A Novel Decision Tree Approach for the Prediction of Precipitation using Entropy in SLIQ

Narasimha Prasad L V	Prudhvi Kumar Reddy K	Naidu M M
Vardhaman College of Engineering	Vardhaman College of Engineering	Vignan University, Vadlamudi
Hyderabad, India	Hyderabad, India	Guntur, India
e-mail: lvnprasad@yahoo.com	e-mail: cience.future4u@gmail.com	e-mail: mmnaidu@yahoo.com

***Abstract* — The rising tendency of population of every nation is one of the severe stumbling blocks to arrest its economic growth, particularly in the third world countries like India, not able to address even the basic needs of its people. It is high time to have introspection for the deficiencies and find a remedy. The major basic need is food, a product of agriculture. Agriculture mainly depends on rainfall. Prediction of precipitation is a complex phenomenon. Till now many of the researchers have tried their best for predicting the precipitation but in vain, since the prediction is quite complex with neural networks, back propagation, fuzzy logic etc. Hence, we found that data mining is an emerging, efficient, easily implementable tool, which predicts the useful patterns for the prediction of rainfall in a very short time. Supervised Learning In Quest, an efficient data mining decision tree algorithm is applied in the prediction of precipitation. The present research illustrates the Supervised Learning In Quest decision tree algorithm using entropy, which estimates the prediction of precipitation with an average accuracy of 74.92% and the knowledge extraction is purely based on historical data.**

Keywords - Data mining, Decision tree, Meteorology, Precipitation, Prediction, Rainfall, SLIQ, Soft Computing.

I. INTRODUCTION

Since the dawn of modern era, prediction of precipitation has been one of the most interesting, complex, challenging and exploring domain, because rain is the fundamental source of agriculture, which is in turn a major economy of developing and under developed nations across the vast universe.

Now a day's, the cultivation of crops has turned rather a difficult process as the lands are becoming less cultivatable due to scarcity of rains and deforestation prevails. The prediction of precipitation plays a vital role, as it predicts whether there is a rain or no rain. So, the crux of the problem lies in rains. Rainfall is due to thick layers of clouds in the atmosphere that would have attained melting point [26].

In order to predict the precipitation, many techniques are applied till date like back propagation model of neural network, linear regression, Bayesian networks, genetic algorithm, fuzzy logic etc. which are not on to their true extent in giving the accurate predictions because the accuracy rate is very less. In the continuous research it is found that data mining provides with accurate results of prediction, which is the primary motivation behind and the approach used for predicting the precipitation is SLIQ (Supervised Learning in Quest) [1, 4] decision tree algorithm.

SLIQ uses entropy as its attribute selection measure which provides the high accuracy. SLIQ algorithm can be applied for huge data, in diminutive time. The success rate for the prediction of the precipitation by employing different data mining tools reported in the literature is 43.6% [29]. Recently, Prasad et. al proposed to employ Supervised Learning In Quest (SLIQ) decision tree using Gini index for the prediction of the precipitation which resulted in an accuracy of 72.3% [2]. This paper proposes to employ SLIQ decision tree using entropy that improves the accuracy from 72.3% to 74.92%.

The rest of the paper is illustrated as follows: Section II describes relevant work. Section III describes briefly regarding decision trees. In Section IV, Gausses SLIQ decision tree algorithm. Section V describes the procedure for entropy based SLIQ decision tree algorithm and also the procedure for selection of root node. Section VI illustrates the performance measures of the SLIQ decision tree algorithm. Section VII displays the results of the model and finally section VIII concludes the paper.

II. RELEVANT WORK

In the literature, there are many research findings which are reported for predicting the precipitation with accurate possible rate. Some of them used the traditional methods of the artificial neural networks for the prediction while other methods include the recent developments like Image Processing, Linear Regression and Fuzzy logic and so on.

Frank Silvio Marzano, Giancarlo Rivolta, Erika Coppola, Barbara Tomassetti and Marco Verdecchia used a fully neural network approach to the rainfall field Nowcasting from infrared and microwave passive-sensor imagery aboard [6]. K.Richards and G.D. Sullivan, combined the features of Bayesian scheme for texture analysis of the cloud images which are taken from the ground [7]. C. Jareanpon, W. Pensuwon, R.J. Frank and N. Davey formed radial basis function neural network with a specially designed genetic algorithm [8]. K. Ochiai, H. Suzuki, S. Suzuki, N. Sonehara and Y. Tokunaga stated that the computational time for learning with an acceleration algorithm can be reduced about 10 percent by introducing a pruning algorithm [9]. I.F.

