Abstract— Precipitation forecasting is used for multiple reasons in multiple areas like agriculture, energy supply, transportation, etc. Accuracy of precipitation conditions shown in forecast reports is very necessary. In this paper, the review was conducted to investigate a better approach for forecasting which compares many techniques such as Artificial Neural Network, Ensemble Neural Network, Back propagation Network, Radial Basis Function Network, General Regression Neural Network, Genetic Algorithm, Multilayer Perception, Fuzzy clustering, etc. which were used for different types of forecasting. Among which neural network with the back propagation algorithm performs prediction with minimal error, the model derived is run on that basis. Neural network is a complex network which is self-adaptive in nature. It learns by itself using the training data and generates some intelligent patterns which are useful for forecasting. This paper reviews various techniques and focuses mainly on neural network with back propagation technique for precipitation-forecasting.

Keywords— Neural Network, Back propagation Algorithm, ANN, Precipitation Prediction, Multilayer Neural Network, Quantitative Forecast, rainfall forecast.

1. INTRODUCTION

Weather forecasting is a process of identifying and predicting to certain accuracy the climatic conditions using multiple technologies. Many of the live systems rely on precipitation conditions to make necessary adjustments in their systems. Weather Forecasting helps to take necessary measures to prevent damage to life and property to a large extent. Quantitative forecast like temperature, humidity and rainfall are important in agriculture area, as well as to traders within commodity markets. Temperature forecasts are used by utility companies to estimate demand over coming days. Since outdoor activities are severely restricted by heavy rain, snow and the chill; forecasts can be used to plan activities around these events, and to plan ahead and survive them [11].

Nowadays multiple computing techniques are available which can be used for forecasting enhancing its accuracy. Different categories of forecasting methods are Naive approach, Judgmental methods, Quantitative and Qualitative method, Causal or econometric forecasting methods, Time series methods, Artificial intelligence methods, etc.

The precipitation forecast reports needs some intelligent computing which can read the nonlinear data and generate some rules and patterns to study and train from the observed data to predict the precipitation in future. Use of ANN will give results which are more accurate. Here, the error may or may not reduce completely. But, the accuracy will improve as compared to previous forecasts.

The precipitation forecasting is live forecasting where the output of the model may be required for precipitation guide our weekly or monthly precipitation plans. Thus, the accuracy of the result is a very important aspect in this forecasting. Multiple issues were discussed which can be considered to get the accurate results. In Section two, a reviews multiple literature on precipitation forecasting. Section three introduces different terms about the neural network. Section four proposes a neural network model with all the specifications for forecasting precipitation with a high degree of accuracy. The object of the study is to develop precipitation (Rainfall) forecasting model. In the present research the rainfall data was selected for the area of shetrunjaya dam, near palitana, Gujarat-India.

2. LITERATURE SURVEY

This section reviews literature on different Precipitation forecasting techniques using neural network.

A. Jun Han & Shastri Annambhotla (May/June 2005)

Author had developed an Artificial Neural Networks Model for Forecasting Watershed Runoff and Stream Flows. The selected study area was Greensboro, North Carolina, which is small urban watershed. A storm water runoff prediction was done by watershed runoff Model at a gauged location near the watershed release. Another Model known stream flow Model was developed to forecast river flows at downstream locations along the same groove. Both models include the current and preceding records of rainfall and stream flow gathered at the watershed outlet and downstream locations. The application of ANN to hydrologic forecasting models in the case of above two models was successful.

B. Sanjeev Kumar, Ajay Indian, Zubair Khan (2009)

A neural net model was developed for ground water level prediction. The models consider rainfall and current Groundwater layer as input parameter also considers rainfall-runoff as an important factor.

C. Ch. Jyosthna Devi et al., 2012 [1]

The study presented an ANN based algorithm for predicting the temperature. The BPN is used because it can fairly approximate a large class of functions. Authors propose a model which takes real time dataset with fifteen
parameters as input, which was then normalized using min-max normalization to scale data between zeros to one. Then it was trained and tested using the Backpropagation Neural Network. The results were compared with the meteorological department to check the least error and accuracy of the model. It was found that the model had the potential for temperature forecasting. Anomaly in the paper is that, it does not predict the complete precipitation condition like rainfall, clouds, etc.

D. Rainfall forecasting was done using Ensemble neural network (ENN) by Harshani R. K. Nagahamulla et al., 2012 [2].

In ENN, finite numbers of ANN are trained for the same task and their results were combined using the weighted average method. Here, each ANN was assigned a weight to minimize mean square error. The study area selected to be Colombo, where daily observed data of forty one years was used by dividing it in four climatic seasons every year with twenty six variables. The performance was compared with Backpropagation neural network (BPN), radial basis function network (RBFN) and general regression neural network (GRNN). Results show that, ENN model predicts rainfall more accurately than individual BPN, RBFN and GRNN.

