I J C T A, 9(24), 2016, pp. 277-289 ISSN: 0974-5572

ISLGAS: Improved Supervised Learning in Quest Using Gain Ratio as Attribute Selection Measure to Nowcast Snow/No-Snow

Kishor Kumar Reddy C.* and Vijaya Babu**

ABSTRACT

The main crux of the present research is to nowcast the presence of snow/no-snow more accurately by making use of historical weather datasets and decision tree approach. In this paper, a new algorithm named Improved Supervised Learning in Quest using Gain Ratio as Attribute Selection Measure (ISLGAS) is proposed. The proposed algorithm is compared with existing decision tree algorithms such as SLIQ, SPM, SLGAS, ISLIQ and ISPM in terms of the overall classification performance defined over four different performance measures namely accuracy, specificity, precision and error rate. Experimental results show that the ISLGAS algorithm scales up well to both large and small datasets with large number of attributes and class labels.

Index Terms: Atmosphere, Classification, Decision Trees, Forecasting, Gain Ratio, Snowfall

1. INTRODUCTION

Weather is one of the effective environmental constraints in every phase of our lives. People started adjusting themselves with respect to weather condition from their dressing habits to strategic organizational planning activities, since the adverse weather conditions may cause a considerable damage on lives and properties. One should be alert to these adverse weather conditions by taking some precautions and using prediction mechanisms for early warning of hazardous weather phenomena [1].

There is a general and increasing interest on weather information, since every day we habitually give an ear to weather forecast news for local and large-scale long-term or short-term weather predictions. Leading weather research institutions such as World Meteorological Organization (WMO), International Research Institute (IRI), World Climate Application and Services Program (WCASP), academicians, scientists, meteorologists and researchers have been developing weather prediction systems capable of detecting, predicting and forecasting weather phenomena and hazards by utilizing state-of-the-science technologies. Thus weather prediction utilizations fields and prediction accuracy increases monotonically by the time [1-4].

A wide range of weather forecast methods are employed at regional, national and global levels. Fundamentally, there are two approaches in weather prediction: empirical method and dynamical methods. The empirical approach is based on analysis of historical data of the rainfall and its relationship to a variety of atmospheric and oceanic variables over different parts of the world. The most widely use empirical approaches are regression, artificial neural network, fuzzy logic and group method of data handling [5] [7-8] [24] [26-35]. In dynamical approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions.

* Research Scholar, Department of Computer Science and Engineering, K L University, Guntur, *Email: Kishoar23@gmail.com*

** Professor, Department of Computer Science and Engineering, K L University, Guntur, *Email: Vijaymtech28@gmail.com*

Snow disasters cause immeasurable losses to human society each year, threatening people's lives and properties and therefore attaching much importance to the measurement of snow cover. Snowfall, however, is one of the most difficult to be measured among meteorological elements, especially on the transport sector such as road, air and rail ways. To solve this problem, the real-time nowcasting of snow more accurately is to be developed. This can effectively avoid railway, road and air accidents caused by snowstorms, improving transport safety as well as providing a quantitative reference for the safe operation [6]. Technological improvements in the computational power are still not sufficient to handle the weather prediction efficiently. Although the current abilities of computer systems helped the meteorologist to implement more advance model that requires high computation and improves the prediction capabilities; the accuracy and timely prediction of weather phenomena is still a major issue. Further, the global climate changes and incident of some disastrous weather events increased the importance of timely and accurate weather prediction [1].

In 2011, the hazardous weather events caused more than $50 billion loss to US and 675 deaths. Pakistan faced deadly flood of its history which caused 1985 human loss and $9.8 billion financial loss. Russia gone through the worst ever drought and the deadliest heat wave in human history. Australia and Columbia faced a record natural disaster due to heavy rain. Another Red signal is issued by the World Economic Forum, who reported that the 21st century climate change will be one of the greatest global challenges for human. Despite the fact that none can control the natural disasters but accurate and timely prediction can help to minimize loss of human lives and other financial cost, which became the motivating factors for the development of snow nowcasting model. [4-5].

The present research employs empirical method: decision tree approach, which tries to make a short-term nowcast of snowfall over different parts of the globe. In this paper, we try to give readers an overview about weather prediction phenomena, expert systems approaches, main domain specific problems, and solution methodologies. The present research is the enhancement to our previous papers such as SLIQ [31] [39], ISLIQ [46], SLEAS [38], SPM [41], ISPM [47], SLGAS [45] to nowcast snow/no-snow [26-35].

Our specific contributions in this paper are listed below:

a. The model is capable to nowcast snow/no-snow based on the weather attributes: humidity, temperature, pressure, wind speed, dew point and visibility more effectively.

b. The proposed model has the capability to predict weather before 4 hours more effectively.

c. A detailed evaluation against other prediction decision tree algorithms is performed, that provide a fair comparison to show the effectiveness of the proposed model.

d. The proposed model is computationally efficient and makes it suitable for small devices such as android environment.

e. The proposed model is evaluated with various performance measures such as accuracy, specificity, precision, error rate and also in terms of number of split points.

The rest of the paper is organized as follows: the next section reviewed the literature, which is followed by the description and working of the new model in section III. Section IV analyzes and discusses the result and finally section V concludes the paper with future directions and references.

2. RELEVANT WORK

Accurate and timely weather forecasting is a major challenge for the scientific community. Snowfall nowcasting modeling involves a combination of computer models, observation and knowledge of trends and patterns. Using these methods, reasonably accurate forecasts can be made up. Several recent research studies have developed snowfall prediction using different weather and climate forecasting methods [8] [17] [24] [26-35].

