Implementation and Comparative Analysis of Rough Set, Artificial Neural Network (ANN) and Fuzzy-Rough classifiers for Satellite Image Classification

Mamta Juneja [1], Ekta Walia[2] , Parvinder Singh Sandhu[3], Rajni Mohana[4]

[1,3,4] Department of Computer Science and Engineering, Rayat and Bahra Institute of Engineering and Bio-Technology, Punjab Technical University, Punjab, India

[2] Department of Information Technology, Maharishi Markendeshwar Univesity, Mullana, Haryana, India.

***Abstract*— Geospatial information we gather through different sensors and from the concepts of the geographic objects, is generally vague, imprecise and uncertain. Also, the imprecision becomes obvious due to the multi-granular structure of the multi-sensor satellite images and that leads to error accumulation at every stage in geo-processing.**

It has been observed that the ground truth data, forming a prime decision system, an essential ingredient for a supervised learning, may itself contain redundant / inconsistent / conflicting information. Moreover, there may be superfluous attributes that warrants a fast mechanism to identify & discard them and at the same time keep the information content compatible to the original data set. Recently the Rough Set Theory - proposed by Zdzislaw Pawlak, has emerged as an effective measure to resolve imprecise knowledge, analysis of conflicts, evaluation of data dependencies and generating rules.

In this study, we have applied the Rough Set Theory, to handle the imprecision due to granularity of the structure of the satellite image. The objective is how the decision system required for any supervised classification, is made consistent and free from superfluous attributes. We compared the results of performing land cover classification of the LISS-III image pertaining to Alwar (Rajasthan) area by the rough set, artificial neural networks, and rough-fuzzy theory. Our findings are that, in the era of internet GIS, time and accuracy is the prime requirement in classification and interpretation of images for any critical application. Rough set and rough-fuzzy theory offer a better and transparent choice to have faster, comparable and effective results.

Keywords-***Roughset theory; Artificial Neural Network (ANN) and Fuzzy-Rough classifiers; Image Classification.***

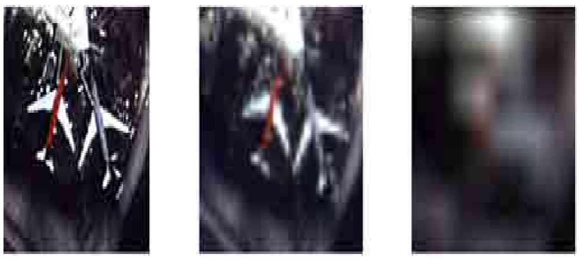
I. INTRODUCTION

InternetGIS, like Intergraph Geo-media Web Map server, ESRI Map Object IMS and MapInfo MapXtreme, initiates a worldwide mechanism for geo-spatial information distribution. Geographic Markup Language (GML) is a schema for modeling, transport and storage of the geographic information. GML is developed as an implementation specification by the Open GIS Consortium to foster data interoperability and exchange between different systems. Also, spatial information, continuously, is being gathered by the Earth Observation Satellites (EOS) around the globe. Both these technologies are working in synergy to help developing faster & efficient apparatus for decision making. The geo-spatial information is received in different windows of the electromagnetic spectrum and at different resolution. This present selective look of the geospatial objects under view of the satellite sensor. Therefore, the totality of capturing the truth / facets of the objects seems to be very difficult.

This implies that at a given set of parameters of observation, we have limited capability to discern two objects. It is equivalent to say that the knowledge generated from the satellite image at a given resolution and spectrum band, is granular. It is, therefore, imperative to have more observational parameters to decompose this granule, i.e. finer view of the objects. The effect is that based on the observational parameters, any two objects, may appear same, whereas, the ground truths about the objects forces us to have different opinion on them. This phenomenon introduces the uncertainty into the information system due to imprecision inducted by the observation system. Therefore, given the nature of digital computation, all data, spatial or otherwise, can be represented at only a finite precision. It should be emphasized that it is not just narrowly concerned with pixel resolution in a raster image, but take a wider view, where any computational spatial data model is seen based upon some resolution structure. The well known triangulated irregular network (TIN) representation and the 'realm' representation of Guting and Schneider, both provide example of representation with respect to particular resolution structure.