Grimes, E. Coppola, M. Verdecchia and G. Visconti presented an approach to cold cloud duration imagery derived from meteosat thermal infrared imagery is used in conjunction with numerical weather model analysis data as an input to an ANN [10]. Thiago N. de Castro, Francisco Souza, Jose M.B. Alves, Ricardo S.T. Pontes, Mosefran B.M. Firmino and Thiago M. de Pereria forecasted seasonal Rainfall using Neo-Fuzzy neuron model [11]. Tuan Zea Tan, Gary Kee Khoon Lee, Shie-Yui Liong, Tian Kuay Lim, Jiawei Chu and Terence Hung IEEE treated the series of rainfall as a continuous time series [12]. Jiansheng Wu Integrated linear regression with ANN. The linear regression extracts linear characteristics of the rainfall [13]. Hui Qi, Ming Zhang and Roderick A. Scofield developed a Multi-Polynomial High Order Neural Network (M-PHONN) [14]. Wint Thida Zaw and Thinn Thu Naing stated that the Multi variables polynomial regression (MPR) is one of the statistical regression methods used to describe the complex nonlinear input and output relationships [15]. C. Kidd and V. Levizzani stated that the rainfall is spatially and temporally highly variable [16]. Sanjay D. Sawaitul, Prof. K.P. Wagh and Dr. P.N. Chatur used the parameters of the weather like wind direction, wind speed, humidity, rainfall and temperature and so on for the classification and prediction of the future weather by using the back propagation algorithm [17]. Soroosh Sorooshian, Kuo-lin Hsu, Bisher Imam and Yang Hong made global precipitation estimation from satellite image by using artificial neural networks [18]. Kesheng Lu and Lingzhi Wang used a bagging sampling technique is used to generate the training sets for combination model based on support vector machine for the rainfall prediction [19]. Grant W. Petty and Witold F. Krajewski discussed in their research methods based on infrared, visible and passive microwave radiation measurements [20].

III. Decision Tree

Decision tree extracts the useful patterns from huge dataset. There exists a particular pattern among the attributes in predicting the class label. It establishes relationship between the various datasets by discovering the hidden patterns among the datasets which are huge and complex [3-5], [27]. Hence decision trees are used for knowledge discovery by excavating and establishing the hidden patterns among the attributes in a dataset.

For prediction of precipitation many approaches like back propagation model of neural network [24, 25], [29-32], linear regression [33], fuzzy logic, linear discriminant, Bayesian networks and so on when implemented provides less accuracy consuming huge amount of time for learning. So, in the present research decision trees are used which provides the maximum accuracy with less learning time and scalability. For prediction of precipitation accuracy and scalability plays a significant role.

Decision trees are constructed with nodes, comprises of root node and child nodes, and the nodes are selected based on the entropy which is explained in later sections. It generates rules which are perfectly comprehensible. As the data size becomes huge the size of decision tree increases and becomes quite complex to understand. The illustration regarding SLIQ decision tree algorithm is described in Section IV.

IV. SLIQ Decission Tree Algorithm

SLIQ is used for prediction of data and it is developed at IBM Almaden Research Center, increases the accuracy in prediction with maximum peak. SLIQ can handle both numeric and categorical attributes, shown in Table I.

The attributes considered for prediction of precipitation are humidity, temperature, pressure, wind speed, dew point. The amount of water vapor in the air is referred as humidity and is invisible in nature. Winds and rainfall are the major causes for humidity. Humidity mainly depends on condensation and water vapor that exists in air. Humidity when combines with hot temperature, the weather becomes more dangerous. Temperature is a property in precipitation which expresses matter in hot stage and cold stage. When the temperature is minimum it is referred as cold and when the temperature is maximum it is referred as hot stage. Temperature is measured in terms of ^{o}C. Pressure is the force per unit area exerted against a land surface by the weight of air above the land surface and is measured by using barometer and is measured in atmospheres. The amount at which wind is flowing is referred as wind speed and mainly affects the weather forecasting, aircrafts, marines and etc. wind speed is measured by using