Irene Y.H Gu et al. [6], put forward a full automatic image analysis system for detection and analysis of snow/ice coverage on electric insulators of power lines using images which were captured by visual cameras in a remote outdoor laboratory test bed. Jinmei Pan et al. [7], put forth a passive microwave remote sensing techniques that detected wet snow in the south of china. Yajaira Mejia et al. [8], gave an approach for estimating the snowfall using neural networks on multi source remote sensing observations and ground based meteorological measurements. Melanie Wetzel et al. [9], projected a technique that supports the snowfall forecast and for the verification of radar limited mountainous terrain that includes matching the output parameters and graphics from high resolution mesoscale models to surface mesonets. Pascal Sirguey et al. [10], made use of ASTER and MODIS sensors, both on the TERRA platform by implementing the ARSIS concept so as to fuse the high spatial content of the two 250m spectral bands of MODIS into five 500m bands using wavelet based multi resolution analysis in the mountainous environment.

Michael A. Rotondi [11] illustrated a Markov chain models across eight national weather stations using historical data from the global historical climatology network to predict a ‘snow day’. Gail M. Skofronick Jackson et al. [12], in their research interpreted how instruments like the W-band radar of Cloudsat, Global Precipitation Measurement Dual-Frequency Precipitation Radar ku- and Ka-bands, and the Microwave Imager can be used in the simulations of lake effect and synoptic snow events in order to determine the minimum amount of snow. Gail M. Skofronick Jackson et al. [13], demonstrated thresholds for detecting falling snow from satellite-borne active and passive sensors.

Andrea Spisni et al. [14], presented an operational chain developed in the Emilia-Romagna region to monitor snow cover and snow water equivalent over the area managed by the Regional Catchment Technical Service. Alberto Martinez Vazquez et al. [15] presented an algorithm using GB-SAR imagery for the automatic recognition and classification of snow avalanches. Jeremie Bossu et al. [16], made use of a structure, based on computer vision which detects the presence of snow or rain. Noel Dacruz Evora et al. [17], used brightness temperature data, provided by seven channels SSM/I aboard the Defense Meteorological Satellite Program F-11 and F-13 spacecrafts. Using which a modelling framework was put forth by combining passive microwave data, neural network based models and geostatistics for snow water equivalent retrieval and mapping. Hossein Zeinivand et al. [18], enforced a spatially distributed physically based model to detect snow and melting in the Latyan dam watershed in Iran.

Xiaolan Xu et al. [19], developed a model that can be used for both active and passive microwave remote sensing of snow. B.B Fitzharris et al. [20], presented three case studies on the usage of satellite imagery for mapping seasonal snow cover in New Zealand, and also explored the effectiveness of using AVHRR imagery in order to obtain the presence of snow, snow covered area and snow line elevation on the mountain ranges of New Zealand. Ashok N.Srivastava et al. [21], in their research discussed the results based on kernel methods for unsupervised discovery of snow, ice, clouds and other geophysical processes based on data from the MODIS instrument.

G. Singh et al. [22], developed a Radar Snow Index model to identify snow using SAR polarimetry techniques. In their research, full polarimetric L-band ALSOS-PALSAR data of snow cover area in Himalayan region have been analyzed based on various component scattering mechanism models and all model results are compared. Fan Ke et al. [23], developed a model to identify winter time heavy snow over Northeast China by using a inter annual increment prediction approach. Folorunsho Olaiya [24] investigated the use of artificial neural networks and decision tree algorithms in forecasting maximum temperature, rainfall, evaporation and wind speed using meteorological data collected from the city of Ibadan, Nigeria through Nigerian Meteorological Agency, Oyo state office. Manjeet Singh et al. [25] forwarded an attempt to develop an automatic technique for avalanche area identification and also its severity index. For detailed relevant work refer our earlier papers [38] [39] [41] [45-47].

3. ISLGAS DECISION TREE ALGORITHM

Classification is the task of learning a target function *f* that maps each attribute set *x* to one of the pre-defined class labels *y*. The input for the classification is the training dataset, whose class labels are already known. Classification analyzes the training dataset and constructs a model based on the class label, and aims to assign a class label to the future unlabelled records [5] [35]. A set of classification rules are generated by such a classification process, which can be used to classify future data and develop a better understanding of each class in the database. Some of the classification models are decision trees [31] [38-41] [45], neural networks [8] [17], genetic algorithms, statistical models like linear/geometric discriminates [35]. In the present research we are introducing a novel decision tree algorithm entitled ISLGAS for the now casting of snow/no-snow.

Procedure for evaluating the split points and decision tree generation

a. Read the training dataset T

b. Sort T in ascending order and choose the initial attribute along with the associated class label.

c. Evaluate the interval range, as shown in equation 1.

$$Interval\ Range = \frac{a\max - a\min}{Group\ size} \tag{1}$$

Where amax is the maximum value for the particular attribute, amin is the minimum value for the particular attribute and Group size is to be fixed by user. Upon the experimentation, it is identified that group size 3 is giving maximum accuracy when compared with other size. Based on the interval range, evaluate the split points and it is shown in equation (2) [38].

i. Initially check for change in the class label.

ii. If there is a change in the class label, evaluate the split points and the midpoint of changed class labels is the split point. For instance, Let V be the initial record and V_i be the second record: such that take Mid Point (V, V_i) only when there is change in the class label, shown in formula (2).

$$Split\ Point = Midpoint\,(V, V_i) \tag{2}$$

d. Choose the split point 1 and apply gain ratio attribute selection measure and evaluate the gain ratio value and continue this for all the split points obtained for initial attribute and the procedure is as follows:

i. Initially, consider attribute and also along with its associated class label and evaluate attribute entropy and it is shown in formula (3) [38].

