Volume 7, Issue 4, April 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Forecasting of Rain Fall Prediction Using PCA and Hybrid Neuro Fuzzy Classification System

S. Kumaresan, S. Omprakash M.sc, M.phil
Department of Computer Science, Kovai Kalaimhal Collage of Arts & Science, Coimbatore, Tamilnadu, India

Abstract: Forecast is being disputed in a large vicinity of rules in day these days lifestyles. The weather measures like maximum and minimal temperature, moisture are predicted the use of the features extracted over numerous intervals and additionally from the climate measure sequence of facts factors itself. The technique implemented here makes use of feed forward artificial neural networks. Earlier study of the previous techniques isn't capable of reflect the experimental inter-annual variability of rainfall. Obscure truth of the strategies to sea floor temperatures can be one some of the feasible reality for the modest implementation of these strategies to forecast seasonal rainfall disputes. To enhance forecasting the information units, synthetic neural network (Ann) with the combination of fuzzy common sense as proposed technique in this look at. The proposed technique intention is to use combination of artificial neural community with fuzzy common sense to forecasts rainfall in south India. This proposed approach as compared to other strategies fuzzy good judgment based is locating prediction stage is excessive. Fuzzy common sense in comparison to present technique accuracy, time period additionally very green.

Keywords: Data mining, rainfall in south India, ANN, fuzzy logic North east monsoon, forecast of rainfall.

I. INTRODUCTION

Information mining [1] is the giant reference of absolute, formerly unknown, and possibly useful data from data. Forecasting is one of the foremost elements which could prevent lives and damage which by preventative arrangements and additionally facilitates the farmers to plan for flora. In this paper, we investigate in predicting the context of climate and climatology.

a massive wide variety of people take benefit of weather forecasts each day and often make decisions based totally on the anticipated weather situation. This is specifically critical for excessive events like storms or droughts: in these instances, forecasts are one of the main elements that could help save lives or prevent damage which might be avoided by means of preventative arrangements. Moreover, in many sectors low in cost fulfillment based at the climate: prediction is a crucial factor for business planning.

Rainfall is possibly the maximum vital variable in the phenomenon of monsoon. The quantity of rainfall in a selected day, week, month or season vary from each year over a huge variety which raises the query: is there a recognizable version in these versions, or is the range in basic terms threat or chance. Variability can be described as a motion of rainfall to ebb and waft or oscillate around ordinary dimensions. it effects that one can take this modifications on several time scales like days, weeks and months, and also on elements like stations, districts, towns , villages and coastal regions.

Monsoon is fact to be regulated associated with space on a big scale and is decided in time for wide variety of months, it could be beneficial to research the facts on a few top values. Meanwhile, the top time and area dimension for rainfall are undistinguished and as a result one has to agree the data as they may be and forecast the lifestyles of fashions. in the proposed studies, this is undertaken for the rainfall facts of Andhra Pradesh. a diffusion of statistical analyses of rainfall at the monthly dimensions, had been made already via previous investigators and with the statistics on the suggest, general deviation, coefficient of version additionally exists.

The northeast monsoon season occurs among October, November and December is the time of major rainfall activity over south India, particularly in andhra pradesh, tamil nadu and Pondicherry. The ne monsoon (nem) season purpose nearly 50% of annual rainfall within the east coast of Indian peninsula. Giant dying and belongings loss are reported each 12 months inside the visible coastal states of orissa, andhra pradesh, and Tamil nadu.

II. BACKGROUND STUDY

Rainfall play a vital role in water resource management planning, agriculture, economical planning and therefore, various types of patterns with various angle of difficulties have been developed for this purpose. The amount of rainfall [6] differs each and every year based on various aspects like location, environment, temperature, monsoon failure and unexpected storms. Rainfall forecast is vital for an area where collection of water or rainfall had been carried out, exactly for flood alert methods. The earlier research or existing system based on statistical systems for wide range

forecasts of Indian monsoon rainfall has outcome precise application, for past years. Method for combining long-established and scientific weather forecast methods would give better idea of reservation and drawbacks to the agriculture and also to make preventive measure to save lives and properties.