The paper compared ANN with BPN, RBFN and GRNN and the comparison shows that ANN had given more accuracy compared to others.

E. M. Nasseri et al., 2008 [3]
A research was conducted by M. Nasseri et al., 2008 [3], in which they developed Feed forward type network to forecast rainfall using the backpropagation algorithm coupled with the Genetic Algorithm (GA). The study area selected was Sydney, Australia consisting fourteen recording rain gauges of four years data. Measurements of rainfall were at intervals of five minutes. After preliminary data analysis, twenty six storm events were selected for synchronization. Among which eighteen events were selected for training and four for testing. The paper concluded that the multilayer perceptions (MLP) type network coupled with GA performs better than MLP type network alone.

It compares the performance of Multilayer Perceptron type networks and itself combined with GA. Thus, found that adding GA to MLP improves performance.

F. Lee et al., 2004 [4]

Author proposed an innovative, intelligent, multi-agent based environment named as intelligent Java Agent Development Environment (iJADE). It is used for precipitation prediction of eleven precipitation stations in Hong Kong using five years data which provided more than 7300 data records. The model uses GA for input node selection, a fuzzy classification for rainfall parameters and neural network for training using a BPN. Its experimental results were more promising than single point sources using similar network and other networks like Radial Basis Function Network, Learning Vector Quantization and Naive Bayesian Network. The authors compared ANN with RBF, LVQ and Naive Bayesian Network and proved that ANN’s results were better as compared to others.

G. Temperature Forecasting based on Neural Network Approach, Moshe hayati et al., 2007 [5]

The study contained use of ANN for one day prediction of temperature. Here multilayer perception (MLP) was trained using 65% of patterns and tested using 35% of patterns for ten years meteorological data from Iran which was split into four seasons namely spring, summer, fall and winter. MLP network of three layers with sigmoid transfer function for hidden layers and linear transfer function for output layer was used. Number of hidden neurons and epochs were decided using the trial and error method. The paper concludes that MLP with the structure had minimum prediction error, good performance and reasonable prediction accuracy. The above paper used historical data for ten years and achieved minimum error and reasonable accuracy.

H. Kumar Abhishek et al., 2012 [6]

Author developed an ANN model to forecast average monthly rainfall. He selected data from Udupi, Karnataka which was eight months data for fifty years making 400 entries for input and output. The data was normalized by finding mean and standard deviation of each parameter. Then training was done on 70% of data and the remaining 30% data was used for testing and validation. The model used a three layered ANN with back propagation learning. Later after testing, the results were compared with actual output. It showed higher degree of similarity in output. Thus it was proved that ANN model is accurate in prediction. Authors concluded that, learngdm is the best learning function for training whereas trainlm is the best training function. The data used for research was huge enough i.e. of fifty years, thus, larger input data reduced the mean square error.

I. Yamin Wang et al., 2013 [7]

Author proposes a novel wind speed forecasting method based on ensemble empirical mode decomposition (EEMD) and Genetic algorithm – backpropagation Neural network. The wind speed data recorded after every ten minutes was taken for five days which gave a total of 721 data for testing. The wind speed data was first decomposed by EEMD in eight different IMF’s and residue. Later GA – BPN was used for forecasting each IMF. Then IMF results were combined and then the forecasting result for wind speed was obtained. Thus, the method proved to forecast wind speed more precisely as compared to traditional GA-BP and hybrid of EMD and GA-BP. But here the forecasting was done considering only wind speed, which was not enough for precipitation forecasting.

J. Saima H. et al., 2011[8]

The study reviewed many hybrid methods, that has been used for precipitation forecasting with their merits and demerits. They used consortiums of statistical and artificial
intelligent methods for achieving accuracy. Different models studied are

1. Hybrid autoregressive moving average (ARMA) model and ANN for wind speed prediction
2. Adaptive Neuro-Fuzzy Inference system (ANFIS) which is the integration of Fuzzy Inference System and Neural network used for rainfall runoff prediction
3. Fuzzy clustering and Type-2 Fuzzy Logic
4. A Neuro evaluative Interval Type -2 TSK Fuzzy System
5. Grey Relational Analysis.

In all forecasting models, the major concern was with accuracy not the processing time. It was observed that a single forecasting model can forecast with 100% accuracy, but they could reduce the accuracy error by various techniques. Authors had compared all the techniques with their accuracy and found that no model can be totally accurate but near optimum results were expected.