TABLE I Training Dataset

Day	Humidity (H)	Temperature (T)	Pressure (P)	Wind Speed (W)	Dew Point (D)	Class (C)
1	97	24	1005	14	21	Rain
2	85	26	1004	16	21	No Rain
3	91	27	1004	14	21	Rain
4	82	27	1006	16	20	Rain
5	81	26	1007	18	19	No Rain
6	95	26	1007	18	20	Rain
7	95	26	1007	16	20	Rain
8	93	26	1008	18	21	Rain
9	87	24	1005	13	21	Rain
10	88	24	1005	11	21	Rain
11	80	26	1005	14	21	Rain
12	89	26	1005	14	21	Rain
13	86	27	1006	14	21	No Rain
14	86	28	1007	10	22	Rain
15	94	27	1006	14	21	Rain
16	88	26	1004	13	21	No Rain
17	92	27	1005	13	21	Rain
18	86	27	1007	11	21	Rain
19	82	27	1006	11	21	Rain
20	76	27	1007	14	19	No Rain
21	79	27	1008	11	20	No Rain
22	75	27	1008	13	20	No Rain
23	84	27	1007	13	20	No Rain
24	88	26	1006	11	21	Rain
25	86	25	1005	16	19	Rain
26	78	28	1006	13	21	No Rain
27	79	27	1008	13	19	No Rain
28	80	28	1008	8	20	No Rain
29	84	29	1009	6	21	No Rain
30	76	27	1009	6	22	Rain

anemometer. Pressure gradient, Rossby waves and jet streams and local weather conditions mainly affect the wind speed which leads to destructions. The units of wind speed are meters per second. Dew point is the temperature at which the air present in the atmosphere can no longer hold all of the water vapor which is mixed with it and some of the water vapor must condense in to liquid water.

Initially in SLIQ, the data needs to be sorted at the tree growth phase [2], which decreases the cost. As the sorting process takes place at the initial stage and is not repeated at each node. Where as in other algorithms the sorting of the data is carried out at each and every node which increases the cost.

Entropy is generally used to measure the inequalities among the statistical data and its frequencies. So far, its use is permitted for the analysis of wealth and income of the economic countries. Due to the inequalities present in the probabilities, there may be some glaring. But, irrespective of its limitation present it has a wide variety of applications in statistical analysis.

Entropy is used here for the construction of decision tree where the roots and sub-roots are classified based on entropy. The use of entropy for the rainfall analysis is quite appropriate because of the irregularities present in the weather data of precipitation. The precipitation data does not follow an order, which may be due to the inequalities of the present attribute with former attribute. This may change to a great extent or to some extent depending on the Mother Nature.

The algorithm for the construction of SLIQ decision tree for the prediction of precipitation is presented below. The notations used are given in Table II.

Overview of SLIQ Decision tree growth and split points

1. Read dataset into the root node of the SLIQ decision tree
2. Generate an attribute list for each attribute of the dataset
3. Sort the attribute lists on attribute value in non-decreasing order
4. Compute the entropy for the root node

$$Entropy(D) = -\sum_{i=1}^{N} P_i \log_2 P_i \quad (1)$$

5. Compute the Info of attribute list 'V'

$$Info(V) = \sum_{j=1}^{N} P_j \left[-\sum_{i=1}^{N} P_i \log_2 P_i \right] \quad (2)$$

6. Compute the Gain for each attribute list

$$Gain\ (V) = Entropy\ (D) - Info\ (V) \quad (3)$$

7. Determine maximum gain from among the gain values which become the basis for the best split as shown in Table III.

$$Best\ Split = Max.\ Gain\ value\ of\ attribute \quad (6)$$

TABLE II NOTATIONS USED IN PRESENTING SLIQ ALOGRITHM

Symbols	Description
D	Set of training tuples with associated class labels
|D|=N	The number training tuples in D
D_J	The set of data tuples in D satisfying outcome J
C	The class label
Entropy (D)	The information needed to classify a tuple in D
Split point	Midpoint of V_i and V_{i+1}
V	An attribute list
V_i	Change in class label attribute V
P_i	The probability that a tuple in D belongs to class C_i
D_i	Values which are greater than or equal to the Split point
D_j	Values which are less than the Split point

8. Partition the root node into leaf nodes based on the best split point.
9. Repeat the steps 5 through 8 reading the root node as leaf node until all leaf nodes contain the same class labels.

The ideal goal is to pressure a compact and accurate minimizing its construction time and facilitating scalability.

The decision tree shown in Fig. 1 is constructed with a sample of 30-day training tuples given in Table 1 using SLIQ Decision Tree algorithm employing entropy for determining precipitation.