$$Attribute\ Entropy = \sum_{j=1}^{N} P_j \left[-\sum_{i=1}^{M} P_i \log_2 P_i \right] \tag{3}$$

Where P_i is the probability of class entropy belonging to class i. Logarithm is base 2 because entropy is a measure of the expected encoding length measured in bits.

ii. Further, consider class label and evaluate class entropy and is as follows:

Class entropy is a measure in the information theory, which characterizes the impurity of an arbitrary collection of examples. If the target attribute takes on M different values, then the class entropy relative to this M-wise classification is defined in formula (4) [38].

$$Class\ Entropy = -\sum_{i=1}^{M} P_i \log_2 P_i \tag{4}$$

Where P_i is the probability of class entropy belonging to class i. Logarithm is base 2 because entropy is a measure of the expected encoding length measured in bits.

Now, compute the entropy: it is the difference of class entropy and attribute entropy and is shown in formula (5) [38].

$$Entropy = Class\ Entropy - Attribute\ Entropy \tag{5}$$

e. Compute Split Info for each and every attribute and is shown in formula (6).

$$Split\ Info = -\sum_{i=1}^{N} \frac{|T_i|}{|T|} \log_2 \left(\frac{|T_i|}{|T|} \right) \tag{6}$$

f. Compute Gain ratio: it is the proportion of Entropy generated by the split and is shown in formula (7) [5].

$$Gain\ Ratio = \frac{Entropy}{Split\ Info} \tag{7}$$

g. The maximum Gain Ratio is considered to be the best split attribute and becomes the root node, shown in formula 8 [5].

$$\text{Best split point} = \text{Maximum (Gain Ratio)} \tag{8}$$

h. Finally, if the number of attributes are N, we will get N best split points for individual attributes. As decision tree is a binary tree, there will be only one root node and for this reason, among the N Gain Ratio values choose one best Gain Ratio value to form the root node.

i. Now, consider the maximum Gain Ratio value attribute as the root node and take its split point and divide the tree in binary format i.e. keep the values which are lesser to split point at the left side of the tree and keep the values which are greater and equals to the right side of the tree, and continue the process till it ends with a unique class label.

4. RESULTS AND DISCUSSION

Technological improvements in the computational power are still not sufficient to handle the weather prediction efficiently. Although the current abilities of computer systems helped the meteorologists to

Table 1
Dataset Description

City Name	*Instances*	*Training*	*Testing*	*Attributes*	*Classes*
Bangkok	5740	4305	1435	6	2
Barcelona	6013	4510	1504	6	2
Botswana	6047	4535	1512	6	2
Cairo	6143	4607	1536	6	2
Delhi	6015	4511	1504	6	2
Eglinton	6318	4738	1580	6	2
Humberside	1036	777	259	6	2
Hyderabad	5849	4387	1462	6	2
Iceland	3512	2634	878	6	2
Lahore	4887	3665	1222	6	2
Manchester	6338	4753	1585	6	2
Norway	6105	4579	1526	6	2
Perth	6182	4636	1546	6	2
Sellaness	5412	4059	1353	6	2
Tiruptahi	6039	4529	1510	6	2

implement more advanced models that requires high computation and improves the prediction capabilities; the accuracy and timely prediction of weather phenomena is still a major issue. The proposed model has been tested on 15 international locations historical datasets of snow/no-snow, collected from www.wunderground.com [36]. We conducted experiments by implementing our proposed algorithm in Java Net Beans IDE 7.2. All experiments were performed on intel i3 core processor and 4 GB RAM with windows 7 operating system. We also divided our data set in to two parts: training set (75%), which is used to create the model, and a test set (25%), which is used to verify that the model is accurate and not over fitted [35]. Table 1 summarizes the characteristics of the datasets, arranged in alphabetical order, presenting the number of instances, training instances, testing instances, and classes.

A common but poorly motivated way of evaluating results of Machine Learning Experiments is using specificity, accuracy, precision, dice and error rate. Specificity relates to the test's ability to exclude a condition correctly [35] [37]. Precision is defined as the proportion of the true positives against all the positive results [35] [37]. Prediction error is a measure of the performance of a model to predict the correct output, given future observations used as predictors. In order to reveal the performance of our proposed ISLGAS algorithm, we present comparison between SLIQ, SPM, SLGAS, ISLIQ, ISPM and ISLGAS in terms of classification accuracy, specificity, precision and error rate, using 15 international locations datasets collected from www.wunderground.com [36].

The comparison in terms of split points is presented in Table 2. Apparently, almost all results for ISLGAS are better than those of SLGAS. The results clearly show that the proposed ISLGAS reduces nearly 51.48 % when compared with SLGAS decision tree algorithm. Figure 1 presents the comparison of number of split points of the SLGAS and proposed ISLGAS algorithms graphically.

The comparison in terms of classification accuracy is presented in Table 3. The proposed method yielded an average accuracy of 93.25%, better, when compared with SLIQ, SPM, SLGAS, ISLIQ, and ISPM. For some of the cities, the accuracy levels are more for other algorithms when compared with ISLGAS. But, on an average the ISLGAS model outperforms when compared with other algorithms, graphically shown in Figure 2.

