Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 47 (2015) 374 – 382

Procedia Computer Science

Optimistic Multi-Granulation Rough set based Classification for Medical Diagnosis

S. Senthil Kumar[a], H. Hannah Inbarani[b]

[a]Research Scholar, [b]Assistant Professor,
Department of Computer science, Periyar University, Salem-636011
pkssenthilmca@gmail.com, hhinba@gmail.com

Abstract

Medical analysis has been approached by several machine learning methods for many years.Pattern recognition algorithms are capable of improving the quality of prediction, early diagnosis of diseases and disease classification. The classification complications in medical area are solved based on the outcome of medical analysis or report of medical treatment by the medical specialist. This research focuses on applying Rough set based data mining techniques for medical data to discover locally frequent diseases. This work applies Optimistic Multi-granulation rough set model (OMGRS) for medical data classification. Multi-granulation rough set provides efficient results than single granulation rough set model and soft rough set based classifier model. The results of applying the OMGRS methodology to medical diagnosis based upon selected information. The performance of the proposed optimistic multi granulation Rough set based classification is compared with other rough set based (RS), K^{th} Nearest Neighbor (KNN) and Back propagation neural network (BPN) approaches using various classification Measures.

Peer-review under responsibility of organizing committee of the Graph Algorithms, High Performance Implementations and Applications (ICGHIA2014)

Keywords:Rough set; Optimistic Multi-granulation rough set; medical data; classification; comparative analysis of classification measures

1. Introduction

Rough set introduced by Pawlak, is a mathematical tool for dealing with uncertainty or incomplete data and knowledge [5, 6, 7, 21]. Rough Set theory has become a valuable tool in the resolve of several problems, such as: representation of inexact or vague data; data investigation; estimate of excellence and ease of use data with respect to reliability; proof of identity and evaluation of data dependency; reasoning based an uncertain and reduct of

*S.Senthil Kumar Email:pkssenthilmca@gmail.com

Peer-review under responsibility of organizing committee of the Graph Algorithms, High Performance Implementations and Applications (ICGHIA2014)
doi:10.1016/j.procs.2015.03.219

uncertainty data [19]. The scope of rough set applications used these days is much comprehensive than in the earlier, essentially in the varieties of remedy, investigation of dataset features and procedure control. In the past 15 years, numerous extensions of the rough set model have been offered in terms of many requirements, such as the, the rough set model based on tolerance relation, equivalence relation and reflexive relation, the soft rough set model, the fuzzy set model, the rough soft set model, the fuzzy rough set model and the rough fuzzy set model [2, 20]. In the vision of granular computing, a universal concept considered by a set is all the time characterized via the so-called upper approximations and lower approximations under a single granulation, i.e., the idea is presented by known data talk into from a single equivalence relation (such as tolerance relation, reflexive relation and equivalence relation) on the universe. Multi-granulation rough set approximations are defined by using multiple equivalence relation on the universe [3, 8, 9, 10, 12, 15, 16, 17, 18 21, 22, 26, 27]. In Example 1, the approximation of a set is discussed by multiple features (two attributes) using with multiple equivalence relations on the universe, that is the target model is defined by two granulation approximation spaces. This paper discusses how optimistic multi-granulation rough set method can be utilized for the analysis of medical data [4, 11, 13, 14, 24, 25], and for generating classification rules from a set of observed samples of the medical data. This paper is organized as follows. Section 2 describes theoretical concepts of rough set and multi-granulation rough set data analysis. Section 3 describes the overall structure of methodology. Section 4 describes proposed algorithm, which are related to the work and rule generation algorithm is presented. Section 5 describes data set information. Experimental results are reported and Comparison with rough set, KNN, SVM and BPN classification algorithms are given in Section 6. Section 7 describes discussion about experimental results. Finally, conclusion is discussed in Section 8.

2. Preliminaries

2.1 Rough Sets

Rough set model was initiated by Pawlak for dealing with ambiguity and granularity in information and knowledge systems. This concept handles the approximation spaces of an arbitrary subset of a universe by two definable or noticeable subsets called lower approximation and upper approximation. It has been effectively applied to system learning methods, intelligent methods, inductive reasoning, pattern recognition algorithms, image processing applications, signal analysis, data discovery, decision analysisand etc.., [5, 6, 7, 21].

Definition 1:Let Rbe an equivalence relation on U. The pair (U, R) is called a Pawlak approximation space. The equivalence relation $R$is often called an indiscernibility relation. Using the indiscernibility relation R, one can define the following two rough approximations:

$$R_*(x) = \{x\epsilon U: [x]_R \subseteq \mathrm{X}\} \quad \text{----} \quad (1)$$

$$R^*(x) = \{x\epsilon U: [x]_R \cap X = \Theta\} \quad \text{----} \quad (2)$$

$R_*(x)$ and $R^*(x)$ are called the Pawlak lower approximation and the Pawlak upper approximation of X, respectively[5, 6, 7, 21].

2.2 Optimistic Multi-Granulation Rough sets

Rough set models are based on single granulation; they also called the single equivalence rough sets. In the optimistic multi-granulation rough set (OMGRS), the objective is approached through the multiple equivalence relation. In lower approximation, the word optimisticis used to precise the idea that in multiple independent granulations, we need only at least one of theequivalence to satisfy with the inclusion condition between data granule and target concept. The upper approximation of OMGRS is defined by the complement of the lower approximation [3, 8, 9, 10, 12, 15, 16, 17, 18, 21, 22, 26, 27].

