ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Rough Set Reduct Algorithm based Feature Selection for medical domain

T.Keerthika[1*], Dr.K.Premalatha[2]
[1]Department of Information Technology, Sri Krishna college of Engineering and Technology, Coimbatore, Tamilnadu, India.
[2]Department of CSE, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India.
***Corresponding author: E-Mail: keerthiglitter@gmail.com**

ABSTRACT

With the real time data, results in increasing in size. Feature selection (FS) approach chooses the most informative features from the original features according to a selection method. Also current methods are inadequate. In particular, this has found successful application in tasks that involve datasets containing huge numbers of features (in the order of tens of thousands), which would be impossible to process further. Rough set theory has been used as such a dataset preprocessor with much success, but current methods are inadequate at finding *minimal* reductions. This paper proposes a new feature selection mechanism based on Ant Colony Optimization and genetic algorithm. Proposed work is applied in the medical domain to find the minimal reducts and experimentally compared with the Quick Reduct, Entropy Based Reduct, and other hybrid Rough Set methods such as Genetic Algorithm (GA), Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) are considered for comparative performance analysis in which the experimental results shows that feature selection is best for minimal reductions.

KEY WORDS: Feature selection, rough set theory, Genetic Algorithm, Particle Swarm Optimization, Ant Colony Algorithm.

1. INTRODUCTION

The solution to the dimensionality reducing problem has been of prior importance and worked in a variety of fields like statistics, pattern recognition, discovery through knowledge and machine learning. Two major techniques done for reducing the various dimensionality inputs are: feature extraction feature selection. The concept behind feature extraction is that the lower dimensionality is used when a primitive space is being mapped onto a new space Principle component analysis, partial least squares technique are the two important approach used by the feature extraction method. The various applications of feature extraction is used in variety of fields that include literature, where image processing, visualization and signal processing plat an important role. Unlike the feature extraction process, feature selection(FS)involves the selection methods of t-statistic, f-statistic, correlation, reparability correlation measure, or information gain which chooses the most pinpointed and informative features from the original one using the methods above. The low accuracy and slow learning difficulties occurs because of redundancy and irrelevancy in the dataset.

Finding the subset of features that are informative enough is NP complete. Heuristic algorithm is administered to invoke a search process through feature space. The most intricacy of learning an algorithm and defining its accuracy are some issues used for evaluating the selected subset.

The rough set (RS) (Pawlak, 1982; 1991; 1993) is a helping tool that reduces the crisis of input dimensionality and finds a better solution for correcting the vague and uncertain datasets. The reduction of attributes is meant for data dependencies.

The RS theory splits the dataset into few equivalent (indiscernibility) classes, and approximates uncertain and vague concepts resting on the partitions. The purpose of approximation is mainly used to analyze the measure of dependency. The measure of dependency is regarded as a heuristic guide for the approach of FS.

Proper approximations of concepts are very essential to obtain a significant measure which makes the initial partitions to be vital in this matter. For the sum of discrete dataset, finding the indiscernibility classes are feasible ,but in the case of real-valued attributes, one can't be sure if the two objects mentioned as the same, or by what relation they are same using the above mentioned indiscernibility relation. A team of research persons extended this RS theory by the usage of tolerant or similarity relation (termed tolerance-based rough set). The likeliness measure between two objects is delineated for all attributed based on the distance function. When it is quantified by exceeding a similarity threshold value, the objects are said to be similar.

The important and challenging job is to identify the finest threshold boundary. The subsets of attributes (features) are referred by the feature selection process. On the basis of some criteria, the objects are being classified In advance, the class of each object is given its easy to group the objects in to their particular classes. This type is called supervised classification. No class is attached to any object and by grouping them on the basis of some particular similarity based method like type, size, colour or similar type of attributes. This type of classification is known as unsupervised. Clustering is one of the important method of unsupervised classification. The main purpose of the feature selection is to find the similar features, its use to delete the irrelevant or unnecessary features. This mainly reduces the burden on various learning models and as a result of it will be helpful in building the learning model. The use of feature selection depends on two types of folds: it will decrease the computation time of algorithm and next, it will increase the accuracy of the model. For the past one decade, feature selection has been studied. (Khoo, 1999; Roman, 2003; Shoji Hirano, 2005).