Table 2
Split Points Comparison

City Name	*SLGAS*	*ISLGAS*
Bangkok	112	50
Barcelona	174	41
Botswana	195	88
Cairo	165	162
Delhi	281	162
Eglinton	360	43
Humberside	171	34
Hyderabad	116	72
Iceland	385	207
Lahore	190	51
Manchester	499	211
Norway	765	560
Perth	246	136
Sellaness	391	154
Tirupathi	154	108

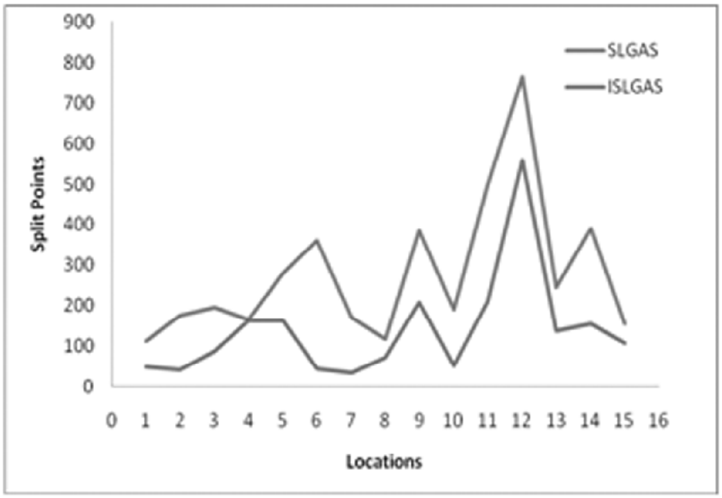

Figure 1: Split points comparison

Table 3
Accuracy Comparison

City Name	*SLIQ*	*SPM*	*SLGAS*	*ISLIQ*	*ISPM*	*ISLGAS*
Bangkok	96.09	94.49	95.19	98.11	98.32	98.9
Barcelona	95.8	95.14	95.67	96.07	96.07	96.07
Botswana	93.78	96.16	93.58	95.43	96.29	96.62
Cairo	88.99	89.7	89.77	89.98	89.32	90.1
Delhi	96.14	94.94	96.8	93.15	96.34	96.8
Eglinton	89.24	90.06	90.06	89.56	89.75	90.87
Humberside	93.05	94.59	94.98	93.82	94.2	94.82
Hyderabad	96.5	97.8	94.79	96.4	97.8	97.67
Iceland	89.17	88.49	86.2	88.49	88.95	89.81
Lahore	84.82	86.05	85.89	84.65	86.38	87.06
Manchester	92.74	92.87	89.58	93.43	91.29	93.36
Norway	88.99	90.89	90.62	90.89	90.69	90.3
Perth	94.3	94.43	93.31	94.43	94.24	94.37
Sellaness	75.9	77.67	79.45	84.4	84.18	84.7
Tirupathi	97.54	97.41	97.41	97.48	97.54	97.35
Mean	90.53	91.04	91.55	92.41	92.75	93.25

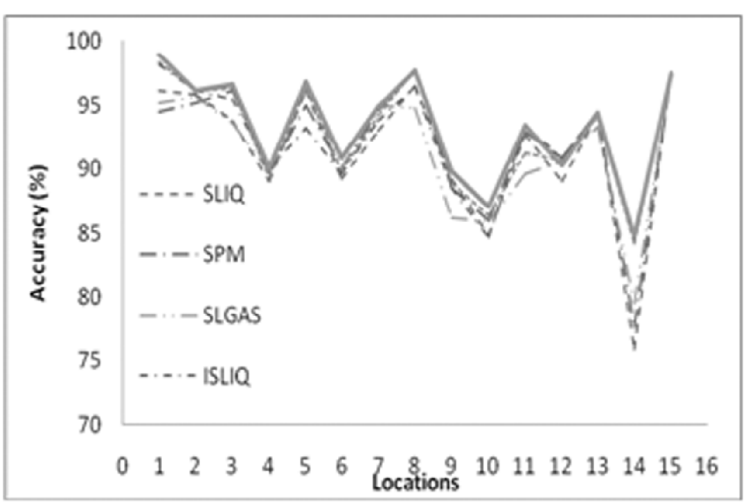

Figure 2: Accuracy comparison

The comparison in terms of specificity is presented in Table 4. For some of the cities, the specificity levels are more for other algorithms when compared with ISLGAS. But, on an average the ISLGAS model outperforms when compared with other algorithms. The proposed method yielded an average specificity of 96.76%, better, when compared with SLIQ, SPM, SLGAS, ISLIQ, and ISPM. Figure 3 presents the classification specificity of the SLIQ, SPM, SLGAS, ISLIQ, ISPM and proposed ISLGAS algorithms graphically. As can be observed, ISPM obtained better results than SLIQ, SPM, SLGAS and ISLIQ i.e. larger specificity

Table 4
Specificity Comparison

City Name	*SLIQ*	*SPM*	*SLGAS*	*ISLIQ*	*ISPM*	*ISLGAS*
Bangkok	96.36	94.75	95.45	98.53	98.74	98.32
Barcelona	97.59	97.8	97.87	98.83	98.83	98.28
Botswana	94.68	97.24	94.41	96.56	97.3	97.64
Cairo	95.53	96.95	96.74	96.43	96.1	96.52
Delhi	97.83	96.34	98.44	94.58	97.83	98.3
Eglinton	97.52	97.87	97.66	98.44	98.72	98.86
Humberside	95.49	98.36	98.36	97.13	97.13	97.13
Hyderabad	97.77	99.16	95.97	98.16	99.16	99.02
Iceland	96.51	93.29	95.48	93.67	96	94.32
Lahore	86.93	89.5	90.59	89.7	89.5	89.4
Manchester	95.69	95.9	92.94	96.9	94.01	95.83
Norway	95.92	98.14	97.78	97.92	97.71	97.42
Perth	96.73	96.87	98.8	96.87	96.73	96.53
Sellaness	78.74	82.96	85.45	93.02	92.42	88.46
Tiruptahi	98.06	97.86	98.06	97.93	98.06	97.93
Mean	94.75	95.53	95.6	96.31	96.54	96.76