III. METHODOLOGY

Climate measures like maximum & minimal temperature, expectation and amount of rainfall, cyclone and its directions, are predicted the usage of meteorological satellites pictures to degree upcoming actions. The satellite-based systems are more expensive and need complete help. Those types of techniques are capable of enabling most effective such facts, which may be possible handiest with an extensive geographical vicinity. Frequent versions in climate patterns are authorized than preceding structures. Exactness in forecasting rainfall can be performed only through strengthening the information base of diverse existing systems for improving development in diverse packages. Considered one of its strength is its effective coping with and manipulation of massive records sets and their character graphical illustration, similarly it permits several possibilities for the enter and output of outside records files.

ANN:

The regular technique of a neural network [8] includes ‘‘training’’ the community with a giant instances of common information. The network has specific amount of enter and output ‘‘nodes’’ characterize the analyst and count on and variables, respectively (fig.3). The education degree of the neural community is established to decorate the weights so that the imply-squared mistakes of the output are decreased. for each node at a particular layer, the input node factors from the preceding layer are extended by using the burden of the hyperlinks between the nodes after which common of all of the specific connections are calculated to make the value at that node. This method is repeated for all nodes and repeated for every layer [4]. the community then can be used to make estimations depend upon sparkling input values.

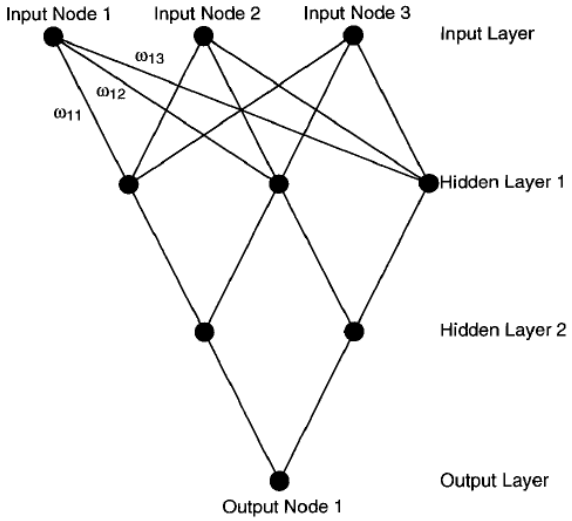


Fig 2. Schematic of a neural network with input nodes

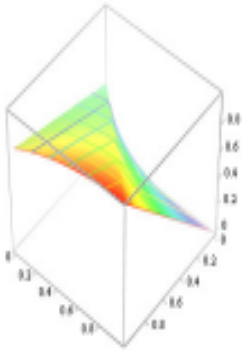
Fuzzy Logic:

the proposed fuzzy type set of rules builds at the extended Boolean ir version, that is primarily based on the fuzzy set idea and fuzzy good judgment [3], [4]. Inside the extended Boolean ir version, files are interpreted as fuzzy units of indexed terms. In each document, every listed term has a weight from the range [0, 1] expressing the diploma of significance of the term for report illustration. Many extraordinary weighting processes can be used to assign weights to index terms, e.g. the tf idft term weighting scheme [5]. The entire report collection can be represented by using a real valued index matrix d, in which every row di represents i−th document and value dij j−th term in i−th file. the query language inside the extended Boolean model of ir is stepped forward with the opportunity of weighting question terms which will characteristic unique levels of importance to those in a search request and with the aid of weighting (parameter zing) aggregation operators (most usually and, or, and now not) to melt or blur their impact on question assessment [3], [4]. bear in mind q to be the set of person queries over a set; then the load of time period t in question q is denoted as a(q, t) pleasant a : $q \times t \rightarrow [0, 1]$. To assess the atomic question of one time period representing single seek criterion the characteristic g: $[0, 1] \times [0, 1] \rightarrow [0, 1]$ might be used. The value of g(f(d, t), a) is referred to as the retrieval status price (rsv). For RSV assessment the interpretation of the query time period weight a is important.

The most usually used interpretations see the question time period weight because the significance weight, threshold or perfect record description [3], [4]. The theory for the evaluation of rsv inside the case of threshold interpretation is shown and (1) respectively [3], [4], in which p (a) and q(a) are coefficients used for tuning the brink curve. An example of p (a) and q (a) can be as follows: p (a) = 1+a 2 and q(a) = 1−a 2 4 . The RSV formula from (1) is illustrated in fig. 1. adopting the threshold interpretation, an atomic query containing time period t of the burden a is a request to retrieve documents having f(d, t) same or more to a. documents satisfying this circumstance can be rated with excessive rsv and contrariwise documents having f(d, t) smaller than a can be rated with a small rsv..

