Journal of Science and Technology
ISSN: 2456-5660 Volume 5, Issue 6, Nov-December 2020

www.jst.org.in *DOI: https://doi.org/10.46243/jst.2020.v5.i6.pp26-36*

Rainfall Forecasting Based on Surface Data of Chennai Region Using Artificial Neural Networks

P. Dhandapani[1], Dr. T. Anuradha[2]
[1](Assoc. Professor, S.V. College of Computer Sciences, R.V.S. Nagar, Chittoor (AP), India)
[2](Professor, Dept of Computer Science, Dravidian University, Kuppam (A.P) India)
[1]Corresponding Author: pdpani@yahoo.com

To Cite this Article
P. Dhandapani and Dr. T. Anuradha, "Rainfall Forecasting Based on Surface Data of Chennai Region Using Artificial Neural Networks", Journal of Science and Technology, Vol. 05, Issue 06, Nov-December 2020, pp26-36

Article Info
Received: 15-06-2020 Revised: 20-08-2020 Accepted: 15-09-2020 Published: 24-09-2020

Abstract: *In this study, we developed user friendly rainfall forecasting system based on Back propagation Neural Network using MATLAB 7.10 to forecast Hourly rainfall in Chennai region. The dataset of 31488 samples has been collected from Nungambakkam Meteorological Station, Chennai for the period of 2005 to 2015. The data was organized into day-wise hourly recordings as well as day-wise, maximum, minimum, average data of Relative Humidity (RH), Temperature, Pressure and Wind Speed along with Rainfall data. The collected dataset has been used both for training and for testing the data. The developed system gives more accuracy of 94.8197% when the training data set is 55% and the testing data set is 45% with least Mean Squared Error (MSE) value 0.012437.*
Keywords*: Back propagation Neural Networks, Rainfall Forecasting, Artificial Intelligence, Data Mining.*

I. Introduction

Forecasting of rainfall is a tedious and complicated process, timely forecasting and the accuracy of forecasting is the need of the hour. More rains leads to floods and other natural calamities whereas less rainfall leads to drought. Forecasting of rainfall helps in agriculture, aqua farming and other water resource management. Based on rainfall forecasting, farmers are able to choose which crop to raise, to reap maximum agricultural products. Many of the existing forecasting methods were able to forecast seasonal monsoons like southwest monsoon, northeast monsoon, summer, winter monsoons etc.

Moreover the existing methods were forecasting rainfall over large regions such as state-level, zone levels like Rayalaseema, Coastal Andhra Pradesh, North Karnataka and South Karnataka etc. In the next level, existing systems predict rainfall over district-wise but not over smaller regions like villages, hamlets etc. In addition, existing systems uses advanced climatology indices such as Southern Oscillation Index (SOI), El Nino Southern Oscillation (ENSO), East Atlantic (EA), North Atlantic Oscillation (NAO), Inter-decadal Pacific Oscillation (IPO), in this study, the proposed system is to develop to forecast daily rainfall using surface data such as relative humidity, pressure, temperature and wind speed. Extraction of surface data over smaller regions is much easier and less expensive than the climatology indices. It is more essential to forecast rainfall over smaller regions such as mandal level, village level and hamlet level.

In this study, the proposed system is to forecast daily rainfall using neural network over smaller regions using surface data. Backpropagation neural network is proposed which is multi layered feedforward network and is trained by Delta Learning rule. The proposed system is designed and developed using MATLAB 7.10. The surface data is obtained from India Meteorological Department (IMD), Pune, by registering with it and the surface data collected were pertaining to Nungambakkam Weather Station, Chennai.

II. Surface Data and Data Collection

Surface data consist of Relative Humidity (RH), Temperature, Pressure and Wind Speed along with Rainfall data. The surface data is collected from India Meteorological Department (IMD), Pune. Data is collected through research registration with IMD, Pune. Data is collected related to Nungambakkam Meteorological Station

(Station Index No. 43278), Chennai, for the period 2005 to 2015. The data was organized into day-wise hourly recordings as well as day-wise, maximum, minimum, average data. The data file organization and sample data of Temperature, Pressure, Wind Speed and Rainfall are shown in Fig 1. In this study to forecast rain fall, Relative Humidity, Temperature, Pressure and Wind Speed are taken as input parameters and Rainfall as target parameter.