V. PERFORMANCE MEASURES AND RULES OF THE DECISION TREE

The primary metric for evaluating the prediction of precipitation is accuracy and scalability.

a. Accuracy

The accuracy of a predictor refers to how well a given predictor can gives the value of the predicted attribute for new or previously unseen data.

$$Accuracy = Correct\ predictions / Total\ predictions \quad (7)$$

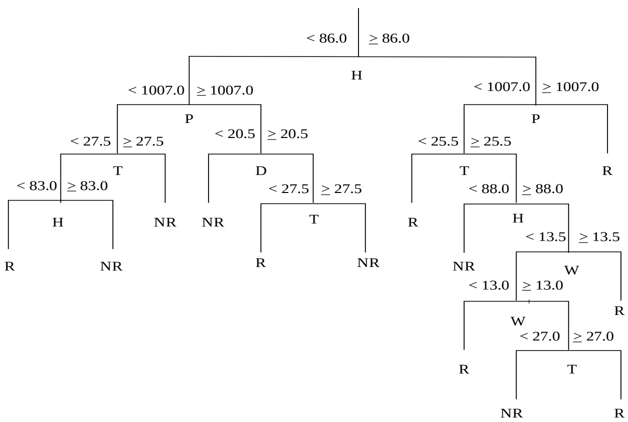

Fig. 1. Decision Tree based on Entropy

TABLE III. GAIN BASED SPLIT VALUE FOR VARIOUS ATTRIBUTES

Iteration	Humidity		Temperature		Pressure		Wind Speed		Dew Point	
	Split Value	Gain	Split Value	Gain	Split Value	Gain	Split Value	Gain	Split Value	Gain
Step 1	**86.0**	**0.2782**	25.5	0.1079	1007.5	0.0898	13.5	0.0348	20.5	0.0976
Step 2	83.0	0.1201	27.5	0.1201	**1007.0**	**0.1928**	17.0	0.0361	20.5	0.1239
Step 3	83.0	0.3219	**27.5**	**0.3219**	1006.0	0.0199	13.0	0.1709	20.5	0.1709
Step 4	**83.0**	**0.8112**	27.0	0.3112	1006.0	0.3112	15.0	0.3112	20.5	0.1225
Step 5	77.0	0.1971	27.5	0.0431	1008.5	0.28103	7.0	0.2811	**20.5**	**0.2811**
Step 6	83.0	1.0	**27.5**	**1.0**	1007.5	0	15.0	0	20.5	0
Step 7	88.0	0.0158	25.5	0.0561	**1007.0**	**0.0732**	15.0	0.0566	20.5	0.0403
Step 8	88.0	0.0382	**25.5**	**0.1347**	1006.0	0.03827	13.0	0.0587	20.5	0.0277
Step 9	**88.0**	**0.3059**	27.0	0.0059	1006.0	0.0059	13.0	0.0760	20.5	0
Step 10	83.0	0	27.0	0.1908	1006.0	0.1091	**13.5**	**0.1908**	20.5	0
Step 11	83.0	0	27.0	0.2516	1006.0	0.2516	**13.0**	**0.2516**	20.5	0
Step 12	83.0	0	**27.0**	**1.0**	1007.5	0	15.0	0	20.5	0

b. Scalability

This algorithm takes N input attributes and N number of classes as an input and produces the minimized decision tree.

c. Rules for Decision Tree

Once the decision tree is constructed, there is a possibility that the tree is very large to understand. Hence, to simplify the understanding of the large decision tree the rules are generated.

Rule 1: If [(humidity < 86.0) and (pressure < 1007.0) and (temperature < 27.5) and (humidity < 83.0)] Then (Prediction = Rain)

Rule 2: If [(humidity < 86.0) and (pressure < 1007.0) and (temperature < 27.5) and (humidity >= 83.0)] Then (Prediction = NoRain)

Rule 3: If [(humidity < 86.0) and (pressure < 1007.0) and (temperature >= 27.5)] Then (Prediction = NoRain)

Rule 4: If [(humidity <86.0) and (pressure >=1007.0) and (dew-point <20.5)] Then (Prediction=NoRain)

Rule 5: If [(humidity < 86.0) and (pressure >= 1007.0) and (dew-point >= 20.5) and (temperature < 27.5)] Then (Prediction = Rain)

Rule 6: If [(humidity < 86.0) and (pressure >= 1007.0) and (dew-point >= 20.5) and (temperature >= 27.5)] Then (Prediction = NoRain)

Rule 7: If [(humidity >= 86.0) and (pressure < 1007.0) and (temperature < 25.5)] Then (Prediction = Rain)