$$g(F(d,t),a) = \begin{cases} P(a)\frac{F(d,t)}{a} & \text{for } F(d,t) < a \\ P(a) + Q(a)\frac{F(d,t)-a}{1-a} & \text{for } F(d,t) \geq a \end{cases} \quad (1)$$

The operators AND, OR, and NOT can be evaluated with the,



Assist of fuzzy set operations. Fuzzy set operations are extensions of crisp set operations on fuzzy sets [6]. A feature function uniquely defines a fuzzy set and hence fuzzy set operations are defined the usage of function capabilities [7]. In [6] l. zadeh defined simple strategies for the complement, union and intersection of fuzzy sets. Subsequent to this preferred (zadeh's) fuzzy set operations, whole classes of prescriptions for outlining the complements, intersections and unions on fuzzy units were later designed [8]. In this examine, we use the brink interpretation of RSV and fashionable t-norm (2) and t-conform (three) for the implementation of and, or operators and fuzzy supplement for the assessment of no longer operator (four). c(x) = 1 − x (2) t(x, y) = min(x, y) (three) s(x, y) = max(x, y) (4) but, using other commonplace t-norm and t-conorm pairs is possible. An ir evaluation the effectiveness of an information retrieval machine can be evaluated the use of the measures precision p and do not forget r. precision corresponds to the chance of retrieved file to be relevant and don't forget may be seen as the probability of retrieving applicable document. Precision and keep in mind within the prolonged Boolean ir model can be defined using the σ−count number kak [9]:

$$\rho(X|Y) = \begin{cases} \frac{\|X \cap Y\|}{\|Y\|} & \|Y\| \neq 0 \\ 1 & \|Y\| = 0 \end{cases} \quad (5)$$

$$P = \rho(REL|RET) \qquad R = \rho(RET|REL) \quad (6)$$

For an easier IR effectiveness evaluation, measures combining precision and recall into one scalar value were developed. The F-score F is among the most used scalar combinations of P and R: $F = \frac{(1+\beta^2)PR}{\beta^2 P + R}$

The index matrix D can be seen as a general data matrix with m rows (data samples) and n columns (data features). The evaluation of Extended Boolean query over the document collection generates an ordering of the documents (i.e. it assigns a real value from the range [0, 1] to each document). The ordering can be also interpreted as a fuzzy set of documents. If we abandon the IR terminology, we can call the extended Boolean query a general fuzzy classifier and use it to describe fuzzy sets or fuzzy sub sets of data by its features.

The evolution of fuzzy classifiers for data mining utilizes genetic programming. In this section, we provide brief introduction into the area of evolutionary computing and genetic programming in particular. Moreover, we describe the application of genetic programming to evolutionary query optimization. A. Genetic algorithms and genetic programming Genetic algorithms are a popular member of the wide chapter of evolutionary algorithms. They are based on the programmatic implementation of genetic evolution and they emphasize selection and crossover as the most important operations in the evolutionary optimization process [10], [11]. Genetic algorithms evolve a population of chromosomes representing potential problem solutions encoded into suitable data structures. The evolution is performed by genetic operators modifying the chromosomes, i.e. the encoded forms of problem solutions. Genetic programming (GP) is an extension to genetic algorithms, allowing work with hierarchical, often tree-like, chromosomes with an unlimited length [10], [12]. In GP, the chromosomes take the form of hierarchical variably sized expressions, point-labeled structure trees. The trees are constructed from nodes of two types, terminals and functions.

The chromosomes are evaluated by the recursive execution of instructions corresponding to tree nodes [13]. Terminal nodes are evaluated directly (e.g. by reading an input variable) and functions are evaluated after left-to-right depth-first evaluation of their parameters. Genetic operators are applied to the nodes in tree-shaped chromosomes. A crossover operator is implemented as the mutual exchange of randomly selected sub-trees of the parent chromosomes. Mutation has to modify the chromosomes by pseudo-random arbitrary changes in order to prevent premature convergence and broaden the coverage of the fitness landscape. Mutation could be implemented as: i) removal of a sub-tree at a randomly chosen node ii) replacement of a randomly chosen node by a newly generated sub tree iii) replacement of node instruction by a compatible node instruction (i.e. a terminal can be replaced by another terminal, a function can be replaced by another function of the same arity) iv) a combination of the above Genetic programming facilitates the efficient evolution of symbolic expressions, even whole computer programs. In this work, we use genetic programming for fuzzy classifier optimization.