Fig 1: File Organization of Rainfall Data (Data Credit: IMD)

```
ABBREVIATIONS USED FOR SRRG DATA :
 -------------------------------
  INDEX = INDEX No. OF THE STATION
  MN    = MONTH (01-12)
  DT    = DATE (01 to 31)
  HRF01 = HOURLY RAINFALL IN MM FOR HOUR 01 IST
  HRF02 = HOURLY RAINFALL IN MM FOR HOUR 02 IST
   ...    ...
  HRF23 = HOURLY RAINFALL IN MM FOR HOUR 23 IST
  HRF24 = HOURLY RAINFALL IN MM FOR HOUR 24 IST
  TOTRF = TOTAL R/F IN 24 HOURS IN MM
  MAXRF = MAXIMUM R/F IN ONE HOUR IN A DAY IN MM
  H1 H2 = OCCURRENCE OF HOURLY MAXIMUM R/F BETWEEN HOURS H1 AND H2
  HR MI = DURATION OF TOTAL R/F IN A DAY" IN HOURS AND MINUTES

AUTOGRAPHIC  (SRRG) DATA LISTING :
------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------
INDEX YEAR MN DT HRF01 HRF02 HRF03 HRF04 HRF05 HRF06 HRF07 HRF08 HRF09 HRF10 HRF11 HRF12 HRF13 HRF14 HRF15
HRF16 HRF17 HRF18 HRF19 HRF20 HRF21 HRF22 HRF23 HRF24  TOTRF MAXRF H1 H2 HR MI
------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------
43278 2005 01
01    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0
.0    .0    .0    .0    .0    .0     .0
43278 2005 01
02    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0
.0    .0    .0    .0    .0    .0     .0
43278 2005 01
03    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0
.0    .0    .0    .0    .0    .0     .0
43278 2005 01
04    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0
.0    .0    .0    .0    .0    .0     .0
43278 2005 01
05    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0    .0
.0    .0    .0    .0    .0    .0     .0
```

When 4 input parameters (Relative Humidity, Pressure, Temperature and Wind Speed) are considered, total number of records obtained after data pre-processing for the period 2005 to 2015 are 31488. These records are used both for network training and testing.

III. Artificial Neural Network

Artificial Neural Network is an information processing paradigm that is inspired by the biological nervous system. Artificial Neural Networks allow the systems to recognize the input patterns by learning from past experiences or examples. The major tasks of Artificial Neural Networks are Function Approximation, Classification, Clustering, Decision Support Systems, etc,.

There are different models of Neural Networks namely Perceptron Neural Network, ADALINE, MADALINE, Back propagation Neural Network, etc,. Among these models Back propagation Neural Network Model is the most commonly and widely used model.

Back propagation Neural Network is a Multilayer Feed Forward Neural Network which uses the extended gradient-descent based delta learning rule and also known as back propagation (of errors) rule for training. Back propagation rule provides a computationally efficient method for changing the weights in a Feed Forward Network with differentiable activation function units to learn a training set of input-output examples. Being a gradient descent method it minimizes the total squared error of the output computed by the net. The network is trained by supervised learning method. The aim of this network is to train the net to achieve the balance between the ability to respond correctly to the input patterns that are used for training and ability to provide good responses to the input that are similar. In the present study, for rainfall forecasting user friendly system is developed using MATLAB 7.10 by implementing Back propagation Neural Network model.

IV. Rain Fall Forecasting With Hourly Surface Data

In this study, to improve the accuracy in forecasting of daily rainfall, an attempt is made to forecast the rainfall using hourly data rather than daily average data. The input data file consisting of data related to relative humidity, pressure, temperature, wind speed (4 input parameters) and rainfall (one output parameter) is loaded into workspace for training and testing of the network. In this developed system, network is trained using Delta Learning

Rule, sigmoid function is used as an activation function for adjusting the weights and two hidden neurons are used. An example input data file consisting of hourly surface rainfall data is shown in figure 2.