Related Works

Rough-Set Oriented Incremental APPROACH: In this approach (Chen, 2013) the variable precision rough-set model (VPRS) usually differ under a certain dynamic information system environment. Therefore by utilizing a previous data structures, approximation is using the incremental updating approximations by utilizing previous data structures. In this paper a new incremental approach is being used for VPRS by updating approximations whereas objects present in the information system will be alter dynamically. It discusses the granulation information properties and approximations under the environment which is dynamic in nature while the objects may alter dynamically. Finally, a wide experimental assessment validates the efficiency of the proposed method for dynamic maintenance of VPRS approximations.

Rough set theory used in novel dynamic incremental rules extraction algorithm: In this paper (He-Ping Yan, 2005) a novel incremental rules extraction algorithm which is called "RDBRST" (Rule Derivation Based on Rough Set And Search Tree) is proposed. This is belongs to Width first heuristic search algorithm.

Based on this algorithm (Liang, 2012) the incremental rules are extracted and the existing rule set is updated based on this algorithm (Liang, 2012). Incremental Induction Decision Rules are obtained from Dominance-Based Rough Approximations. It is extended to handle preference-ordered domains of attributes (called criteria) within Variable Consistency Dominance-based Rough Set Approach. It deals, moreover, with the crisis of missing values in the data set.

For medical applications, the algorithm has been designed which require: (i) medical experience is being represented with the careful selection of the set of decision rules and (ii) an easy update of these decision rules because of data set evolving in time, and (iii) The collective decision rules are not only a high predictive capacity but also a thorough explanation of a proposed decision. To gratify all these requirements, therefore we propose an incremental algorithm for satisfactory set of decision rules to induction and a post-processing technique on the generated set of rules.

A Distance Measure Approach to Exploring the Rough Set Boundary Region of a Distance Measure Approach for Attribute Reduction method: This paper examines a information gathered from both the lower approximation dependency value and a distance metric which uses a rough set FS technique and it considers the amount of objects which present in the boundary region and the space of those objects from the lower approximation In rough set feature selection, the use of this measure can result in smaller subset sizes than those obtain from the dependency function alone. This explains that there is much valuable data to be extracted from the boundary region (Parthalain, 2010).

Incremental Learning of Decision Rules Based on Rough Set Theory: In this paper, based on the rough set theory, the concept of ∂-indiscernibility relation is put forward in order to transform an incompatible decision table to one that is consistent, called ∂-decision table, as an initial preprocessing step (Hu, 2011).

Then, the ∂-decision matrix is constructed. On the basis of this, by means of a decision function, incremental learning algorithm rules are presented. The following algorithm can also incrementally change some mathematical measures of a rule.

Proposed System: The Proposed system idea is to develop a new novel feature selection mechanism based on Ant Colony Optimization method to battle this difficulty. Original rough set-based approach is being modified to produce a new entropy. These are applied to the problem of finding minimal rough set reducts, and evaluated experimentally.

- Feature selection methods, the Importance Score (IS) which is related to greedy-like search method and a genetic algorithm-based (GA) method, in order to better understand.
- It also presents a new entropy based modification of the original rough set-based approach. These are applied to the problem of finding minimal rough set reducts, and evaluated experimentally.
- This proposed work is useful in the medical domain to identify the experimentally with the Quick Reduct, minimal reducts and Entropy Based Reduct, and other hybrid Rough Set methods such as Genetic Algorithm (GA) method , Ant Colony Optimization (ACO) method and Particle Swarm Optimization (PSO) method.

Advantages: For reducing the complexity of the problem we have to reduce the dimensionality of the attributes and allows researchers to focus more clearly on the relevant attributes.

- Simplifying data explanation may facilitate physicians to create a prompt diagnosis.

Having fewer features means less data need to be collected, result in consuming more time and it's also costly. The proposed work can be explained with the help of the system flow diagram in Figure-1as below,

Figure.1. Diagram ATIC APPROACH OF System flow

Feature Selection APPROACH: The main objective of FS (feature selection) is to pick out from the problem domain the minimal feature subset, such that it represents the unique features with an outstanding accuracy given. FS plays a predominant role in real world issues and since the problem is irrelevant, noisy and misleading features of the data that are plenty in numbers.

In case of reducing these irrelevant data's, the method of learning from the data technique can be useful for the users. The work of FS is to search a feature subset that is the most optimal (that varies depending on the trouble to be solved) from the given n size of feature set by competing with candidate subset of size n. But this method is not possible even though an exhaustive methodology is used.