B. Evolutionary query optimization Genetic programming has been recently used for the optimization of extended Boolean queries [14], [15]. It was shown that genetic programming was able to optimize search queries so that they described a set of relevant documents. In the fuzzy information retrieval model, the relevant documents formed a fuzzy subset of the set of all documents and the extended Boolean queries were evolved to describe them. An information

retrieval system based on the extended Boolean IR model was implemented to validate evolutionary query optimization. The t idft term statistics [5] were used for document indexing and query weights (RSV) were evaluated using (1). The query language in the IRS supported the standard Boolean operators AND, OR, and NOT.

The information retrieval system served as a test bed for evolutionary query optimization and allowed genetic programming over extended Boolean queries. The GP evolved tree representations of search queries with Boolean operators as function nodes and terms as leaves. Both operator nodes and term notes were weighted. In order to generate a random initial population for the GP, the system was able to generate random queries. The particular settings of the random query generator showing the probabilities of generating a particular query node are summarized in Table Ia. The implementation of a crossover operator for GP is straightforward. In the experimental information retrieval system, it was implemented as a mutual exchange of two randomly selected branches of parent tree chromosomes. The mutation operator in query GP aims to perturb the content and structure of the chromosomes randomly. In our implementation, it selects a node from the processed chromosome at random and performs one of the mutation operations summarized in Table Ib. The query mutation types that were implemented included: i) change of selected node weight. ii) Replacement of selected node type by a compatible node type (i.e. operator OR replace by operator AND, term replaced by another term). iii) Insertion of NOT operator before selected node. iv) Removal of NOT operator if selected. v) Replacement of selected node by a randomly generated branch. The IR measure F-Score (7) was used as a fitness function. This study extends the framework for genetic evolution of extended Boolean queries to the evolution of general fuzzy classification rules. For test each data set, a fuzzy rule that would describe known classes can be found. Rules found using such a supervised learning procedure can be subsequently used to classify new data samples. The learning of a classifier can be long time process (depending on the dimension of the data). On the other hand the classification (i.e. evaluation of the classifier) of each data sample is very fast and can be used also in real time.

District	Year	January	February	March	April	May	June	July	August	September	October	November	December	Annual Total
Khammam	2009	0	0	10.1	2.8	68.3	36.4	185.2	172.1	144.2	81.4	33.6	0	734.1
Khammam	2010	29.5	4.7	2.2	23.8	134.2	153.6	615	304.6	303.6	120	67.5	45	1803.7
Krishna	2009	0	0	0.8	0	105.1	36.1	87.2	175.7	154.8	75.6	99.1	1.7	736.1
Krishna	2010	16.6	0	0	20.1	202.3	131.7	373.1	296.2	318.1	201.8	152.8	92.5	1805.2
Kurnool	2009	0.1	0	12.8	0	67.5	88.6	24.7	88.5	227.3	197.2	57.8	3.4	767.9
Kurnool	2010	4.1	0	0	18.8	41.4	91.7	194	208.5	124.5	80.6	48.2	11.4	823.2
Medak	2009	0	0	0	6.6	21	52.3	64.3	157.6	142.7	41.2	13.8	4.6	504.1
Medak	2010	9.7	13.5	0	5.5	5.3	83.8	337	269.9	199.6	95.2	18.3	0.6	1038.4
Nellore	2009	0.1	0	0.1	0	28.5	33.2	40.3	111.6	66.9	49.6	465.7	109.6	905.6
Nellore	2010	4.9	0	0	17.8	108.3	112.5	116.3	203.7	109.4	244.7	343.9	154.7	1416.2
Vizamabad	2009	0	0	0.1	9.9	15.7	78.7	173.9	188.3	155.5	48	15.8	3.8	689.7
Vizamabad	2010	8.4	30.3	0.9	7.3	0.5	104.1	299.2	303.8	245.4	157.5	34.9	0.4	1192.7
Prakasam	2009	1.9	0	4.5	3.7	40.6	49.3	22.6	98.8	126.5	50.3	216.6	20.3	635.1
Prakasam	2010	17.7	2	4.3	1	265.4	106.9	170.3	205.1	131.7	163.4	227.1	108.1	1403
angareddi	2009	0	0	3.3	32.9	20.2	89.9	67	299.8	151.5	77.1	34.7	2.4	778.8
angareddi	2010	9.6	3.2	0	13.9	21.2	162.2	272.4	286.6	205.4	99.4	31.7	4.8	1110.4
rikakulam	2009	0	0	10.2	7.2	26	99.3	215.9	162.5	145.4	122.6	85.2	0	874.3
rikakulam	2010	0.4	16.4	17.5	3.1	172.7	142.1	236.8	263	207.9	333	178.6	129	1700.5
nakhapatn	2009	0	0	6.3	6.6	24.3	94.9	132.7	92	171.6	116.9	89	0	734.3
nakhapatn	2010	13.6	0.1	2.1	27.5	173.1	156.3	321	167	188	203.5	287.4	163	1702.6
zianagara	2009	0	0	13.3	11.3	50.5	145.1	182.1	242.9	144.2	125	118	0	1032.4
zianagara	2010	46.2	9.6	14.8	5.5	128.9	179.9	293	234	245.9	189.6	167.1	133.4	1647.9
Warangal	2009	0	0	0	0.9	22	50.6	123.7	149.2	180.4	81.1	46.9	0.4	655.2
Warangal	2010	7.8	5.7	0	0	56.2	119.5	499.3	223.7	280.3	99.7	39.2	21.9	1353.3