The images of Figure-1 show the effect of perception of information at different resolutions/ granularity.

The focus of this study is an analysis of the effect of the granularity on indiscernability relation of objects. A formalism on which reasoning of this kind may be based is provided by the theory of rough sets. The Rough-Set Theory (Pawlak, 1982) is a new tool for discovery relationship hidden in data & expresses them in the natural language of decision rules. It has been applied in multi-criteria decision analysis, allowing the recognition of relationship ol variables (Slowinski, 1993); relevant application of the Rough-Set theory covers field of medicine, engineering, banking, environment management (Pawlak, 2000). Very few applications of the Rough-Set theory cover aspects of water resource planning and management (Chen et.al. 2003), geographic knowledge discovery from the chloropleth maps (Aldridge, 2001). Landuse/landcover classification application of rough-set is a challenging area. The motivation for this paper was to explore the granular knowledge embedded in the geo-spatial information and the reasoning mechanism offered by the Rough Set Theory framework and therefore, deriving a landcover classification



-Meter This one-meter solution picture simulates IKONOS satellite image. It zoomed-in to focus on two planes parked at the course.

Five-Meter This is the same picture, but it has been altered to simulate what you would see in a five-meter resolution image. The airplanes look blurry at this resolution.

25-Meter At 25-meter resolution, this same picture is unreadable. The airplanes are not recognizable.

Figure-1

Rough set concept (Pawlak, 1982) is a new mathematical tool to reason about uncertainty cause by vagueness, imprecision due to granularity in the domain of discourse. The theory is one of the important constituent of the computational theory of perception, generally known by the term -Soft Computing. Perceptions are described from a natural language and are basically imprecise. Soft Computing encompasses the Fuzzy Logic (FL), Artificial Neural Networks (ANN) and Genetic Algorithm (GA) beside Rough Set (RS) theory. Probability theory has its own limitations in dealing with perceptions, because it foundations rests on bivalent logic, that is logic of measurement. Fuzzy logic deals with imprecision and vagueness rather than randomness. ANN is the machinery for learning and adaptation, whereas GA deals with optimization and searching.

Rough set analysis uses only internal knowledge and does not rely on prior model assumption as fuzzy set methods or probabilistic models do. That is, instead of using external numbers or addition parameters it utilizes solely the granularity structure of the given data, expressed as classes of suitable equivalence relations (Duntsch & Gedia, 1998). But that does not mean that rough set data analysis does not have any model assumption - Statistical model behind it is the principle of indifference. However model assumptions are such that we admit complete ignorance of what happens within the region of indiscernability, given by the granularity of information. The RS theory is appropriate for data reduction, discovery of data dependency, evaluating consistency of decision system, discovering data significance, similarity and approximate classification.

The empirical learning system called Learning from Examples based on Rough Sets (LERS), developed at the University of Kansas, has been used for two years by NASA's Johnson Space Center as a tool to develop expert systems of the type most likely to be used in medical decision- making on board the space station Freedom.

The paper is organized into five sections. Following the introduction, a section illustrates the Rough-Set Theory. The third section describes the methodology of implementation. The fourth section reports the results of the case-study implementation. The last section summarizes the important findings.

II. THE ROUGH SET THEORY

The Rough Set theory introduced by (Pawlak, 1982; Pawlak, 1991) The concept of Rough Set theory is based on the assumption that with every object of the universe (U) there is associated a certain amount of information (data, knowledge), expressed by means of some attributes (Q) used for object description. In particular, the decision table illustrates all the possible relationships between the objects (also called decision attributes D) and the corresponding descriptors (condition attributes C) in the form of logical statements" if . . . , then . . ."; the antecedent condition part (if) specifies the value(s) assumed by one o more condition attributes, and the consequence decision part (then) specifies the values assumed by the decision attribute(s). Objects having the same description are indiscernible (similar) with respect to the available information. The indiscernability relation induces a partition of the universe into blocks of indiscernible objects (elementary sets) that can be used as "bricks" to build knowledge about a real or abstract world (Greco, et al., 1999). Any subset X of a universe may be ex- pressed in terms of these elementary sets either precisely (as a union of elementary sets) or ap- proximately only. In the latter case, the subset X may be characterized by two ordinary sets, called lower and upper approximations. The lower approximation of X is composed of all the elementary sets completely included in X (whose elements x, therefore, certainly belong to X):