Rule 8: If [(humidity >= 86.0) and (pressure < 1007.0) and (temperature >= 25.5) and (humidity < 88.0)] Then (Prediction = NoRain)

Rule 9: If [(humidity >= 86.0) and (pressure < 1007.0) and (temperature >= 25.5) and (humidity >= 88.0) and (wind-speed < 13.5) and (wind-speed < 13.0)] Then (Prediction = Rain)

Rule 10: If [(humidity >= 86.0) and (pressure < 1007.0) and (temperature >= 25.5) and (humidity >= 88.0) and (wind-speed < 13.5) and (wind-speed >= 13.0) and (temperature < 27.0)] Then (Prediction = NoRain)

Rule 11: If [(humidity >= 86.0) and (pressure < 1007.0) and (temperature >= 25.5) and (humidity >= 88.0) and (wind-speed < 13.5) and (wind-speed >= 13.0) and (temperature >= 27.0)] Then (Prediction = Rain)

Rule 12: If [(humidity >= 86.0) and (pressure < 1007.0) and (temperature >= 25.5) and (humidity >= 88.0) and (wind-speed >= 13.5)] Then (Prediction=Rain)

Rule 13: If [(humidity >= 86.0) and (pressure >= 1007.0)] Then (Prediction = Rain)

VI. EXPERIMENTAL RESULTS

Decision Tree is a conceptual data mining method where the data is partitioned based on nodal values called split points. The split points are evaluated in SLIQ using various approaches like gini index, entropy and gain ratio. A split point is defined as the midpoint of data where there is a change in class label. In entropy based SLIQ decision tree algorithm gain values are used for identifying best split point. The split point with maximum gain value is the best split.

Pre-sorting is the technique in which the data is sorted at the beginning of the decision tree. In entropy based SLIQ decision tree the data is sorted at the beginning of the tree. The dataset is sorted at each attribute along with its class label before identification of split points. For example, the humidity dataset is sorted as shown in Table IV.

The dataset after sorting contains the attribute values and their corresponding class labels. The dataset is analyzed and the points where there is a change in class label are excavated and the midpoint of the data becomes a split point.

From the Table IV, containing the sorted dataset of the attribute humidity, there are points where the class label is changed and their midpoints become split points of the dataset for that attribute. For example, in Table IV at positions 2 and 3, the variation of class label is identified. Hence, the midpoint of positions 2 and 3 i.e. midpoint (76, 76) = 76 become the split point of the dataset. In the similar manner remaining 8 more split points are calculated, shown in Table IV.

Split point = midpoint (data where class label changes) (8)

Repeat the procedure to find out the split points for the attribute temperature shown in Table V, attribute pressure shown in Table VI, attribute wind speed shown in Table VII and attribute dew point shown in Table VIII. Only best split point is chosen for every attribute and the selection of that best split point is illustrated in Table III.

The order in which the attributes are chosen plays an important role. In Fig. 1 the attribute dew point is chosen first. The pressure attribute must be examined second. The construction of tree continues till the training data is perfectly classified. Fig. 1 shows a complete decision tree for the training dataset given in Table. I

TABLE IV. DATASET SORTING ON HUMIDITY

Humidity	Class	Split Point
75	No Rain	
76	No Rain	76
76	Rain	77
78	No Rain	
79	No Rain	
79	No Rain	
80	No Rain	80.0
80	Rain	80.5
81	No Rain	81.5
82	Rain	
82	Rain	83.0
84	No Rain	
84	No Rain	
85	No Rain	
86	No Rain	86.0
86	Rain	
86	Rain	
86	Rain	
87	Rain	87.5
88	No Rain	88.0
88	Rain	
88	Rain	
89	Rain	
91	Rain	
92	Rain	
93	Rain	
94	Rain	
95	Rain	
95	Rain	
97	Rain	

TABLE V. DATASET SORTING ON TEMPERATURE

Temperature	**Class**	Split Point
24	Rain	
24	Rain	
24	Rain	
25	Rain	25.5
26	No Rain	
26	No Rain	
26	No Rain	26
26	Rain	
26	Rain	
26	Rain	
26	Rain	
26	Rain	
26	Rain	26.5
27	No Rain	
27	No Rain	
27	No Rain	
27	No Rain	
27	No Rain	
27	No Rain	27
27	Rain	
27	Rain	
27	Rain	
27	Rain	
27	Rain	
27	Rain	
27	Rain	27.5
28	No Rain	
28	No Rain	28.0
28	Rain	28.5
29	No Rain	