The searching for the datasets are done randomly regulate to cease this complexity. But in that case the extent of getting best results is drastically brought down. The degree to which a feature subset or may be a feature may be useful is based on two important factors: 1. relevancy 2.redundancy.

Relevancy depends on its capability to predict the decision feature(s), if not the datas are said to be irrelevant. Redundant feature must be correlating with other features. So an optimal search to find the best feature subset must be its ability to have a correlation between the decision features but must not be correlating apart from that.

When it comes to subset minimalist and subset suitability, a tradeoff occurs with these non-exhaustive techniques and it becomes likely to choose between the two so that one will benefit over the other. Choosing this optimality is a challenging one. Involving situations when the inspection of many features is not possible, it is better to switch to subset feature that is much smaller and have lesser accuracy amount.

For instance, classification rate that is a feature of modeling accuracy should be very high when the user is using selected features, by taking the expense of a non-minimal feature subset. On the basis of evaluation procedure, there are two important classification in feature selection algorithm. The filter approach is one of the first approach where the FS works independently and which is a separate pre-processor to any type of learning algorithm. This approach is being practical in all the domains as they are very effective in filtering all irrelevant attributes before induction and no any specific induction algorithm is used. The next is wrapper approach which involves tying up of evaluation procedure to a task of any type of learning algorithm as in the case of classification. This method employs an accuracy estimation that can search through the spaces of feature subset with the help of an induction algorithm that measures suitability of subsets.

Wrapper are the ones that produce good results when compared by other but faces the difficulties of a break down when huge number of features are fed into it and also makes it expensive to run since the learning algorithm used, that invokes the problem when large datasets are used.

Rough Set-Based Feature Selection APPROACH: Rough set theory (RST) can be discovering data dependencies. They can curtail the attributes found in the dataset by using only the data and not any additional information. This is a topic in trend that lures many researches to work on it and has been applied in various domains and fields over the past decade. Using RST it is likely to search for the right subset that is often termed as reduct when discretized attribute values are given in a dataset; the remaining attributes can be taken out from the dataset with minimal loss of information.

From the view of dimensionality, the one with the predictive nature of class attribute are often called the informative feature. Finding rough set reducts are put into two approaches: one for to estimate the degree of dependency and the other for discemibility matrix consideration. This section describes the fundamental ideas behind both of these approaches. To finding rough set reducts there are two main approaches: those that consider the degree of dependency and those that are anxious with the discernibility matrix. This section explains the basic ideas behind both the approaches.

To illustrate the operation of these, an example dataset (Table 1) will be used. Table 1

Table 1. An example dataset

xU	a	b	c	d	e
0	0	0	2	2	0
1	1	2	1	1	2
2	0	1	0	1	1
3	1	2	1	0	1
4	0	1	2	0	1
5	2	1	0	2	2
6	1	2	1	1	2
7	1	1	1	0	1

Rough Set Attribute Reduction: Central to Rough Set Attribute Reduction (RSAR) is the concept of indiscernibility. Let I = (U, A) be an information system, where U is a non-empty set of finite objects (the universe) and A is a non-empty finite set of attributes such that $a:U \rightarrow Va$ for every $a \in A$. Va is the set of values that attribute a may take. With any $P \in A$ there is an associated equivalence relation IND (P).

Information and Decision Systems: An information system can be viewed as a table of data, consisting of objects (rows in the table) and attributes (columns). in medical datasets, for example, objects are identified as a patients and blood pressure are some of the measurements such as blood pressure, form attributes. The attribute value for a particular patient is their specific reading for that measurement. Throughout this paper, the terms attribute, feature and variable are used interchangeably.

An information system may be extended by the inclusion of decision attributes. Such a system is termed a decision system. For example, the medical information system mentioned previously could be extended to include patient classification information, such as whether a patient is ill or healthy. A more abstract example of a decision system can be found in table 1. Here, the table consists of four conditional features (a; b; c; d), a decision feature (e) and eight objects.

Ant Colony Optimization Used For Feature Selection: Swarm Intelligence (SI) is the property of a system whereby the collective behaviors of simple agents.

Without centralized control or the provision of a global model it is possible to explore collective (or distributed) problem solving Particle Swarm Optimization is one area of interest in SI, a population-based assumptive optimization technique. Here, the system is initialized with a population of random solutions, called particles.