Fig 2: Input Data File with Hourly Data (4 input parameters)

```
85,1013.9,22.1,0
88,1013.7,21.6,0
91,1013.2,21.2,0
93,1013.1,20.8,0
93,1013.1,20.8,0
93,1013.1,21.2,0
93,1013.6,21.3,0
89,1014.1,22.7,0
86,1016,24.1,0
68,1016,25.6,0
60,1016,27.2,0
56,1015.5,28.5,0
```

V. Experiments and Results

By using the developed system different types experiments are performed for forecasting of hourly rainfall data. In all the experiments the surface data of Relative Humidity, Pressure, Temperature and Wind Speed are taken as input parameters and the occurrence of the rainfall is taken as the target parameter. From the collected data, some of the data are taken for training and some of the data for testing. All the experiments performed along with results obtained are shown in the following. Again to validate the accuracy rate obtained from this neural network model, Heidke Skill Score (HSS) has been computed.

Experiment 1:

This experiment is implemented using 90% of data for training and 10% of data for testing which is shown in fig 3. The experiment and the results obtained are shown in fig 3 and fig 4. The graphical representation of the result is also shown in fig 5. In the below figures 4 and 5, it can be observed that the accuracy rate of rainfall forecasting is achieved as 88.1512 and the least Mean Squared Error (MSE) is achieved at epoch 2 with MSE value as 0.010962.