TABLE VI. DATASET SORTING ON PRESSURE

Pressure	**Class**	Split Point
1004	No Rain	
1004	No Rain	1004
1004	Rain	
1005	Rain	
1005	Rain	
1005	Rain	
1005	Rain	
1005	Rain	
1005	Rain	
1005	Rain	1005.5
1006	No Rain	
1006	No Rain	1006
1006	Rain	
1006	Rain	
1006	Rain	
1006	Rain	1006.5
1007	No Rain	
1007	No Rain	
1007	No Rain	1007
1007	Rain	
1007	Rain	
1007	Rain	
1007	Rain	1007.5
1008	No Rain	
1008	No Rain	
1008	No Rain	
1008	No Rain	1008
1008	Rain	1008.5
1009	No Rain	1009
1009	Rain	

TABLE VII. DATASET SORTING ON WIND SPEED

Wind Speed	Class	Split Point
6	No Rain	6
6	Rain	7
8	No Rain	9
10	Rain	10.5
11	No Rain	11
11	Rain	
11	Rain	
11	Rain	
11	Rain	12
13	No Rain	
13	No Rain	
13	No Rain	
13	No Rain	
13	No Rain	13
13	Rain	
13	Rain	13.5
14	No Rain	
14	No Rain	14
14	Rain	
14	Rain	
14	Rain	
14	Rain	
14	Rain	15
16	No Rain	16
16	Rain	
16	Rain	
16	Rain	17
18	No Rain	18
18	Rain	
18	Rain	

TABLE VIII. DATASET SORTING ON DEW POINT

Dew Point	Class	Split Point
19	No Rain	
19	No Rain	
19	No Rain	19
19	Rain	19.5
20	No Rain	
20	No Rain	
20	No Rain	
20	No Rain	20
20	Rain	
20	Rain	
20	Rain	20.5
21	No Rain	
21	No Rain	
21	No Rain	
21	No Rain	
21	No Rain	21
21	Rain	
21	Rain	
21	Rain	
21	Rain	
21	Rain	
21	Rain	
21	Rain	
21	Rain	
21	Rain	
21	Rain	
21	Rain	
21	Rain	
22	Rain	
22	Rain	

Some experiments have been conducted on real data to analyze the accuracy of the tree. We have used the dataset from the accuweather.com of Indian Meteorological Department. The goal is to predict the precipitation for rainfall. The training dataset consists of 1 to 15 years of data from the year 1997 to 2012 containing 365 to 4992 examples. The test dataset consists of 6 years consisting of 2073 records.

In this paper, the analysis had been made to test the efficiency of entropy in the prediction of precipitation by constructing a decision tree. It is observed that the average efficiency has been found to be 74.92% for the entire 14 year dataset. Though, this contributes a decent efficiency or success rate, the other methods of back propagation neural networks, linear discriminate statistical analysis and J48 are analyzed to select the best performing method of prediction of precipitation.

It has been found in Table IX, the distinction between the success rate of prediction and time. It can also be observed, that the efficiency obtained is 74.1% on one year dataset, 77.47% for two years dataset, 75.18% for three years dataset, 74.14% for four years dataset, 75.56% for a five 5 years dataset and 77.78 % for a six years dataset. The average efficiency has been found to be 74.92%. Though, this contributes a decent efficiency or success rate, the other methods of back propagation neural networks [7,8], [12-15], linear discriminate statistical analysis [16] and J48 are analyzed to select the best performing method of prediction of the precipitation.

The published results for this dataset are: 64.3% accuracy for back propagation, 58% for a linear discriminant and 68.6% for J48. Also the recently proposed model by Prasad et. al has established an accuracy of 72.3%. Using the same training and test datasets, as the average accuracy using SLIQ with entropy is 74.92% as shown in Fig. 2, SLIQ using entropy can be considered as the best performing method for the prediction of precipitation.