Fig 3. Rainfall Dataset

Training of the network

The proposed system has more than one hidden layers, where calculated points are hidden units. These properties are mined from sample weather data set which is given as input and the result is taken as the parameter to be predicted during training of the dataset.

Ann with fuzzy logic in this experiment was trained and simulated using Mat lab 2013 version. The input dataset holds dataset that corresponds to Indian state. Before starting the training, the inputs and results have been measured so that they exist between specific ranges.

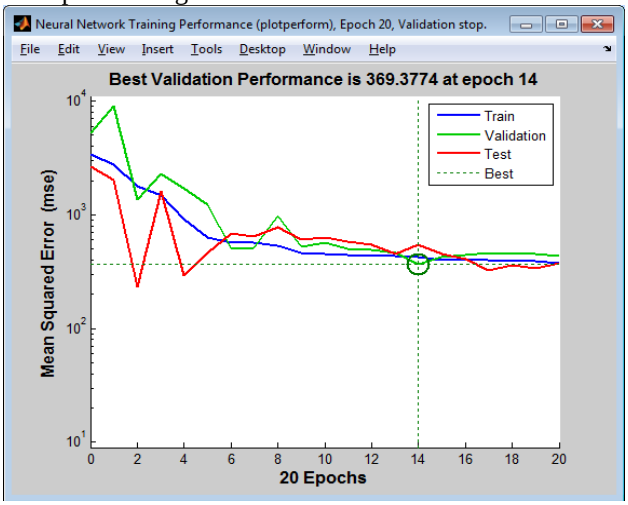


Fig 4. Performance of data

IV. EXPERIMENT RESULTS

The NE monsoon period is the major term of rainfall occurrences various regions of south Agricultural endeavor of the parts is extremely reliable on the extent of rainfall received in every season.

The experimental results are presented to establish the contribution of each factor used to optimize the Ann by using fuzzy logic method. It has been empirical that many permutations may offer with similar validation and accuracy and the final sample can be trained by any one among them. The results in all the data sets prove that the Fuzzy Logic - Ann can achieve 95.8% accuracy. The experimental study shows that the Fuzzy Logic -Ann model is very effective in terms of both evaluation time and estimating performance.

With Fuzzy Logic and Ann [5] is very simple and efficient way to find out the optimality condition with multiple responses and its predicted value. An Ann model predicted the output response as a function of cutting parameters. The approach has been proved to be successful with experimental results with less relative errors which plot it is explored that points are randomly and scattered and therefore the estimated data are significantly fit to the approach.

This experiment has led the approach in integrating the efforts of both original and scientific weather forecast approaches in improving the chances towards their application in various domains.