In SI , Ant Colony Optimization (ACO) is one of the most important area. In nature ,real ants are capable of indentify the shortest route between a food source and their nest without the need of visual information and hence possess no global world model, adapting to changes in the environment. Over a period of time, the deposition of pheromone is the main factor in enabling real ants to find the shortest routes. In this chemical, each ant probabilistically prefers to follow a direction. In excess of time the pheromone decays, which results in much lesser pheromone on lesser popular paths.

Provided that over time the shortest route will have the higher rate of ant traversal, reinforced will happen and it will be diminished until all ants follows the shortest path as a same. During this time, the rates of traversal of ants over various the short paths will be roughly the same, which results in these paths being maintained while the others are ignored. In addition, if there will be a change to the environment (e.g. in a shortest path many obstacle may happen), the ACO system responds to this and will produce a new solution.

ACO is particularly attractive for feature selection as there seems to be no heuristic that can guide search to the optimal minimal subset every time. Additionally, throughout the search space it can be the case that ants discover the best feature combinations as they proceed This section discusses how ACO may be used to the difficult problem of identify optimal feature subsets and, in particular, fuzzy-rough set- based reducts. The ACO-suitable problem is formulated from feature selection task. ACO represents the problem as a graph where the features are represented as a node. In this setup – a is consider as the ant node and next path choice will be chosen in a form of dotted lines. Next feature b is choose based on transition rule, followed by c and d.

Genetic Algorithm For Feature Selection Method: One of the main search heuristic approach is Genetic algorithm (GA), it is mainly used to produce a optimization solutions to the problem, thus the following the techniques such as inheritance, mutation, selection, and crossover. In the genetic algorithm,

- A string population as chromosomes, which encode candidate solution to an optimization problem is taken.
- A proper fitness function is then constructed, and the fitness of the current population is evaluated.
- Two fittest chromosomes are chosen as the parents and (a) crossing over will happen or (b) a production of new children done by the mutation of a parent. Then the new population is produced again.
- Again the fitness function for the new population is estimated.
- The process recurs as long as the fitness function keeps on improving or until the termination condition is attained.

The algorithm of a genetic programming begins with the population which is a collection of arbitrarily created individuals. Each individual represents a potential solution which is further represented as a binary tree.
Each binary tree is build by all the possible compositions of the sets of functions and terminals. A tree fitness value is calculated by a suitable fitness function. In order to find the fitness value, a set of individuals with better fitness will be selected. Using genetic operators, these individuals are used to generate new population in next generation with genetic operators.

Pso for feature selection method: Evolutionary computation technique such as Particle swarm optimization (PSO) is one of the best feature selection method. The original version of PSO was formed from the modified initial simulation. To produce the standard ISO, later she introduced inertia weight into the particle swarm optimizer.

A population of different results which is also called ‘particles’ was initialized by PSO. In S-dimensional space each particle is act as a point. The *I*th particle is represented as $a_i=(a_{i1},a_{i2},a_{i3}\ldots a_{in})$. The best previous position (pbest, the position giving the best fitness $b_i=(b_{i1},b_{i2},\ldots.b_{in})$. The symbol ‘gbest’ is to represent the index of the best particle among all the particles in the population. $E_i=(e_{i1},e_{i2},\ldots.e_{in})$ represents the rate of the position change(velocity) for the particle i. The following equation manipulates the particles value) of any particle is recorded and represented

$e_{id}=w*e_{id}+c1*rand()*(b_{id}-a_{id})+c2*Rand()*(e_{gd}-a_{id})$ $a_{id}=a_{id}+e_{id}$

Where $d = 1, 2, \ldots, S$, w is the inertia weight.

According to the generation iteration time changing will occur. The global and local exploration is provided by suitable selection of inertia weigh and also it is balanced, results is to find a sufficiently optimal solution for a problem. Acceleration constants such as c1 and c2 , which represents the pbest and gbest positions.

In target regions, the result are very high, then the low values allow particles from target regions to roam. Range is specified by the two random functions are rand () and Rand ().On each dimension, particle’s velocities are imperfect to maximum velocity, Vmax. Where as if Vmax is one of the good solution for a high particles.

The “flying particles” act as a first part of equation which produces a degree of memory as a new search space areas. The “cognition” act as a second part which represents the private thinking of the element.
The third party which provides the union among the particles and also known as “social”.