Fig 3: User Interactive to Input Parameter Size to Network

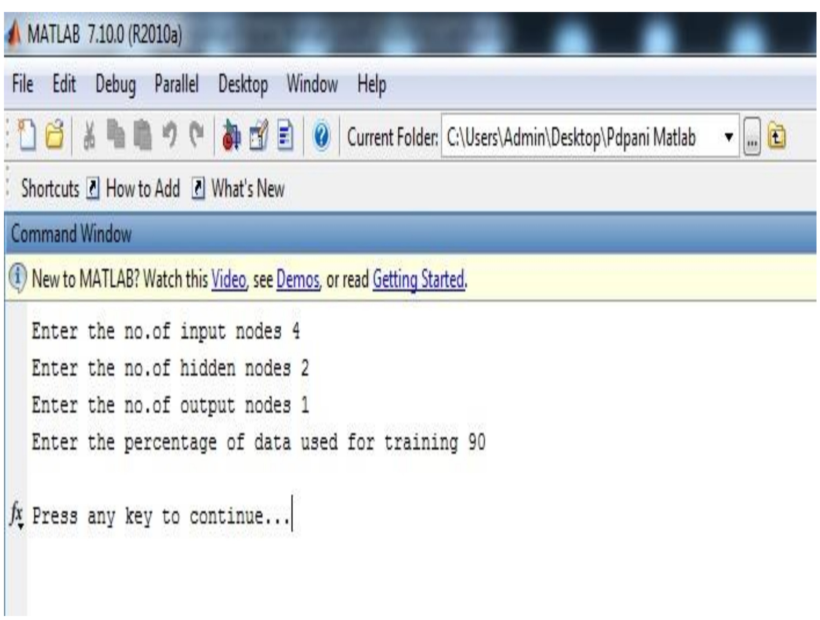


Fig 4: Result of Neural Network Model with 90% Training Dataset

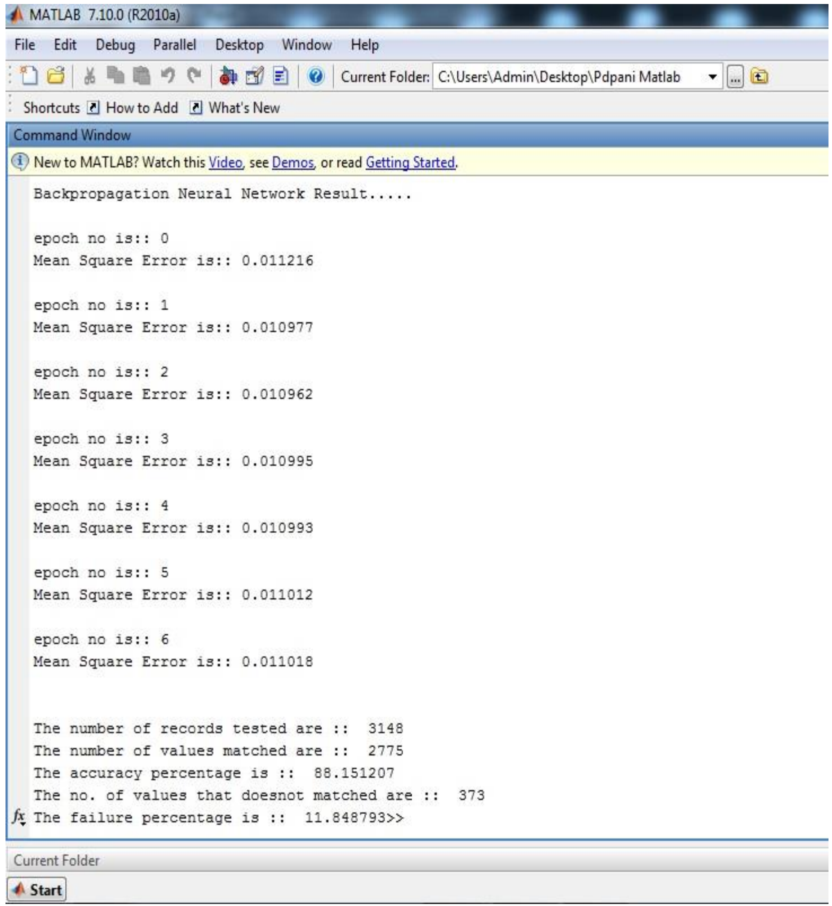
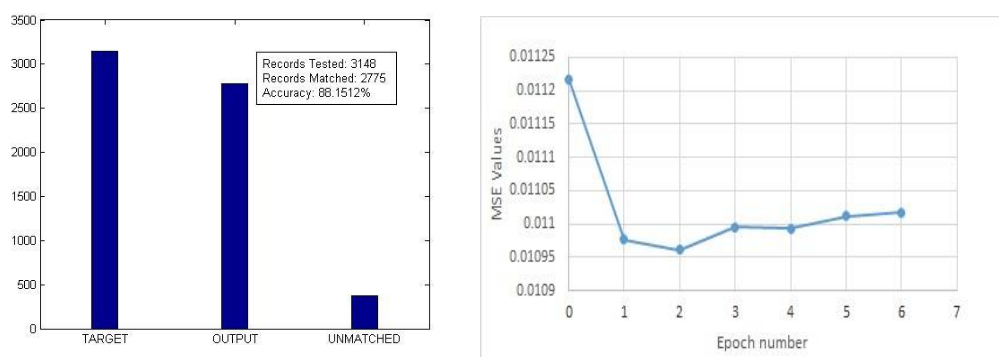


Fig 5: Graphical Representation of Result with 90% Training Dataset



The results obtained from this experiment 1 were tabulated in Table 1 to compute HSS score of this rainfall forecasting. In Table 1, for the total 3148 records tested, 2775 records were matched and the Heidke Skill Score computed for this experiment is 0.73244.

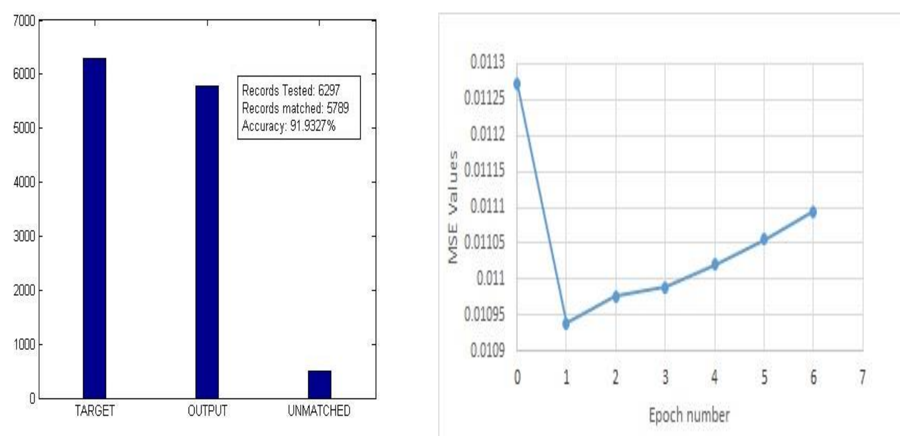
Table 1: Result with 4 input Hourly data, 2 Hidden Neurons and 90% Training Data