Ann layer:

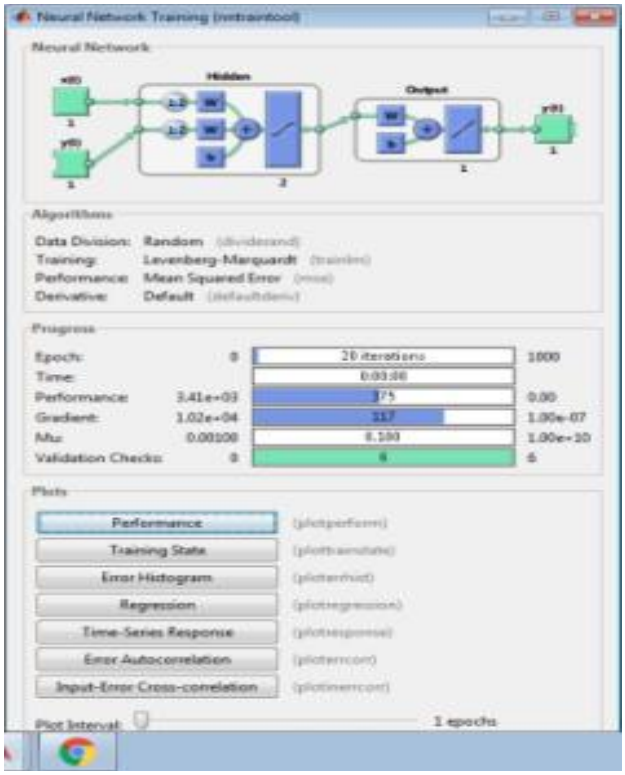


Figure5. Neural net work basic layout

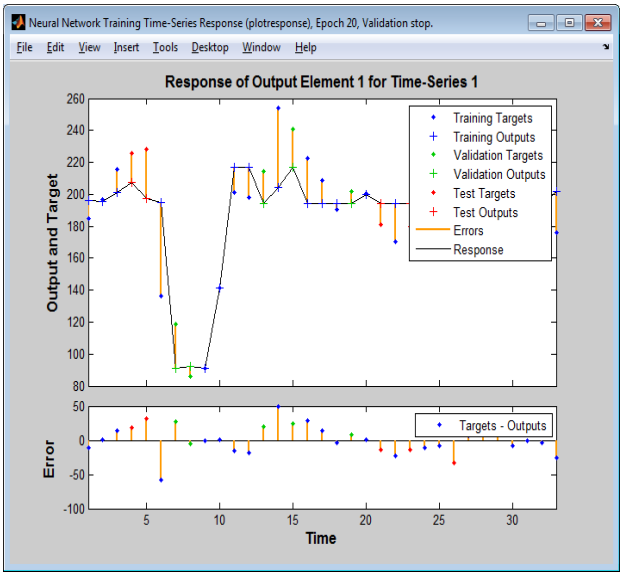


Figure 6 rain Data prediction using ann method

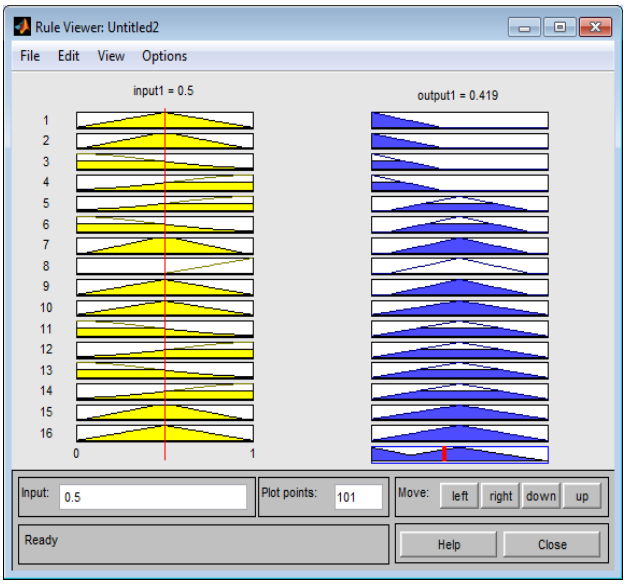


Figure 7. Fuzzy rule mining based data classification methods

The proposed approach Fuzzy logic with artificial neural network has given really a good forecast of both the probability and amount of rainfall.

This Ann combines with Fuzzy logic with important features of self-generation, constraint enhancement, and rule-base popularization and here used the Ann & Fuzzy Logic to popular rainfall data sets.

V. CONCLUSION

Artificial Neural Networks with the integration of Fuzzy logic (Fuzzy Logic -Ann) is able to forecast the weather. Selected mathematical displays are of able to mine the movements, which can be taken into account as phenomena for the enhancement of system. Arithmetic / mathematical displays except measurement of unparallel are found apt to mining the hidden patterns exist in weather dataset. The neural network signal processing and fuzzy logic method is capable of providing better results in forecasting and it is taken into account as a substitute to usual meteorological approaches.