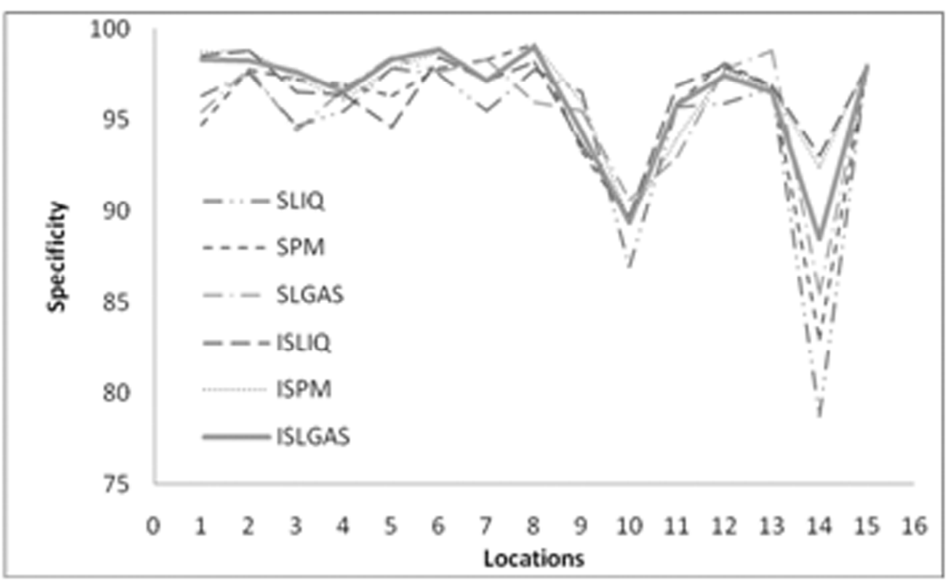

Figure 3: Specificity comparison

The comparison in terms of precision is presented in Table 5. Apparently, almost all precision results for ISLGAS are better than those of SLIQ, SPM, SLGAS, ISLIQ, and ISPM. For some of the cities, the precision levels are more for other algorithms when compared with ISLGAS. But, on an average the ISLGAS

Table 5
Precision Comparison

City Name	*SLIQ*	*SPM*	*SLGAS*	*ISLIQ*	*ISPM*	*ISLGAS*
Bangkok	3.7	2.59	2.98	0	0	0
Barcelona	35.18	15.78	29.54	22.72	22.72	34.21
Botswana	12.22	18	12.63	13.55	20	22.22
Cairo	22.22	17.3	22.03	22.12	21.42	30
Delhi	3.03	8.47	8	4.76	11.11	13.79
Eglinton	46.96	56.52	56	51.11	55	57.89
Humberside	42.1	55.55	60	46.15	50	46.15
Hyderabad	3.03	0	3.33	14.12	14.81	0
Iceland	56.45	50.94	59.77	51	54.41	47.61
Lahore	54.16	57.76	58.14	54.78	58.43	55.6
Manchester	41.81	42.45	26.05	45.88	35.03	38
Norway	20.83	33.33	31.11	35.55	33.33	28
Perth	9.25	9.61	21.73	9.61	7.54	14.75
Sellaness	31.19	30.52	32.66	42.95	42.48	32.32
Tiruptahi	6.45	8.57	0	8.82	6.45	3.12
Mean	25.9	27.15	28.26	28.2	28.84	28.93

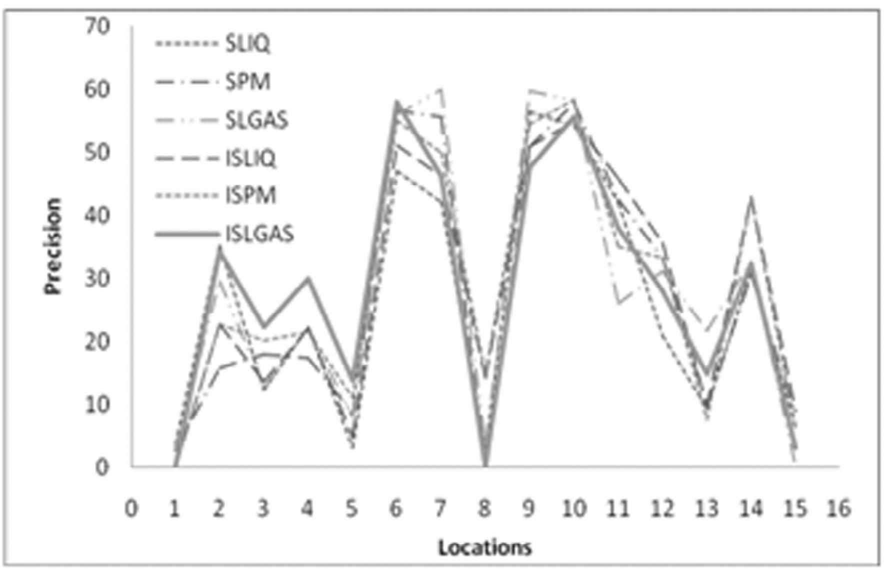

Figure 4: Precision comparison

model outperforms when compared with other algorithms. The proposed method yielded an average precision of 28.93%, better, when compared with SLIQ, SPM, SLGAS, ISLIQ, ISPM. Figure 4 presents the classification precision of the SLIQ, SPM, SLGAS, ISLIQ, ISPM and proposed ISLGAS algorithms graphically.