K. Tony Hall et al., 1999 [9]

The present study discussed about ANN for Probability of Precipitation (PoP) and Quantitative precipitation forecast (QPF) for Dallas-fort worth, Texas area. Neural Network was developed and initially two years data was used consisting nineteen variables. Later for three years the verification reports were generated. In the application, two networks were created, a QPF network for predicting amount of precipitation and a PoP network for probability or confidence in the forecast. The networks were designed with three features, first are a separate network for different seasons like warm and cool, and second are the use of QPF and PoP and the last is the network made interactive so we can rerun the entire network with some changes. The technique improves precipitation forecast dramatically, particularly for applications requiring accurate results.

Authors stated that QPF and PoP together can improve the performance.

3. ARTIFICIAL NEURAL NETWORK FOR PRECIPITATION (RAIN FALL) FORECASTING

Forecasting it is intuitive that accuracy is very important. The input parameters for a precipitation forecasting model are different types of data need different types of methods; and need to be handled accordingly. Statistical methods are usually associated with linear data whereas Artificial Intelligence methods are associated with nonlinear data [13]. Different learning models based on Artificial Intelligence are genetic algorithms, neuron-fuzzy logic and neural networks. Among which neural networks is preferred for time series forecasting for applications such as “stock index forecasting” in financial markets or “fault detection” in machine maintenance-[14]. Precipitation forecasting can be done more accurately using ANN. Because precipitation data has multiple parameters representing temperature, humidity, rainfall amount, cloud distance and size, wind speed and direction, etc. All these parameters are not linear, but they need to be processed together to determine temperature, rainfall, humidity or precipitation status for the next day. Such type of applications use the models which are complex in nature and can produce the required result by generating the patterns on its own by performing self-learning using the training data given to the model.

To develop an ANN model for precipitation forecasting, selection of region for input data and parameters is necessary. The input data is to be taken from a specific area on which the model is trained and tested so that the model is able to generate accurate results. The number of input data given to model also helps to improve accuracy of the model by giving the results with a high degree of similarity between predicted and actual output data. The available data may be noisy but, data should be cleaned. Similarly, it has to be normalized because, all the parameters are of different units and normalization will help the input and output parameters to correlate with each other [6]. The data should be divided in training and testing samples in proper proportion so that the results can be predicted, tested and validated properly. Structure of the NN model also has a great impact on generation of accurate results. The multilayer ANN helps in predicting nonlinear data more efficiently. The activation function will be different for different layers of NN as per need.

4. ARTIFICIAL NEURAL NETWORKS (ANN)

One type of network that sees the node as an ‘artificial neurons’ is called artificial neural networks. Artificial neural network is a software implementation which resembles the biological term central nervous system that is the human brain. Natural neurons receive signals through synapses located on the dendrites or membrane of the neuron. When the signals received are strong enough (surpass a certain threshold), the neuron is activated and emits a signal though the axon. This signal might be sent to another synapse, and might activate other neurons [15].

The complexity of real neurons is highly abstracted when modeling artificial neurons. These basically consist of inputs (like synapses), which are multiplied by weights (strength of the respective signals), and then computed by a mathematical function which determines the activation of the neuron. Another function computes the output of the artificial neuron (sometimes in dependence of a certain threshold).

This network is of a very complex type because it contains multiple neurons associated with each other in a well-formed structure to produce complex output with minimal error.

This network is mostly used to make predictions by training the model using past dataset and experience. Above Figure 1 shows an ANN structure where multiple input parameters are used as input neurons, which are then multiplied by weights and forwarded to a hidden layer.
where activation function is applied and then forwarded to output layer with another activation function where it finally computes the output of artificial neuron.

5. THE BACKPROPAGATION OF ALGORITHM PROPOSED APPROACH

The back propagation algorithm is used in layered feed-forward ANNs. It uses supervised learning, which means the model trains itself with the use of target output. For every set of input data the target output was provided. The neural network model processes the input data with random values for weights and suitable activation function using one or more hidden layer in between and then produces the predicted output. This predicted output was then compared with the target output provided for same input dataset. Thus, error was calculated by subtracting predicted output from target output. Using this error, the weights were adjusted and again the entire process was repeated for multiple epochs until the error was minimal or in acceptable range [15].

The idea of the back propagation algorithm is to reduce this error, until the ANN learns the training data. The training begins with random weights, and the goal is to adjust them so that the error will be minimal.

For practical reasons, ANNs implementing the back propagation algorithm do not have too many layers, since the time for training the networks grows exponentially. Also, there are refinements to the back propagation algorithm which allow a faster learning.

6. BACKPROPAGATION APPROACH

The model proposed in this paper for precipitation forecasting using ANN using BP algorithm is as given below in Figure 2. The area for input data can be any one of a meteorological station area in which all the data was limited to a certain region. The different input parameters are taken viz. temperature, relative humidity, air pressure, wind speed and direction, cloud amount and height, rainfall, etc.