$$\underline{P}(X) = x \in U: I_P(x) \subseteq X$$

The upper approximation of X is composed of all the elementary sets which have a non-empty intersection with X (whose elements x, therefore, may belong to X):

$$\overline{P}(X) = \bigcup_{x \in X} Ip(x)$$

Where $I_{p(x)}$ represents the indiscernability relation on U with respect to a non-empty subset of attributes $P \subseteq Q$

The difference between the upper and lower approximations constitutes the boundary region of the Rough Set, whose elements cannot be characterized with certainty as belonging or not to X, using the available information. The information about objects from the boundary region is, there fore, inconsistent or ambiguous. For this reason, the number of objects from the boundary region may be used as a measure of vagueness of the information about X.

Moreover, it can be defined the following ratio as accuracy of the approximation of X, $0 \neq X \subseteq U$, by means of the attributes from P:

$$\alpha p(X) = |\underline{P}(X)| / |\overline{P}(X)|$$

The result is $0 \leq \alpha p(X) \leq 1$; if $\alpha p(X) = 1$, X is an ordinary (exact) set with respect to P; if $\alpha p(X) < 1$, X is a rough (vague) set with respect to P.

The structure of data that is central point of our work is represented in the form of *information system* or, more precisely, the special case of information system called *decision table*.

III. IMPLEMENTATION

Alwar, (Figure-2), in Rajasthan, is the study area, which contains good variety of Landuse/landcover classes. A, 512X512, LISS-III image, shown below, is selected for demarcating Landuse/landcover classes namely: Crop, Water, Forest,Sandy area and Fellow land.

The software development platform is MATLAB ver 6. The rough set and artificial neural network (ANN) and Fuzzy-Rough classifiers are implemented in MATLAB on a Pentium III / 800 MHz PC under Windows NT 4.0 operating system NT 4.0 operating system.

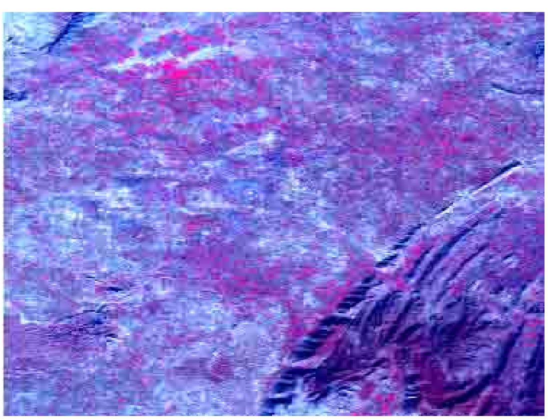
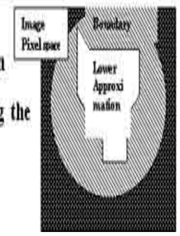


Figure-2 Training Data set:

For supervised classification, a good training set is the key element that has direct consequences on the accuracy of any classifier. Ground truth data, from field survey, corroborate the mapping of training samples on the image. Limited with the image granularity, pure pixel demarcation is a difficult task. This implies that our training data is vague and ambiguous, that may have pixels with same DN values in the spectral bands but have decision class different altogether. The training set when represented in a tabular fashion is termed as decision system as shown below:

Decision System	Conditional Attribute			Decision Attribute
Pixel(x,y)	Band1	Band2	Band3	
x1	68	63	79	crop
x2	46	52	71	Sand
x3	60	66	81	Fallow land

Two types of reduction in the decision table can be carried out, by applying the rough set theory:



i. Removal of redundant tuples in the table and

ii. The superfluous attributes i.e. the conditional attributes, removal of which may not degrade the quality of the information content of the decision system.