Then the PSO defines a particle’s new velocity based on its previous velocity and the distance is calculated based on the current location from its own best knowledge (position) and the group’s best experience. Accordingly, the new position will be identified by the particles flies.

2. EXPERIMENTAL STUDY

The concert of the reduct calculation approaches discussed along with this paper has been tested with different medical datasets obtained from UCI machine learning data repository, (Liang, 2012) to evaluate the performance of proposed algorithm. Weka tool is being used for experimental purpose. Table 2 shows the facts of datasets used in this paper.

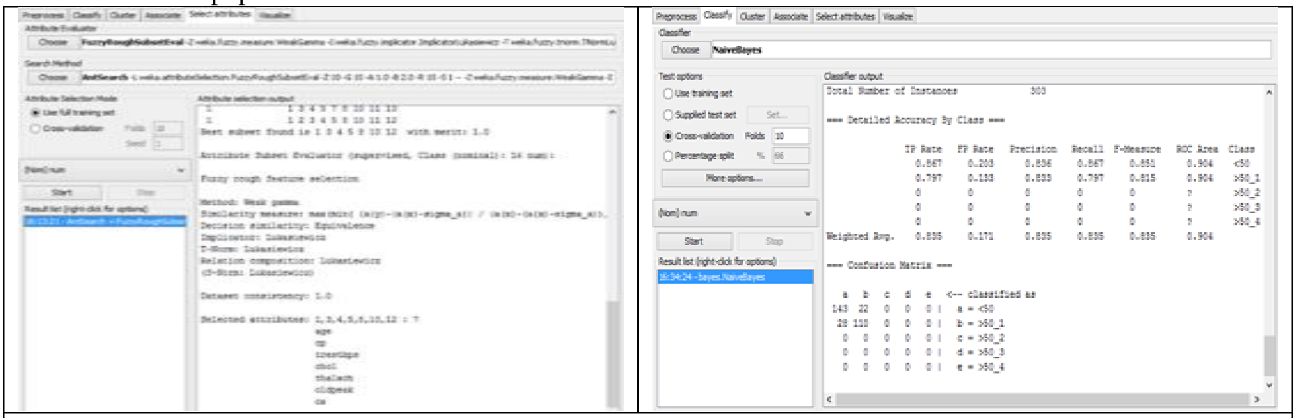

Fig.2. Sample screen shots

Table.2. Detail of Data Sets Used For Experiment

Data Set Name	Total Number of Instances	Total Number of Features	Feature Reduction
Cleveland Heart	303	14	7
Lung Cancer	32	57	5

Performance analysis: By obtain the optimal data reductions here, this paper explains three types of optimization algorithms such as GA , PSO Algorithm, ACO algorithm were used to examine the performance as follows,

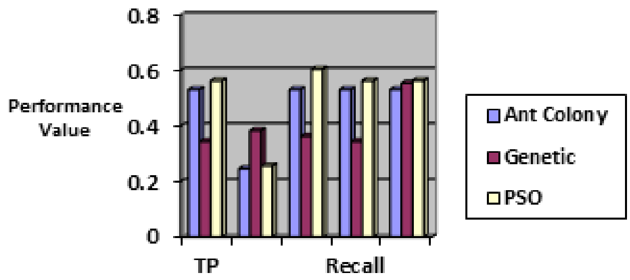

Fig.3. Performance analysis

3. CONCLUSION AND FUTURE WORK

The most valuable preprocessing technique for applications involving large amount of data. It mainly deals with the problem of selecting minimal attribute set that are most predictive to represent the original attributes in data set. This paper discussed the Pros and Cons of various existing feature selection techniques.

The major preprocessor tool for feature selection is Rough Set Reduct algorithm. This paper begins with the basic concepts of rough set theory algorithm and explains basic techniques: Quick Reduct. The minimal reduct set can be produced by this method. The swarm intelligence methods have been used to guide this method to discover the minimal reducts. The fundamental concepts of rough set theory and its basic techniques are explained in this paper : Quick Reduct. These methods can produce close to the minimal reduct set. To find the minimal reducts, the swarm intelligence methods have been used.

Here three different computational intelligence method based reducts: Genetic algorithm, PSO and Ant colony optimization. Though these methods are performing well, there is no consistency since they are dealing with more random parameters. All these methods are analyzed using medical datasets. Experimental results on different data sets have shown the efficiency of the proposed approach.