The comparison in terms of error rate is presented in Table VI. Apparently, almost all error rate results for ISLGAS are better than those of SLIQ, SPM, SLGAS, ISLIQ, and ISPM. The proposed method yielded an average error rate of 3.37%, better, when compared with SLIQ, SPM, SLGAS, ISLIQ, and ISPM. Figure 5 presents the error rate of the SLIQ, SPM, SLGAS, ISLIQ, ISPM and proposed ISLGAS algorithms graphically. As can be observed, ISLGAS obtained better results than SLIQ, SPM, SLGAS, ISLIQ, ISPM i.e. lesser error rate.

Table 6
Error rate Comparison

City Name	*SLIQ*	*SPM*	*SLGAS*	*ISLIQ*	*ISPM*	*ISLGAS*
Bangkok	3.91	5.51	4.81	1.89	1.68	0.55
Barcelona	4.2	4.86	4.33	3.93	3.93	1.97
Botswana	6.22	3.84	6.42	4.57	3.71	1.69
Cairo	11.01	10.3	10.23	10.02	10.68	4.95
Delhi	3.86	5.06	3.2	6.85	3.66	1.60
Eglinton	10.76	9.94	9.94	10.44	10.25	4.57
Humberside	6.95	5.41	5.02	6.18	5.8	2.59
Hyderabad	3.5	2.2	5.21	3.6	2.2	1.17
Iceland	10.83	11.51	9.8	11.51	11.05	5.10
Lahore	15.18	13.95	14.11	15.35	13.62	6.47
Manchester	7.26	7.13	10.42	6.57	8.71	3.32
Norway	11.01	9.11	9.38	9.11	9.31	4.85
Perth	5.7	5.57	3.69	5.57	5.76	2.82
Sellaness	24.1	22.33	20.55	15.6	15.82	7.65
Tiruptahi	2.46	2.59	2.59	2.52	2.46	1.33
Mean	8.46	7.95	7.98	7.58	7.24	3.37

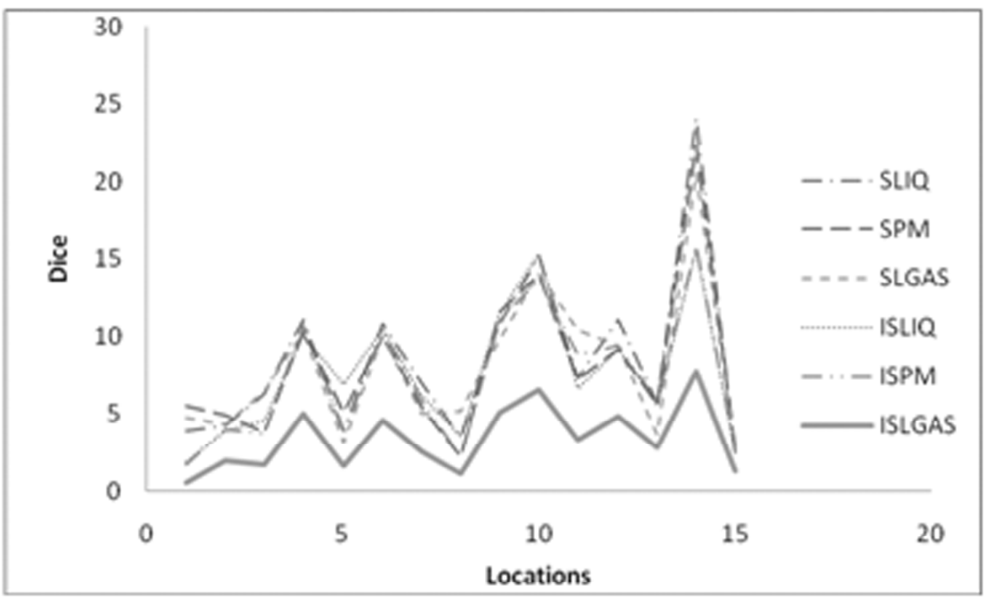

Figure 5: Error rate comparison

5. CONCLUSION

Weather prediction is inherently complex process, so it impossible to wait 100% accurate forecast results since we cannot measure all factors that may be local and global scales. Weather prediction systems are more likely decision support system than expert systems because they need guidance and weather predictions must be evaluated by human interference. Hybrid systems are very promising for integration of current expert systems on large scale.

Experimental results show that the ISLGAS algorithm scales up well to both large and small datasets with large number of attributes and class labels. We compare our proposed method with SLIQ, SPM, SLGAS, ISLIQ, ISPM decision tree algorithms in terms of the overall classification performance defined over four different performance measures namely accuracy, specificity, precision and error rate. Results on the snow/no-snow 15 international locations datasets show that:

a. the ISLGAS decision tree outperforms in terms of classification accuracy over 15 international locations of snow/no-snow datasets. The proposed method yielded an average accuracy of 93.25%, better, when compared with SLIQ, SPM, SLGAS, ISLIQ, and ISPM. For some of the cities, the accuracy levels are more for other algorithms when compared with ISLGAS. But, on an average the ISLGAS model outperforms when compared with other algorithms.

b. the ISLGAS decision tree outperforms in terms of classification specificity over 15 international locations of snow/no-snow datasets. For some of the cities, the specificity levels are more for other algorithms when compared with ISLGAS. But, on an average the ISLGAS model outperforms when compared with other algorithms. The proposed method yielded an average specificity of 96.76%, better, when compared with SLIQ, SPM, SLGAS, ISLIQ, and ISPM.