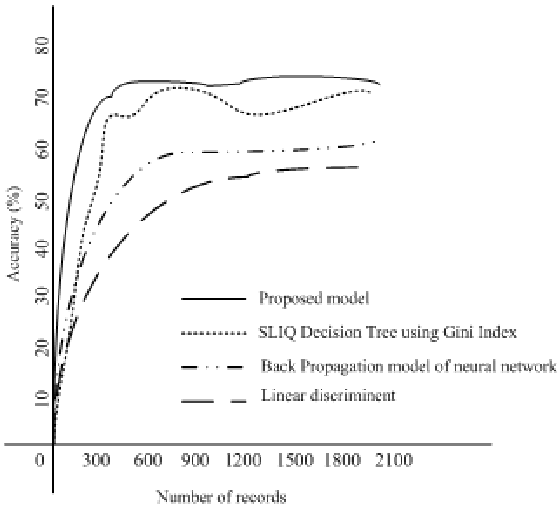

FIG. 2. AVERAGE TESTING ACCURACY GRAPH

TABLE IX. RESULT SHOWING THE ACCURACY AND TIME OF RESPONSE

No. of records	Correct predictions	In correct predictions	Accuracy (%)	Time (Sec)
363	269	94	74.104	37
728	564	164	77.471	40
995	765	230	75.181	42
1354	1031	323	74.141	44
1711	1010	401	75.563	46
2073	1579	494	74.164	47

VII. CONCLUSION

In this vast universe, food is the primary need for all living organisms. Agriculture is the main source for food; rain plays a prominent role for a better agriculture. Due to irregular rainfall the cultivated lands turn into barren lands. The prediction of rainfall depends on various attributes like humidity, temperature, pressure, wind speed, dew point etc. For prediction of precipitation many approaches like neural networks, back propagation, linear discriminant, Bayesian networks etc. are used but the success rate is not appreciable. So in the research decision tree algorithm is applied based on entropy for the prediction of precipitation.

This model provides the accurate predictions based on the historical climate data, obtained from meteorological department. With the results, it is clearly visible that SLIQ decision tree algorithms predicts at a greater accuracy rate i.e. matching with actual prediction. An entropy based SLIQ decision tree algorithm on an average provides 74.92% accuracy.

VIII. FUTURE ENHANCEMENTS

In this paper, we mainly focused on entropy based SLIQ decision tree algorithm, which gives maximum accuracy and in future implementation various other decision tree algorithms like CART, SPRINT, ELEGANT, EC4.5 with additional parameters can be used for better prediction accuracy.

REFERENCES

[1] Manish Mehta, Rakesh Agrawal, Jorma Rissanen, “SLIQ, A Fast Scalable Classifier for Data Mining,” EDBT '96 Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology, Springer-Verlag London. UK, pp.18-32, 1996.

[2] Narasimha Prasad, Prudhvi Kumar Reddy, Naidu MM, “An Approach to Prediction of Precipitation Using Gini Index in SLIQ Decision Tree”, 4th International Conference on Intelligent Systems, Modeling & Simulation, Bangkok, pp.56-60, 2013.

[3] Yu-Shan Shih, “Families of Splitting Criteria for Classification Trees”, Statistics and Computing, Vol 9, pp.309-315, 1999.

[4] Mahesh V. Joshi, Eui-Hong (Sam) Han, George Karypis, Vipin Kumar, Parallel Algorithms in Data Mining. CRPC Parallel Computing Handbook. Morgan Kaufmann, 2000.

[5] B. Chandra, P. Paul Varghese, "Fuzzy SLIQ Decision Tree Algorithm", IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics. Vol.38, 2008.

[6] K Frank Silvio Marzano, Giancarlo Rivolta, Erika Coppola, Barbara Tomassetti and Marco Verdecchia, “Rainfall Nowcasting from Multisatellite Passive-Sensor Images Using a Recurrent Neural Network”, 2007, IEEE Transactions on Geosciences and Remote Sensing, Vol. 45.

[7] K.Richards and G.D. Sullivan, “Estimation of cloud cover using colour and texture”, 1992, British Machine Vision Conference, Springer, PP. 436-442.

[8] C. Jareanpon, W. Pensuwon, R.J. Frank and N. Davey, “An adaptive RBF network optimized using a Genetic algorithm applied to Rainfall Forecasting”, 2004, International Symposium on Communications and Information Technologies.

[9] K. Ochiai, H. Suzuki, S. Suzuki, N. Sonehara and Y. Tokunaga, “Snowfall and rainfall forecasting from the images of weather radar with artificial neural networks”, 1995, IEEE International Conference on Neural Networks.

[10] D.I.F. Grimes, E. Coppola, M. Verdecchia and G. Visconti, “A Neural Network Approach to Real-time Rainfall Estimation for Africa and using Satellite Data”, 2003, American Meteorological Society.

[11] Thiago N. de Castro, Francisco Souza, Jose M.B. Alves, Ricardo S.T. Pontes, Mosefran B.M. Firmino and Thiago M. de Pereria, “Seasonal Rainfall Forecast using a Neo- Fuzzy Neuron Model”, 2011, 9th IEEE International Conference on Industrial Informatics.