Rainfall Forecasted	Rainfall Occurred		
	Yes	No	∑Occurred
Yes	2775	336	3111
No	373	1943	2316
∑ Forecasted	3148	2279	5427

Experiment 2:

This experiment is implemented using 80% of data for training and 20% of data for testing. The graphical representation of the result is shown in fig 6. Form the below figure 6, it can be observed that the accuracy rate of rainfall forecasting is achieved as 91.9327 and the least Mean Squared Error (MSE) is achieved at epoch 1 with MSE value as 0.010939.

Fig 6: Graphical Representation of Result with 80% Training Dataset



The results obtained from this experiment 2 are tabulated in Table 2 to compute HSS score of this rainfall forecasting. From the below Table 2 , for the total records 6297 tested, 5789 records were matched and the Heidke Skill Score computed for this experiment with 80% training dataset and with hourly data is 0.81666 which is better than with 90% training dataset.

Table 2: Result (with 4 Input Hourly Data, 2 Hidden Neurons, 80% Training Data)

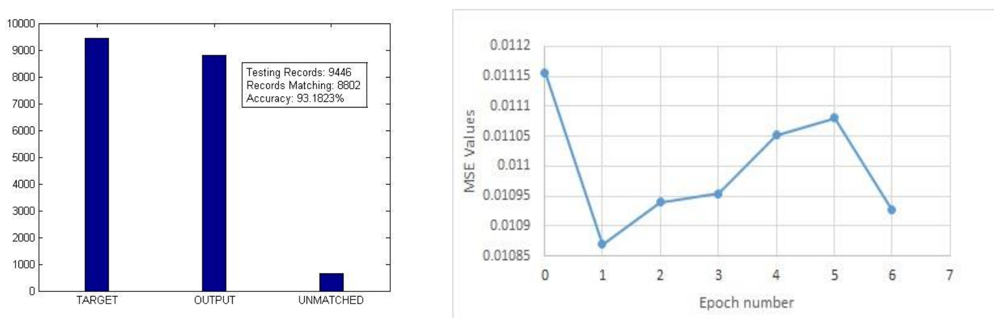
Rainfall Forecasted	Rainfall Occurred		
	Yes	No	∑Occurred
Yes	5789	457	6246
No	508	4052	4560
∑ Forecasted	6297	4509	10806

.

Experiment 3:

This experiment is implemented using 70% of data for training and 30% of data for testing. The results obtained from this experiment are shown graphically in the figure 7.Form the figure 7, it can be observed that the accuracy rate of rainfall forecasting is achieved as 93.1823 which is better than experiment 2 and the least Mean Squared Error (MSE) is achieved at epoch 1 with MSE value as 0.010870.

Fig 7: Graphical Representation of Result with 70% Training Dataset



The results obtained from this experiment 3 are tabulated in Table 3 to compute HSS score of this rainfall forecasting. From the below Table 3, for the total 9446 records tested, 8802 records were matched in forecasting and the Heidke Skill Score computed for this experiment is 0.84463.

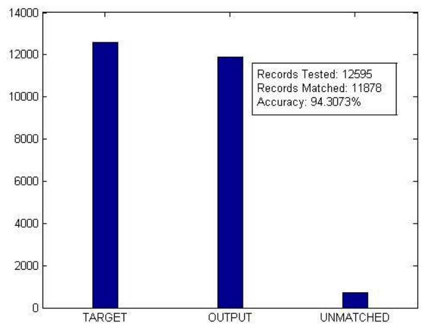
Table 3: Result with 4 Input Hourly Data, 2 Hidden Neurons, 70% Training Data