c. the ISLGAS decision tree outperforms in terms of classification precision over 15 international locations of snow/no-snow datasets. Apparently, almost all precision results for ISLGAS are better than those of SLIQ, SPM, SLGAS, ISLIQ, and ISPM. For some of the cities, the precision levels are more for other algorithms when compared with ISLGAS. But, on an average the ISLGAS model outperforms when compared with other algorithms. The proposed method yielded an average precision of 28.93%, better, when compared with SLIQ, SPM, SLGAS, ISLIQ, and ISPM.

d. the ISLGAS decision tree outperforms in terms of classification error rate over 15 international locations of snow/no-snow datasets. Apparently, almost all error rate results for ISLGAS are better than those of SLIQ, SPM, SLGAS, ISLIQ, and ISPM. The proposed method yielded an average error rate of 3.37%, better, when compared with SLIQ, SPM, SLGAS, ISLIQ, and ISPM.

In future, the most influencing parameters like Humidity, temperature, Pressure, Wind-Speed, Dew-Point and so on that affect the presence of snow/no-snow can be identified using remote sensing of real-time satellite imagery.

ACKNOWLEDGMENT

The authors would like to thank the referees for their valuable comments and suggestions, which greatly enhanced the clarity of this paper. We thank www.wunderground.com, for providing us the historical datasets of snow/no-snow. We are indebted to L V Narasimha Prasad for help and advice on the experimental analysis and paper setup. We are also grateful to the Management and Department of Computer Science and Engineering of Stanley College of Engineering & Technology for Women, Hyderabad and K L University, Guntur for providing their maximum support during the experimentation.

REFERENCES

[1] Robert A. Houze: Cloud Dynamics. Academic Press (1994).

[2] Nabilah Filzah Mohd Radzuan, Andi Putra, Zalinda Othman, Azuraliza Abu Bakar and Abdul Razak Hamdan: Comparative Study – Three Artificial Intelligence Techniques for Rain Domain in Precipitation Forecast. International Journal of Environmental, Ecological, Geological and Mining Engineering (2013) 898-903.

[3] http://en.wikipedia.org/wiki/Weather_forecasting.

[4] http://en.wikipedia.org/wiki/Quantitative_precipitation_forecast.

[5] J. Han: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers. (2001)

[6] Irene Y.H. Gu, Unai Sistiag and Anders Fahlstrom: Online Detection of Snow Coverage and Swing Angles of Electrical Insulators on Power Transmission Lines. IEEE International Conference on Image Processing (2009) 3249–3252.

[7] Jinmei Pan, Lingmei Jiang and Lixin Zhang: Wet Snow Detection in the South of China by Passive Microwave Remote Sensing. IEEE International Geo Science and Remote Sensing Symposium (2012) 4863-4866.

[8] Yajaira Mejia, Hosni Ghedira and Reza Khanbilvardi: A Neural Network Based Approach for Multi-Spectral Snowfall Detection and Estimation, IEEE International Geo Science and Remote Sensing Symposium (2007) 2276 – 2279.

[9] Melanie Wetzel, Michael Meyers, Randolph Borys, Ray Mcanelly, William Cotton, Andrew Rossi, Paul Frisbie, David Nadler, Douglas Lowenthal, Stephen Cohn, and William Brown: Mesoscale Snowfall Prediction And Verification In Mountainous Terrain. AMS Journal Of Weather And Forecasting (2004) 806-828.

[10] Pascal Sirguey, Renaud Mathieu Yves Arnaud, Muhammad M. Khan and Jocelyn Chanussot: Improved Resolution For The Detection Of Snow With Modis Using Wavelet Fusion. IEEE International Geo Science and Remote Sensing Symposium (2007) 3975-3978.

[11] Michael A. Rotondi: Estimating Transition Matrices to Predict Tomorrow's Snowfall Using Real Data. Journal of Statistics Education (2010) 1-14.

[12] Gail M. Sko Fronick Jackson, Benjamin T. Johnson, and S. Joseph Munchak: Detection Thresholds of Falling Snow From Satellite-Borne Active And Passive Sensors, IEEE Transactions On Geo Science and Remote Sensing (2013) 4177-4189.

[13] Gail M. Skofronick Jackson and Benjamin T. Johnson: Thresholds Of Detection For Falling Snow From Satellite-Borne Active And Passive Sensors, IEEE International Geo Science And Remote Sensing Symposium (2007) 2637-2640.

[14] Andrea Spisni, Fausto Tomei, Sara Pignone, Enrico Muzzi, Alessandro Panzacchi, Gabriele Antolini, Giulia Villani, Michele Di Lorenzo, Rosanna Foraci1, Marco Bittelli and Erin S. Brooks: Snow Cover Analysis in Emilia-Romagna. European Journal of Remote Sensing (2011) 59-73.

[15] Alberto Martinez Vazquez and Joaquim Fortuny Guasch: Snow Avalanche Detection and Classification Algorithm for GB-SAR Imagery, IEEE International Geo Science and Remote Sensing Symposium (2007) 3740-3743.

[16] Jeremie Bossu, Nicolas Hautière and Jean Philippe Tarel: Rain or Snow Detection in Image Sequences through Use of a Histogram of Orientation of Streaks, International Journal of Computer Vision (2011) 348-367.

[17] Noel Dacruz Evora, Ddominique Tapsoba and Danielle De Seve: Combining Artificial Neural Network Models, Geo statistics, and Passive Microwave Data for Snow Water Equivalent Retrieval and Mapping. IEEE Transactions on Geo Science and Remote Sensing (2008) 1925-1939.