[12] Tuan Zea Tan, Gary Kee Khoon Lee, Shie-Yui Liong, Tian Kuay Lim, Jiawei Chu and Terence Hung, “Rainfall Intensity Prediction by a Spatial-Temporal Ensemble,” 2008, IEEE International Joint Conference on Neural Networks.

[13] Jiansheng Wu in “A Novel Nonlinear Ensemble Rainfall Forecasting Model incorporating Linear and Nonlinear Regression”, 2008, Fourth International Conference on Natural Computation.

[14] Hui Qi, Ming Zhang and Roderick A. Scofield, “Rainfall Estimation using M-PHONN Model”, 2001, IEEE International Conference on Neural Networks.

[15] Wint Thida Zaw and Thinn Thu Naing, “Modeling of Rainfall prediction over Myanmar using Polynomial Regression”, 2009, International Conference on Computer Engineering and Technology.

[16] C. Kidd and V. Levizzani, “Status of Satellite Precipitation Retrievals”, 2011, Hydrology and Earth System Sciences.

[17] Sanjay D. Sawaitul, Prof. K.P. Wagh, Dr. P.N. Chatur, “Classification and Prediction of Future Weather by using Back Propagation Algorithm - An Approach” International Journal of Emerging Technology and Advanced Engineering.

[18] Soroosh Sorooshian, Kuo- lin Hsu, Bisher Imam and Yang Hong, “Global Precipitation Estimation from Satellite Image using Artificial Neural Networks”, Cambridge University Press, 2007, PP. 21-28.

[19] Kesheng Lu and Lingzhi Wang, “A Novel Nonlinear Combination Model Based on Support Vector Machine for Rainfall Prediction” 2011 Fourth International Joint Conference on Computational Sciences and Optimization.

[20] Grant W. Petty and Witold F. Krajewski, “Satellite Estimation of Precipitation over land”, 1996, Hydrological Sciences Journal.

[21] K. Richards and G.D. Sullivan (2006), “*Estimation of Cloud Cover using Colour and Texture Intelligent Systems Group*”, University of Reading, RG6 2AY.

[22] Koizumi, K. (1999). “An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network, Weather Forecast”, 14, 109-118.

[23] K. Richards, G.D. Sullivan, Estimation of Cloud Cover using Colour and Texture Intelligent Systems Group. University of Reading. RG6 2AY, 2006.

[24] Koizumi. K, An Objective Method to Modify Numerical Model Forecasts with Newly given Weather Data using an Artificial Neural Network, Weather Forecast, vol.14, pp.109-118, 1999.

[25] Luk K. C., Ball J. E, Sharma A, A Study of Optimal Model Lag and Spatial Inputs to Artificial Neural Network for Rainfall Forecasting, Journal of Hydrology, vol.227, pp.56-65, 2000.

[26] Robert A. Houze, Cloud Dynamics. Academic Press, 1993.

[27] J.R. Quinlan, Introduction of Decision Tree, Machine Learning, Vol. 1, pp.81-106, 1986.

[28] Wei-Yin Loh, Classification and Regression Tree Methods. Ruggeri, Kenett and Faltin. Wiley, pp.315–323 2008.

[29] Wang Yong, XU Hong, Guo Zengzhang, Ding Keliang, Liu Yanping, Wen Debao, the Study of Rainfall Forecast Based on Neural Network and GPS Precipitable Water Vapor. IEEE International Conference on Environmental Science and Information Application Technology: pp.17–20, 2010.

[30] Jehangir Ashraf Awan, Onaiza Maqbook, Application of Artificial Neural Networks for Monsoon Rainfall Prediction. IEEE International Conference on Emerging Technologies, pp.27–32, 2010.

[31] Jean Claude Berges, Neural Networks and Tree Classifiers. IEEE International Symposium on Geoscience and Remote Sensing, pp.887–889, 2003.

[32] Yuhui Wang, Yunzhong Jiang, Xiaohui Lei, Wang Hao, Rainfall-Runoff Simulation Using simulated Annealing Wavelet BP Neural Networks. IEEE International Conference on Intelligent Computation Technology and Automation, pp.963–967, 2010.

[33] Jiansheng Wu, A Novel Nonlinear Eensemble Rainfall Forecasting Model Incorporating Linear and Nonlinear Regression. IEEE International Conference on Natural Computation: pp.34–38, 2008.