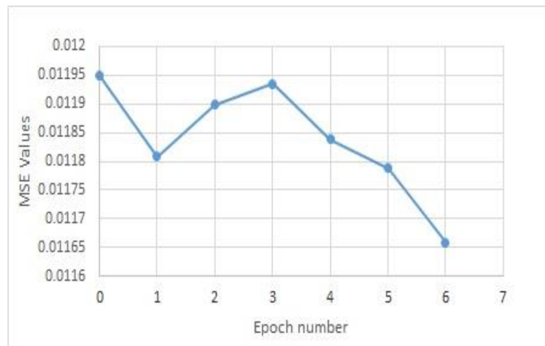
Rainfall Forecasted	Rainfall Occurred		
	Yes	No	∑Occurred
Yes	8802	580	9382
No	644	6161	6805
∑ Forecasted	9446	6741	16187

Experiment 4:

This experiment is implemented with 60% of data for training and 40% of data for testing. The results obtained from this experiment are shown graphically in the figure 8. Form the figure 8, it can be observed that the accuracy rate of rainfall forecasting is achieved as 94.3072 which is better than experiment 3 and the least Mean Squared Error (MSE) is achieved at epoch 6 with MSE value as 0.01165 which is shown in fig 8.

Fig 8: Graphical Representation of Result with 60% Training Dataset





The Heidke Skill Score of this forecast shall be computed with 60% training dataset and 40% of testing data set. The results are shown in table 4. In Table 4, for the total records 12595 tested, 11878 records were matched and the Heidke Skill Score computed for this experiment with 60% training data is 0.87008. The Accuracy rate and HSS scores of this experiment (with 60% training data) is better than with 90%, 80%, and 70% training datasets.

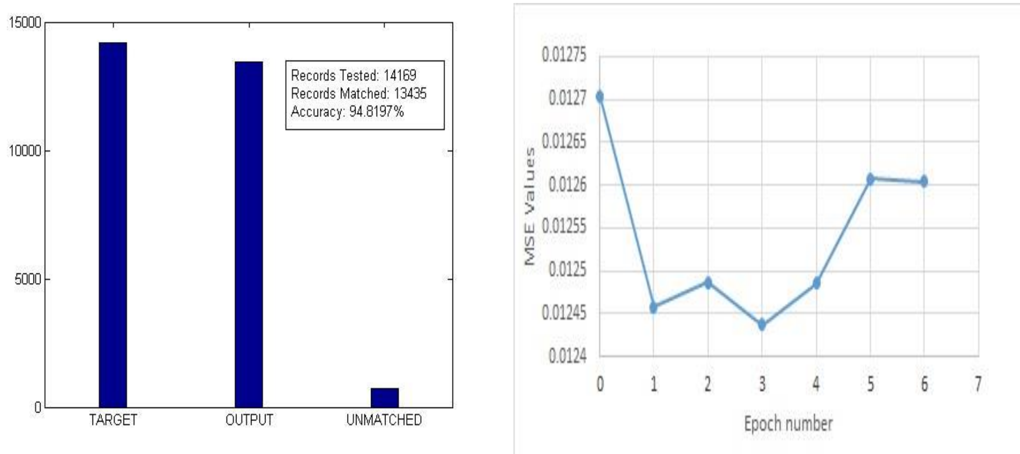
Table 4: Result with 4 Input Hourly Data, 2 Hidden Neurons, 60% Training Data

Rainfall Forecasted	Rainfall Occurred		
	Yes	No	∑Occurred
Yes	11878	645	12523
No	717	8315	9032
∑ Forecasted	12595	8960	21555

Experiment 5:

This experiment is implemented with 55% of data for training and 45% of data for testing. Figure 9 shows the results obtained from this experiment graphically.. Form the figure 9, it can be noted that the accuracy rate of rainfall forecasting is achieved as 94.8197% which is better than experiment 4and the least Mean Squared Error (MSE) is achieved at epoch 3 with MSE value as 0.012437which is shown in fig 9 which is better than the previous experiments.

Fig 9: Graphical Representation of Result with 55% Training Dataset



The Heidke Skill Score of this forecast shall be computed with 55% training dataset and 45% of testing data set. The results are shown in table 5. In Table 5, for the total of 14169 records tested, 13435 records were matched and the Heidke Skill Score (HSS) computed for this experiment with 55% training data is 0.8816. So far, the HSS score achieved in this experiment is higher than in previous experiments.