Comparative performance analysis in which the experimental result shows that feature selection is best for minimal reductions. when compare to other optimization algorithm, PSO algorithm produces higher performance value. As revealed in the results, our proposed method exhibits consistent and better results than the other methods. Future work of this is as comparing the results with some other evolutionary algorithm and performing disease prediction.

REFERENCES

Amit Saxena, Leeladhar Kumar Gavel, Madan Madhaw Shrivas, Rough Sets for Feature Selection and Classification: An Overview with Applications, International Journal of Recent Technology and Engineering (IJRTE), 3 (5), 2014, 2277-3878.

An Liping, Incremental learning of decision rules based on rough set theory, Proceedings of the 4th World Congress on Intelligent Control and Automation (Cat No 02EX527) WCICA-02, 2002.

Chen H.M, Li T.R, Ruan D, Lin J.H, and Hu C.X, A Rough-Set Based Incremental Approach for Updating Approximations under Dynamic Maintenance Environments, IEEE Trans. Knowledge and Data Eng., 25 (2), 2013, 274-284.

Han J, and Kamber M, Data mining concepts and techniques, Academic Press, 2001.

He-Ping Yan, A Novel Dynamic Incremental Rules Extraction Algorithm Based on Rough Set Theory, International Conference on Machine Learning and Cybernetics, 2005.

Hu Q.H, Yu D.R, Pedrycz W, and Chen D.G, Kernelized Fuzzy Rough Sets and Their Applications, IEEE Trans. Knowledge and Data Eng., 23 (11), 2011, 1649-1667.

Khoo L.P, Tor S.B, and Zhi L.Y, A Rough-Set-Based Approach for Classification and Rule induction, Int J Adv Manuf Technol., 15, 1999, 438-444.

Kryszkiewicz M, and Lasek P, FUN: Fast Discovery of Minimal Sets of Attributes Functionally Determining a Decision Attribute, Trans. Rough Sets, 9, 2008, 76-95.

Li M.Z, Yu B, Rana O, and Wang Z.D, Grid Service Discovery with Rough Sets, IEEE Trans. Knowledge and Data Eng., 20 (6), 2008, 851-862.

Liang J.Y, Wang F, Dang C.Y, and Qian Y.H, An Efficient Rough Feature Selection Algorithm with a Multi-Granulation View, Int'l J. Approximate Reasoning, 53, 2012, 912-926.

Pang J.F, and Liang J.Y, Evaluation of the Results of Multi- Attribute Group Decision-Making with Linguistic Information, Omega, 40, 2012, 294-301.

Parthalain N, Shen Q, and Jensen R, A Distance Measure Approach to Exploring the Rough Set Boundary Region for Attribute Reduction, IEEE Trans. Knowledge and Data Eng., 22 (3), 2010, 305-317.

Pawlak Z, Rough Sets, International Journal of Com- puter and Information Sciences, 11, 1982, 341–356.

Pawlak Z, Rough Sets: Present State and the Future, Foundations of Computing and Decision Sciences, 18, 1993, 157–166.

Pawlak Z, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.

Qian Y.H, Liang J.Y, Pedrycz W, and Dang C.Y, Positive Approximation: An Accelerator for Attribute Reduction in Rough Set Theory, Artificial Intelligence, 174, 2010, 597-618.

Rechar O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, 2nd ed., Wiley, 2001.

Roman W. Swiniarski, Andrzej Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Letter, 24, 2003, 833-849.

Shoji Hirano, Shusaku Tsumoto, Rough representation of a region of interest in medical images, International Journal of Approximate Reasoning, 40, 2005, 23-34.

Suguna N, and Dr. Thanushkodi K, A Novel Rough Set Reduct Algorithm for Medical Domain Based on Bee Colony Optimization, Journal of computing, 2 (6), 2010, 2151.

Wei W, Liang J.Y, Qian Y.H, Wang F, and Dang C.Y, Comparative Study of Decision Performance of Decision Tables Induced by Attribute Reductions, Int'l J. General Systems, 39 (8), 2010, 813-838.

Zhao S.Y, Tsang E.C.C, Chen D.G, and Wang X.Z, Building a Rule-Based Classifier-a Fuzzy-Rough Set Approach, IEEE Trans. Knowledge and Data Eng., 22 (5), 2010, 624-638.