This is carried out by finding the Reduct of the information system. In this experiment we have observed that in the 3-band data, all band data is essential, removal of any band drastically affect the decision system quality. Therefore, 3-band attribute is the Reduct of this system. The lower approximation (Figure-3, i.e. the pixel definitely belong to the defined five categories) and upper approximation (i.e. the pixel possibly belong to the categories) are calculated and boundary region is defined. In this study we have collected 4093 pixels of five caegories.

We observed that 368 pixels lies in the boundary region, on which category may not be defined crisply. These ambiguous pixels may be removed from the decision table. This generates a refined decision system.

Rule Generation: The above decision table is nothing but a knowledge representation form and each tuple i.e. row of the table is a rule of the form:

$$\textit{If}\ (Band1_{DN}) \wedge (Band1_{DN}) \wedge (Band1_{DN})\ \textit{then Category}$$

Approximately 3200 rules are obtained, which does not offer a plausible solution. Therefore we descretize the table into close intervals by applying the cut-offs, and assigning new discrete values to the attributes of each tuples. This facilitates a derived decision table with about 86 rules of the form:

If* $(Band1_{DN} = 70) \wedge (Band1_{DN} = 66) \wedge (Band1_{DN} = 86)$ *then (Category = Crop)

*Rough-Set Classification:*Once the rules are extracted from the derived decision system, the whole image is subjected to classification to arrive at a Landuse / Landcover thematic map.

*Rough Fuzzy Classification:*This is achieved through making use of the ambiguous pixels, which rough-set theory is unable to categorize. These pixels are assigned category through membership function, $\mu(x)$ and the pixel with $\mu(x) \geq = 0.5$ are assigned to that category. And again the decision table is update and rules generated through the derived decision table.

*ANN Classification:*Using the original decision system as the training set to the multilayer perceptron architecture of the artificial neural network, we wanted to explore the effect of the inherent ambiguities in the training set over the classification. The network has 3-5-3 architecture of input, hidden and output layer nodes. Network is trained with 4093 pixels from the 3 bands. After 2,00,000 epochs an acceptable error was 0.025. Once the network is trained, the image is subjected to classification.

IV. RESULTS AND DISCUSSION

In this experiment, the image is put to classification by the following three different methods: Rough-Set, Rough-Fuzzy, ANN these three methods are then compared for their accuracy by computing Khat (K) coefficient. This statistics serves as an indicator of the extent to which the percentage correct values of an error matrix are due to "true" agreement versus chance agreement, for example a K value of 0.67 can be thought of as an indication that an observed classification is 67% better than the one resulting from chance. The KHAT statistics is computed as

$$K' = (N\Sigma x_{ii} - \Sigma(x_{i+} \cdot x_{+i})) / (N^2 - \Sigma(x_{i+} * x_{+i}))$$

Where

$i = 1$ to r

r = number of rows in the error matrix

x_{ii} = the number of observations in row i and column i (on the major diagonal)

x_{i+} = total of observation in row i (shown as marginal total to right of the matrix).

x_{+i} = total of observations in column i (shown as marginal total at bottom of the matrix)

N = total number of observations included in matrix.

