vol. 33, No. 1, pp. 3–17, 2006.
- [8] M. Nayebi, D. Khalili, S. Amin, and Sh. Zand-Parsa, "Daily stream flow prediction capability of artificial neural networks as influenced by minimum air temperature data," Bio-systems Engineering, vol. 95, no. 4, pp. 557-567, 2006.
- [9] R. Bustami, N. Bessaih, C. Bong, and S. Suhaili, "Artificial neural network for precipitation and water

- level predictions of bedup river,” IAENG International Journal of Computer Science, vol. 34, no. 2, 2007.
- [10] R. VanRullen, “The power of the feed-forward sweep,” *Advances in Cognitive Psychology*, vol. 3, no. 1-2, pp. 167-176, 2007.
- [11] Y. Bi, J. Zhao, and D. Zhang, “Power load forecasting algorithm based on wavelet packet analysis,” *International Conference on Power System Technology*, vol. 1, no. 21-24, pp. 987-990, November 2004.
- [12] S. Tantane, S. Patamakul, T. Oki, V. Sriboonlue, and T. Prempre, “Downscaled rainfall prediction model (DRPM) using a unit disaggregation curve (UDC),” *Hydrol. Earth Sys. Sci. Discuss.*, vol. 2, pp. 543-568, 2005.
- [13] Wassamon Phusakulkajorn, Chidchanok Lursinsap and Jack Asavanant, "Wavelet-Transform Based Artificial Neural Network For Daily Rainfall Prediction in Southern Thailand", in proceedings of 9th International Symposium on Communications and Information Technology, pp. 432- 437, 2009.
- [14] Kyaw Kyaw Htike and Othman O. Khalifa, "Rainfall Forecasting Models Using Focused Time-Delay Neural Networks", in proceedings of International Conference on Computer and Communication Engineering (ICCCE 2010), May 2010.
- [15] B. Basturk, D. Karaboga, “An artificial bee colony (abc) algorithm for numeric function optimization”, In the IEEE Swarm Intelligence Symposium, Indianapolis, Indiana, USA, May, 2006.
- [16] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, New York, Addison-Wesley, 1989.
- [17] Y. Rahmat-Samii, E. Michielssen, *Electromagnetic Optimization by Genetic Algorithms*, New York: Wiley, 1999.
- [18] P.T. Nastos, K.P. Moustiris, I.K. Larissi, A.G. Paliatsos, “Rain intensity forecast using Artificial Neural Networks in Athens, Greece,” *Atmospheric Research*, vol.119, pp.153-160, 2013.
- [19] Nizar Ali Charaniya, Dr. Sanjay.V.Dudul, “Committee of Artificial Neural Networks for Monthly Rainfall Prediction using Wavelet transform,” *International Conference on Business, Engineering and Industrial Applications (ICBEIA)*, pp.125-129, 2011.
- [20] Lingzhi Wang, Jiansheng Wu, “Application of Hybrid RBF Neural Network Ensemble Model Based on Wavelet Support Vector Machine Regression in Rainfall Time Series Forecasting,” *Fifth International Joint Conference on Computational Sciences and Optimization*, pp.867-871, 2012.
- [21] Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar "A Rainfall Prediction Model using Artificial Neural Network," *IEEE Control and System Graduate Research Colloquium*, pp.82-87, 2012.
- [22] Andrew Kusiak, Xiupeng Wei, Anoop Prakash Verma, and Evan Roz, “Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach,” *IEEE Transactions on Geoscience and Remote Sensing*, no.99, pp.1-6, 2011.
- [23] W.Wang, K.Chau, C.Cheng and L.Qiu, “A Comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series”, *Journal of Hydrology* vol. 374, pp. 294-306, 2009.
- [24] C. L. Wu and K. W. Chau, “Data-driven models for monthly stream flow time series prediction” *Engineering Applications of Artificial Intelligence*, vol. 23, pp. 1350-1367, 2010.
- [25] A. K. Lohani, N. K. Goel and K. K. S. Bhatia, “Comparative study of neural network, Fuzzy logic and linear transfer function techniques in daily rainfall -runoff modeling under different in put domains”, *Hydrological process*, vol. 25, pp. 175-193, 2011.
- [26] P. C. Nayak, et al., A neuro- fuzzy computing technique for modelling hydrological time series” *Journal of Hydrology*, vol. 291. pp. 52-66, 2004.
- [27] M. Z. Kermani and M. Teshnehlab "Using adaptive neuro-fuzzy inference system for hydrological time series prediction.” *Applied Soft Computing*, vol.8, pp. 928-936, 2008.
- [28] A. Jain and A. M. Kumar “Hybrid neural network models for hydrologic time series forecasting”, *Applied Soft Computing*, vol. 7, pp. 585-592, 2007.
- [29] Mehul N Vora, "Genetic Algorithm for Trading Signal Generation Solution to trader’s dilemma: Is it right time to trade? ", *International Conference on Business and Economics Research*, vol.1, 2011.
- [30] Wilder, J.W.,”*New Concepts in Technical Trading Systems*,” Trend Research, Greensboro, 1978.
- [31] Ziqiao Liu, Wenzhong Gao; Yih-Huei Wan; Muljadi, E."Wind power plant prediction by using neural networks, " *IEEE Energy Conversion Congress and Exposition (ECCE)*, pp.3154 - 3160, 2012.