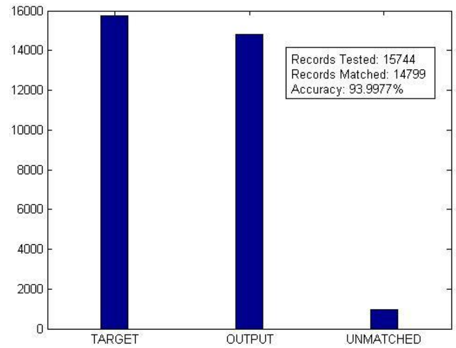
Table 5: Result with 4 Input Hourly Data, 2 Hidden Neurons, 55% Training Data

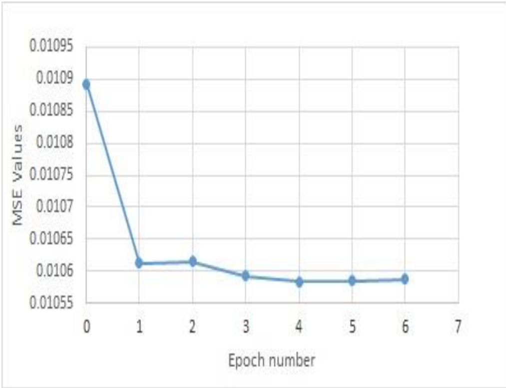
Rainfall Forecasted	Rainfall Occurred		
	Yes	No	∑Occurred
Yes	13435	661	14096
No	734	9405	10139
∑ Forecasted	14169	10066	24235

Experiment 6:

The final experiment is implemented using 50% of data for training and 50% of data for testing and the results achieved through this experiment is shown in fig 10. From fig 10 it is noted that the accuracy rate of rainfall forecasting is achieved as 93.9977% and the least Mean Squared Error (MSE) is achieved at epoch 4 with MSE value as 0.010584.

Fig 10: Graphical Representation of Result with 50% Training Dataset





The validation of the forecast accuracy of this model is also done through Heidke Skill Score, to compute HSS score of this forecast with 4 input parameters, 2 hidden layer neurons and with 50% training dataset, the result of the developed neural network is tabulated in Table 6. In Table 6, out of 15744 total records tested, 14799 records were matched and the Heidke Skill Score of this forecast computed for this experiment is 0.86302.

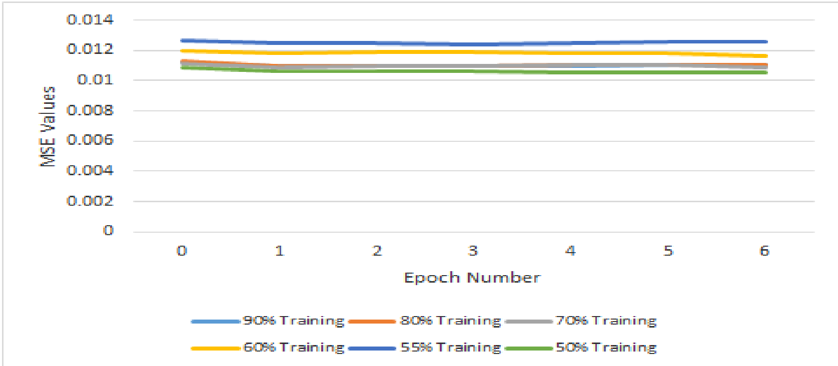
Table 6: Result with 4 Input Hourly Data, 2 Hidden Neurons, 60% Training Data

Rainfall Forecasted	Rainfall Occurred		
	Yes	No	∑Occurred
Yes	14799	851	15650
No	945	10359	11304
∑ Forecasted	15744	11210	26954

VI. Analysis of Results of Hourly Data Rainfall Forecasting

A comparison study was made on the forecasting results obtained from the experiments 1 to 6 which are implemented based on the hourly surface data (relative humidity, pressure, temperature and wind speed) using Backpropagation neural network model. The values of Mean Squared Error (MSE) at each epoch number in experiments 1 to 6, with varying size of training datasets were represented in the graph Fig 11.

Fig 11: MSE values (Hourly Data, 2 neurons with varying training size)



The Least Mean Squared Error and the corresponding accuracy percentage of hourly surface data rainfall forecasting obtained from the experiments 1 to 6 are shown in table 7. From the Table 7, it is observed that more accuracy i.e. 94.8197% is achieved when the training dataset is taken as 55% and testing data is taken as 45%.