Error Matrix Rough Classification KHAT (K) = 0.82696

Error Matrix Rough Classification KHAT (K) = 0.82696

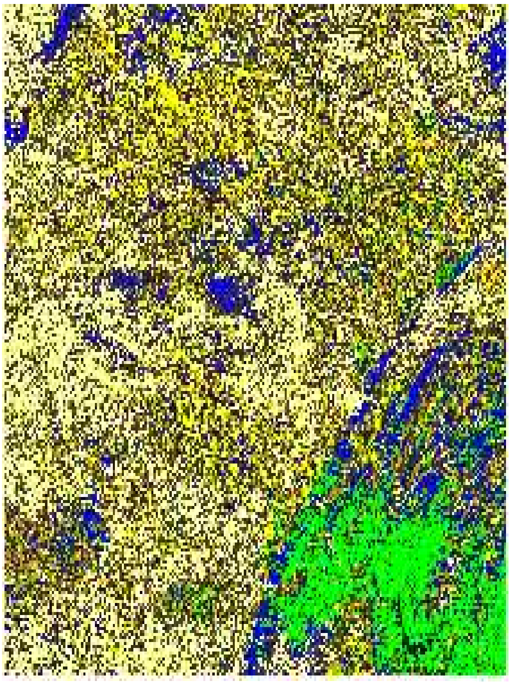
	Water	Sand	Forest	Fallow	Crop	Extra	Row Total
Crop	1497	0	0	0	0	11	1508
Water	0	591	0	0	0	132	723
Sand	0	0	192	0	0	173	365
Forest	0	0	0	414	0	27	441
Fallow	0	0	0	0	853	203	1056
Extra	0	0	0	0	0	0	0
Col Total	1497	591	192	414	853	546	2480

Error Matrix Rough-Fuzzy Classification KHAT (K) = 0.94522

	Water	Sand	Forest	Fallow	Crop	Extra	Row Tot
Crop	1503	0	4	0	1	0	1508
Water	0	627	0	14	82	0	723
Sand	4	0	326	0	35	0	365
Forest	0	3	0	438	0	0	441
Fallow	0	10	14	0	1032	0	1056
Extra	0	0	0	0	0	0	0
Col Tot	1507	640	344	452	1150	0	2480

Error Matrix Evaluating ANN Classification KHAT (K) = 0.89625

	Water	Sand	Forest	Fallow	Crop	Extra	Row Tot
Crop	1480	0	12	0	1	15	1508
Water	0	615	10	7	53	38	723
Sandy area	1	0	290	0	19	55	365
Forest	3	4	6	360	0	68	441
Fallow	0	19	15	0	1017	5	1056
Extra	0	0	0	0	0	0	0
Col Tot	1484	638	333	367	1090	181	2480

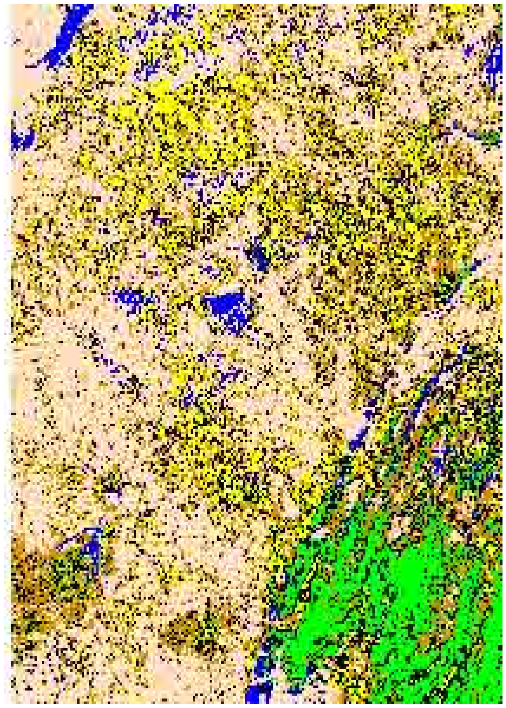


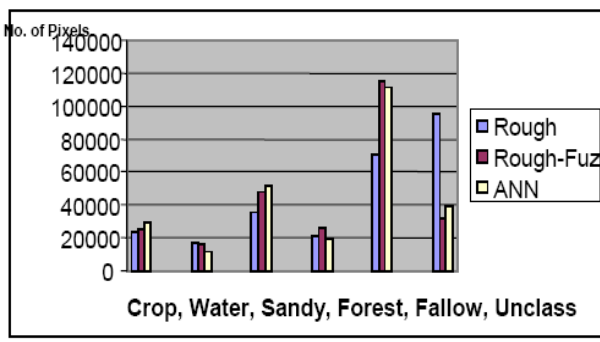
LEGEND

Color	Class	No. of Pixels
Yellow	Crop	23206
Blue	Water	17124
Green	Forest	20794
Brown	Sandy area	35478
Cyan	Fallow land	70423
Black	Unclassified	95119