[18] Hossein Zeinivand and Florimond De SmISLIQ: Simulation of Snow Covers area by a Physical Based Model. World Academy of Science, Engineering and Technology (2009) 428-433.

[19] Xiaolan Xu, Ding Liang, Leung Tsang, Konstantinos M. Andreadis, Edward G. Josberger, Dennis P. Lettenmaier, Donald W. Cline and Simon H. Yueh: Active Remote Sensing of Snow Using NMM3D/DMRT and Comparison With Clpx Ii Airborne Data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2010) 689-697.

[20] B. B. Fitzharris and B. P. Mcalevey: Remote Sensing of Seasonal Snow Cover in the Mountains of New Zealand Using Satellite Imagery. Taylor and Francis Geocarto International (1999) 35-44.

[21] Ashok N. Srivastava and Julienne Stroeve: Onboard Detection of Snow, Ice, Clouds and Other Geophysical Processes Using Kernel Methods. Proceedings of the ICML 2003 Workshop on Machine Learning Technologies for Autonomous Space Applications (2003) 1-5.

[22] G. Singh, Y. Yamaguchi, S. E. Park and G. Venkataraman: Identification of Snow Using SAR Polarimetry Techniques. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science (2010) 146-149.

[23] Fan Ke and Tian Baoqiang: Prediction of Wintertime Heavy Snow Activity in North East China. Springer Chinese Science Bulletin (2013) 1420-1426.

[24] Folorunsho Olaiya: Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, International Journal of Information Engineering and Electronic Business (2012) 51-59.

[25] Manjeet Singh, V. D. Mishra, N. K. Hakur and Jyoti Dhar Sharma: Remote Sensing GIS Based Statistical Modelling for the Prediction of Natural Hazards. International Journal of Engineering Research and Technology, (2012) 1-7.

[26] J.R. Quinlan: Induction of Decision Trees. Journal of Machine Learning (1986) 81-106.

[27] B. Chandra and P. Paul Varghese: Fuzzy Sliq Decision Tree Algorithm: IEEE Transactions on Systems, Man and Cybernetics (2008) 1294-1301.

[28] Anuja Priyama, Abhijeeta, Rahul Guptaa, Anju Ratheeb and Saurabh Srivastavab: Comparative Analysis of Decision Tree Classification Algorithms, International Journal of Current Engineering and Technology (2013) 334-337.

[29] Masud Karim and Rashedur M. Rahman: Decision Tree and Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. Journal of Software Engineering and Applications (2013) 196-206.

[30] Pedro Domingos and Michael J. Pazzani: Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier. International Conference on Machine Learning (1996) 105-112.

[31] Manish Mehta, Rakesh Agarwal and Jorma Rissanen: SLIQ: A Fast Scalable Classifier for Data Mining. International Conference on Extending Database Technology (1996) 18-32.

[32] Rodrigo Coelho Barros, Marcio Porto Basgalupp, Andre C.P.L.F. De Carvalho and Alex A. Freitas: A Survey of Evolutionary Algorithms for Decision Tree Induction, IEEE Transactions on Systems, Man and Cybernetics (2012) 291-312.

[33] J.R. Quinlan: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993).

[34] S. Safavian and D. Landgrebe: A Survey of Decision Tree Classifier Methodology, IEEE Transactions on Systems, Man And Cybernetics (1991) 660 -674.

[35] Arun k pujari: Data Mining Techniques. Universities Press (2004).

[36] http://www.wunderground.com/.

[37] Powers and M. W. David: Evolution: From Precision, Recall and F-Measure to Roc, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies (2011) 37-63.

[38] Kishor Kumar Reddy C, Vijaya Babu B, Rupa C H: SLEAS: Supervised Learning using Entropy as Attribute Selection Measure. International Journal of Engineering and Technology (2014) 2053-2060.

[39] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu B: A Pragmatic Methodology to Predict the Presence of Snow/No-Snow using Supervised Learning Methodologies. International Journal of Applied Engineering Research (2014) 11381-11394.

[40] Chandra, B., Mazumdar, S., Vincent, A. and Parimi, N: Elegant Decision Tree Algorithm for Classification in Data Mining. Proceedings of the Third International Conference on Web Information Systems Engineering (2002) 160"169.

[41] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu: SPM: A Fast and Scalable Model for Predicting Snow/No-Snow. World Applied Sciences Journal (2014) 1561-1570.

[42] Pramote Luenam, Supawadee Ingsriswang, Lily Ingsrisawang, Prasert Aungsuratana, and Warawut Khantiyanan: A Neuro-Fuzzy Approach for Daily Rainfall Prediction over the Central Region of Thailand. International Multi Conference of Engineers and Computer Scientists (2010).

[43] Lily Ingsrisawang, Supawadee Ingsriswang, Pramote Luenam, Premjai Trisaranuwatana, Song Klinpratoom, Prasert Aungsuratana, and Warawut Khantiyanan: Applications of Statistical Methods for Rainfall Prediction over the Eastern Thailand. International Multi Conference of Engineers and Computer Scientists (2010).

[44] Thair Nu Phyu: Survey of Classification Techniques in Data Mining. International Multi Conference of Engineers and Computer Scientists (2009).

[45] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu: SLGAS: Supervised Learning using Gain Ratio as Attribute Selection Measure to Nowcast Snow/No-Snow. International Review on Computers and Software (2015).

[46] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu: ISLIQ: Improved Supervised Learning in Quest to Nowcast Snow/No-Snow. WSEAS Transactions on Computers (2015).

[47] Kishor Kumar Reddy C, Vijaya Babu: ISPM: Improved Snow Prediction Model to Nowcast Snow/No-Snow. International Review on Computers and Software (2015).