It is concluded that fuzzy logic -Ann has proven favorable outputs and is very apt for giving solution to the problem of forecasting rainfall. With the input parameters as network location, the Ann method has been trained to predict Rainfall. This experimental study has come out with that Data Mining approaches deployed could help in providing enhancing data for prediction.

REFERENCE

[1] J. Han and M. Kamber, Data Mining Concepts and Techniques. Morgan Kaufmann, 2001.

[2] Singh.C.V, 2006. "Fuzzy logic of satellite-observed outgoing long-wave radiation during the monsoon period (June-September) over India". Theor.Appl.Climatol.84, 207-211.

[3] Chattopadhyay S.(2007). Multilayered feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India, *Journal Acta Geophysical*, Vol. 55, No.3, 2007, pp. 369-382.

[4] K Poorani, Data Mining Based on Fuzzy logic for Rainfall Forecasting in India, September 2013

[5] Zhang, Y., Li, H., Hou, A., Havel, J., 2006. Artificial neural networks based on fuzzy logic input selection for quantification in overlapped capillary electrophoresis peaks. Chemo metrics and Intelligent Laboratory Systems 82, 165–175.

[6] Ashwini Kulkarni, S S Sabade And R H Kripalani July 2006 Intra-seasonal Vagaries of the Indian Summer Monsoon Rainfall.

[7] Rajeevan M., Bhate J.(2009). A high resolution daily gridded rainfall dataset (1971–2005) for musicale meteorological studies, *Current Science*, Vol. 96, No. 4, February 2009.

[8] Paras, Mathur S., Kumar A., Chandra M. (2007). A Feature Based Neural Network Model for Weather Forecasting. *World Academy of Science, Engineering and Technology,* Vol. 34, 2007, pp. 66-73.

[9] M. Rajeevan and Jyoti Bhate A high resolution daily gridded rainfall dataset (1971–2005) for mesoscale meteorological studies February 2009.

[10] Dr. B. Sudhakar Reddy Regional Variations in Agricultural Development of Andhra Pradesh – A Factor Analysis, Dec 2013.

[11] Mitra, A.K., Nath, Sankar, Sharma, A.K., 2008. Fog forecasting using rule-based fuzzy inference system. Journal of the Indian Society of Remote Sensing 36, 243–253.

[12] Ozkok, Y. 2005, Web Based Ionosphere Forecasting Using Neural Network and Neurofuzzy Models, MS Thesis, supervisor: Tulunay, E., osupervisor: Tulunay, Y., Dept. of Electrical and Electronics Eng., Middle East

Technical University, Ankara, Turkey.Pocernich, Matthew, 2010. Verification package; examples using weather forecasts.

[13] Rich, E., Knight, K., 1996. Artificial Intelligence. McGraw-Hill, New York. Shao, J., 2000. Fuzzy categorization of weather conditions for thermal mapping. J. Applied Meteorology 39, 1784–1790.

[14] Shu, C., Ouarda, T.B.M.J., 2008. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. Journal of Hydrology 349, 31–43.

[15] Sivakumar, B., 2000. Chaos theory in hydrology: important issues and interpretations. Journal of Hydrology 227, 1–20.

[16] Sivakumar, B., 2001. Rainfall dynamics in different temporal scales: a chaotic perspective. Hydrology and Earth System Sciences 5, 645–651.

[17] Sivakumar, B., Liong, S.Y., Liaw, C.Y., Phoon, K.K., 1999. Singapore rainfall Behavior: chaotic? J. Hydrol. Eng., ASCE. 4, 38–48.

[18] Sivanandam, S.N., Sumathi, S., Deepa, S.N., 2007. Introduction to Fuzzy Logic using MATLAB. Springer, pp. 113–145.

[19] Suwardi, A., Takenori, K., Shuhei, K., 2006.Neuro-fuzzy approaches for modeling the wet season tropical rainfall. J. Agric. Inforam. Res. 15, 331–334.

[20] Wilks, D.S., 1999. Multisite generalization of a daily stochastic precipitation generation model. Journal of Hydrology 210, 178–191.

[21] Wong, K.W., Wong, P.M., Gedeon, T.D., Fung, C.C., 2003. Rainfall prediction model using soft computing technique. Soft Comput. Fusion Found. Method. Appl. 7, 434–438.

[22] Zacharov, Petr, Rezacova, Daniela, 2009. Using the fraction skill score to assess the relationship between an ensemble QPF spread and skill.