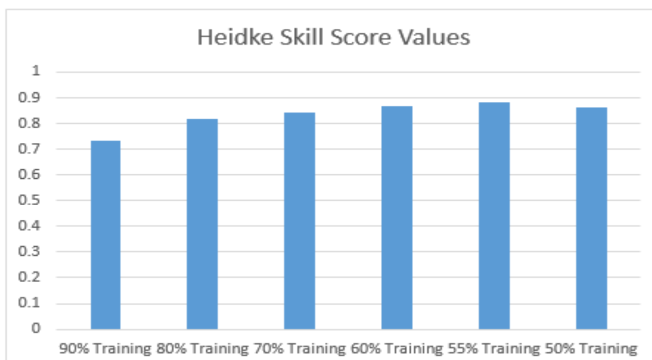
Table 7: Experiment Results with Hourly Data and 2 Neurons

Experiment No.	No. of Input Parameters	No. of Hidden Neurons	Training Dataset %	Mean Squared Error	Accuracy %
7	4	2	90%	0.010962	88.1512
8	4	2	80%	0.010939	91.9327
9	4	2	70%	0.010870	93.1823
10	4	2	60%	0.011659	94.3073
11	4	2	55%	0.012437	**94.8197**
12	4	2	50%	**0.010584**	93.9977

The Heidke Skill Score (HSS) were also computed for the experiments 1to 6 with varying size of training datasets were tabulated and the graph of HSS score at each experiment is shown in Fig 11.

Fig 11: Heidke Skill Scores for various sizes of training data

Training Data Size	HSS Value
90% Training	0.73244
80% Training	0.81666
70% Training	0.84463
60% Training	0.87008
55% Training	0.8816
50% Training	0.86302



From the Fig 11, it is observed that the developed neural network model for the forecasting of daily rainfall with hourly data of 4 input parameters (relative humidity, pressure, temperature and wind speed), 2 hidden layer neurons and with 55% of training dataset (in 5th experiment) has produced highest HSS score which validated and proved the accuracy of the result of this neural network forecasting model.

VII. Conclusion

In the present paper rainfall forecasting of hourly surface data using Back propagation Neural Network has been presented. The experiments and results show that the Back propagation Neural Network gives more accuracy and peak HSS score when the training data is 55% and the testing data is 45%. The Hourly surface data of rainfall produces much better accuracy in forecasting of rainfall rather than daily average surface data.

References

1. Riko Herwanto, Rosyana Fitria Purnomo and Sriyanto, *"Rainfall Prediction Using Data Mining Techniques"*, 3rd International Conference on Information Technology and Business, Dec 2017.
2. D. Nagesh Kumar, M. Janga Reddy and Rajib Maity, *"Regional Rainfall Forecasting using Large Scale Climate Teleconnections and Artificial Intelligence Techniques"*, Journal of Intelligent Systems, Vol. 16, No. 4, 2007.
3. Seema Mahajan and Dr. S.K. Vij, *"Modelling and Prediction of Rainfall Data using Data Mining"*, International Journal of Engineering Science and Technology, Vol 3, No. 7, July 2011.
4. Hafzullah Aksoy and Ahmad Dahamsheh, *"Artificial Neural Network Models for Forecasting Monthly Precipitation in Jordan"*, Stoch Environ Res Risk Access, Vol. 23, 2009.

5. S. Riad and J. Mania, L. Bouchaou, Y Najjar, *"Rainfall-Runoff Model Using an Artificial Neural Network"*, Mathematical and Computer Modelling, Vol. 40, 2004.
6. Norraseth Chantasut, Charoen Charoenjit and ChularatTanprasert, *"Predictive Mining of Rainfall Predictions Using Artificial Neural Networks for Chao Phraya River"*, 4th International Conference of the Asian Federation of Information Technology in Agriculture, Bangkok, Thailand, 2004.
7. Enireddy Vamsidhar, K.V.S.R.P. Varma, P. Sankara Rao and S. Ravikanth, *"Prediction of Rainfall using Backpropagation Neural Network Model"*, International Journal of Computer Science and Engineering, Vol. 02, No. 04, 2010.