Input data was then pre-processed and cleaned. That means it was checked with any outlier and that was removed, missing values were entered, and data was checked if it was in the given range for the given parameter. Later ANN was designed with number of input and output nodes, hidden layers, activation function, and maximum number of epochs, weights, bias, goal and learning function. Neural network was trained with seventy percentages of the input data. Where the model was trained using this observed data to forecast the weather, followed by testing done using remaining thirty percentages of input data. Then the mean squared error and accuracy was calculated for the model by comparing the output of testing with target output.

This model generates output in terms of minimum and maximum temperature of the day, relative humidity and rainfall. The intensity of rainfall was represented by different classes as shown in Table 1. The range of classes for intensity of rainfall and cloud status ranges are taken from the Indian meteorological department because the input dataset is also from same source.

Seventy percentages of the dataset used for training and the other thirty percentages of the dataset used for testing and validation. Hidden layers were required for processing nonlinear data. Number of hidden layers in a network should be selected on a trial and error basis. Because as we increase the number of hidden layers; complexity of the network increases.

Table 1. Rainfall Data

<table>
<thead>
<tr>
<th>Descriptive Term used</th>
<th>Rainfall amount (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Rain</td>
<td>997.83</td>
</tr>
<tr>
<td>Minimum Rain</td>
<td>135.55</td>
</tr>
<tr>
<td>Average Rain</td>
<td>507.16</td>
</tr>
<tr>
<td>Extremely Heavy Rain</td>
<td>997.83</td>
</tr>
</tbody>
</table>

“Over fitting”. Thus, the number of hidden layer neurons can be decided by following the rule-of-thumb [10].

1. The numbers of hidden layer neurons are 2/3 (or 70% to 90%) of the size of the input layer. If this is insufficient then a number of output layer neurons can be added later on.

2. The number of hidden layer neurons should be less than twice of the number of neurons in an input layer.

3. The size of the hidden layer neurons is between the input layer size and the output layer size.

Weights and bias values are initially taken randomly and then during the training period values were adjusted automatically by comparing the mean squared error with the goal value defined. Learning rate trains the network with a constant value provided. Better results can be achieved with high accuracy when learning rate was smaller but its...
performance was slower. Activation functions were applied on each neuron to get the output of neuron on a given input in the neural network. The sigmoid function was a special case of logistic function which had a sigmoid curve. The sigmoid transfer function can be used for hidden layers and for the output layer the linear transfer function can be used.

The correctness and accuracy of the model can be checked using the Mean Squared error (MSE) function. The MSE measures the average of the squares of errors that was, the difference between the actual output and the predicted output of the model. Lesser the MSE value of the model, more accurate the results are.

7. RESULTS AND DISCUSSION

In this paper, different methods for precipitation (Rainfall) forecasting were reviewed. ANN with backpropagation is recommended for precipitation forecasting.

ANN with back propagation uses an iterative process of training where, it repeatedly compares the observed output with targeted output (Figure 3 Shows the observed v/s targeted output of best outfit model) and calculates the error. This error was used to readjust the values of weights and bias to get an even better output. Hence this method tried to minimize the error. Thus, Artificial Neural network with Back propagation algorithm seems to be most appropriate method for forecasting precipitation accurately. Some graphs of best fit model are as said. The Training Testing, Validation and Regression value is shown in figure 4. Training performance is shown in figure 5. Minimum or negligible error graph is shown in figure 6. Results of some best models are shown in table 2.

The precipitation Forecasting has a big challenge of predicting the accurate results which are used in many real time systems like electricity departments, airports, tourism centers, etc. The difficulty of this forecasting is the complex nature of parameters. Each parameter has a different set of ranges of values. This issue is addressed by ANN.
It accepts all complex parameters as input and generates the intelligent patterns while training and it uses the same patterns to generate the forecasts.

8. CONCLUSION

The Artificial Neural Network model proposed in this paper indicates all the parameters for input and output, training and testing data set, number of hidden layers and neurons in each hidden layer, weight, bias, learning rate and activation function. The Mean Squared Error between predicted output and the actual output is used to check accuracy.

It was seen that multiple FFBP simulations were required until obtaining a satisfactory performance criteria. This total duration for FFBP simulations was longer than the unique GRNN application. Another important result of the study is the positive contribution of the utilization of initial statistical analysis results in determining the ANN input layer node number. The rainfall-runoff correglograms point to the number of rainfall values significant in flow estimations. This can be a time saving feature since the input layer nodes are found by trial and error in general. Further suggestion given by the author of this paper, namely Jignesh Joshi, reveals that like this method, the most accurate method of FFBP can also be used for other watershed analysis, too and as it gives good and perfect results, the author also finds feasibility of recommending this method for important projects, too.

ACKNOWLEDGMENT

This work has been supported by the state water data centre (SWDC).

REFERENCES


[15] Carlos Gershenson, “Artificial Neural Network for Beginners”, C.Gershenson@sussex.ac.uk, university of sussex.