Rough Classified Image

Rough-Fuzzy Classified Image





ANN Classified Image

Comparison of Classification

	Rough Set	Rough-Fuzzy	ANN
Crop	23206	25330	29211
Water	17124	16230	11301
Sandy area	35478	48171	51907
Forest	20794	26119	18966
Fallow	70423	114940	111215
Unclassified	95119	31354	39544

V. CONCLUSION

The granularity concept is exploited by applying the rough set theory. Classification results give an insight of the computing time and accuracy by which the original image is been classified using Rough set theory, Rough-Fuzzy theory and Artificial Neural Network.

1. Accuracy level of the classified image by these methods is comparable and quite acceptable.

2. Time consumed in training of ANN in reaching permissible accuracy is in days whereas there is no such training required in Rough set and Rough-Fuzzy theory.

3. Rough set and Rough-Fuzzy theory generate rules which are then used to classify the image; hence there is transparency in between unlike ANN.

4. Training of ANN depends totally on dataset used. ANN requires retraining in case dataset is changed. There is no such requirement in Rough Set and Rough-Fuzzy Theory.

5. Once the algorithms are devised then there is no need of human intervention in Rough Set and Rough-Fuzzy Theory Classification, just the dataset file is to be changed. In ANN, threshold for classification is to be set every time as the dataset is changed.

In Nutshell:
In this competitive world of internet-GIS, time and accuracy is the most important requirement in classification and interpretation of images for any critical application. Rough Set and Rough-Fuzzy Theory are better choice than ANN for image classification with comparable result.

REFERENCES

[1] Aldridge, C.H., A Rough-Set based Methodology for Geographic Knowledge Discovery, Proceedings of the 6th International Conference on Geo-Computation, University of Queensland, Brisbane, Australia, 24 - 26 September 2001.

[2] Chen, .C., Chang, N.B. and Chen, J.C., Rough set-based hybrid fuzzy-neural controller design for industrial wastewater treatment. Water Research 37, 95-107, 2003

[3] Choubey,S.K., Deogun, J.S., Raghavan, V.V. and Sever, H., A Comparison of Feature Selection Algorithms in the Context of Rough Classifiers, Proceedings of Fifth IEEE International Conference on Fuzzy Systems, vol. 2, New Orleans, LA, pp. 1122-1128, September 1996.

[4] Duntch,I., and Gediga,G., Uncertainty Measures of Rough-Set Prediction, Artificial Intelligence ,1998, no. 106, pp. 109-137.

[5] Güting, R.H. and Schneider, M.,. Realms: A Foundation for Spatial Data Types in Database Systems. In D. Abel, B.C. Ooi, Advances in Spatial Databases, Proceedings of SSD'93, Singapore, Lecture Notes in Computer Science 692. Springer-Verlag, Berlin, Germany, pp. 14-35, 1993.

[6] Hayri Sever, Raghavan, V.V. and Johnsten, T.D., The State of Rough Sets for Knowledge Discovery in Databases, ICNPAA-98: 2nd International Conference on Nonlinear Problems in Aviation and Aerospace, Cambridge, UK, 1998, Vol. 2, pp. 673-680.

[7] Pawlak, Z.,. Rough sets. Int. Journal of Inf. and Comp. Sci., 11(5): 341-356, 1982.

[8] Pawlak, Z., Rough Sets - Theoretical Aspects of Reasoning about Data, Kluwer, 1991.

[9] Pawlak, Z. . Rough sets and intelligent data analysis. Information Sciences 147, 1-12., 2002.

[10] Polkowski,L., Rough Sets, Mathematical Foundations, Physics-Verlag,New York, 2002.

[11] Worboys, M.F., Imprecision in Finite Resolution Spatial Data, Geoinformatica 2(3): 257-280, 1998.

[12] Worboys, M.F. and Clementine, E., Integration of Imperfect Spatial Information, Journal of Visual Languages and Computing, 12, 61-80, 2001.

[13] Ziarko,W.P., Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer-Verlag ,New York, 1994