Definition 2: Let S = (U, AT, D, f) is called a complete target information system, where AT is called the conditional attributes and D is called the decision attribute, $\hat{A}$,$\hat{B}$be two partitions on the universe U, and X⊆U. The lower approximation and the upper approximation of X in U are defined by the following:

$$\underline{X}\hat{A} + \hat{B} = \{x: \hat{A}(x) \subseteq \mathrm{X} \text{ or } \mathrm{x}: \hat{B}(x) \subseteq \mathrm{X}\} \quad \text{-----} \quad (3)$$

$$\bar{X}\,\widehat{A}+\widehat{B} = \sim\underline{(\sim X)}\,\widehat{A}+\widehat{B} \qquad \text{-----} \qquad (4)$$

*Example: 1*Table 1depicts a sample information system contains some data objects. A1 and A2 are the conditional attributes of the system, whereas D is the decision attribute.

Table 1: A sample information system

U	A1	A2	D
e1	1	1	0
e2	2	1	1
e3	2	2	0
e4	1	1	0
e5	2	2	0
e6	2	3	1
e7	3	2	1

Let X = {e2, e6, e7}. Three partitions are induced from Table 1 as follows:

$$\widehat{A1} = \{\{e1, e4\}, \{e2, e3, e5, e6\}, \{e7\}\}$$
$$\widehat{A2} = \{\{e1, e2, e4\}, \{e3, e5, e7\}, \{e6\}\}$$
$$\widehat{A1\cap A2} = \{\{e1, e4\}, \{e2\}, \{e3, e5\}, \{e6\}, \{e7\}\}$$

Optimistic Multi-granulation Lower approximation: The lower approximation of a goalnotion in complete sample data objects using multiple equivalence relations is defined as follows:

$$\underline{X}\widehat{A1}+\widehat{A2} = \{x:\widehat{A1}(x) \subseteq \text{X or x}:\widehat{A2}(x) \subseteq \text{X}\} \text{ ----- } (5)$$

For example, for Table 1

$$\underline{X}\widehat{A1}+\widehat{A2} = \{e7\}\cup\{e6\} = \{e6, e7\}$$

Optimistic Multi-granulation Upper approximation: Multi-granulation upper approximation is a reverse of multi-granulation lower approximation.

$$\bar{X}\,\widehat{A1}+\widehat{A2} = \sim\underline{(\sim X)}\,\widehat{A1}+\widehat{A2} \qquad \text{------} \qquad (6)$$

For example for Table 1,

$$\bar{X}\,\widehat{A1}+\widehat{A2} = \{e1, e2, e3, e4, e5, e6, e7, e8\}\cap\{e2, e3, e5, e6, e7\}\}$$
$$= \{e2, e3, e5, e6, e7\}$$

But,Pawlak's rough set model based on lower and upper approximations are as follows:

$$\underline{X}\widehat{A1\cup A2} = \{Y \in \widehat{A1\cup A2}: Y \subseteq \text{X}\} = \{\text{e2, e6, e7}\}$$
$$\bar{X}\widehat{A1\cup A2} = \{Y \in \widehat{A1\cup A2}: Y\cap\text{X} \neq \Theta\} = \{e2, e6, e7\}$$

3. Methodology

Figure 1 show over all structure of classification. First stage of classification is data gaining. Data gaining is the process of taking data which should be acceptable to the computing device for further processing. Data gaining is typically made by devices, digitizing mechanism and scanners. Second stage is data analysis. Later data gaining, the task of analysis begins. During data analysis step, the learning about the data takes place and information is collected about the different actions and object classes available in the data.

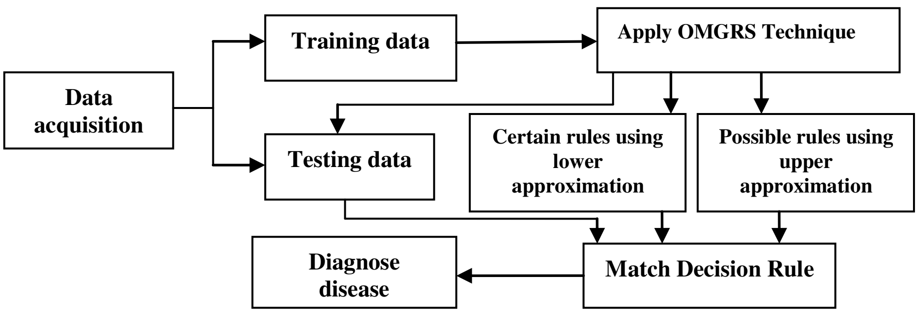

Figure 1: Proposed classification structure

This knowledge about the data is used for further processing. Data set presented to a classification is separated into two sets: training set and testing set. The efficiency of classifier is checked by offering testing set to it. Proposed algorithm is depicted in Table 2. During third stage, multi-granulation rough set classification is applied for training data. Its fortitude is to decide the group of new data on the basis of knowledge generated from certain rules and possible rules. In the next step, matching decision rules are applied for test data. Finally classification measures are applied to evaluate the performance of various classification approaches for disease diagnosis.

4. **Proposed Algorithm**

Optimistic Multi-granulation Rough set based classification approach is presented in Table 2.In this approach, optimistic multi-granulation rough set lower approximation of the given data set based on Decision class X are constructed in step 1. In the second step, optimistic multi-granulation rough set upper approximation of the given data set based on Decision class X are constructed. In the third step, certain rules are generated based on OMGRS lower approximation. In the fourth step, possible rules are generated based on OMGRS upper approximation.

Table: 2 Proposed algorithm

Proposed Algorithm: Optimistic Multi-granulation Rough set based classification

Input:Given Dataset with conditional attributes 1, 2, … , n-1 and the Decision attribute n.
Output: Generated Decision Rules
Step 1: Construct the Optimistic multi-granulation rough set based lower approximation for the given data set

$$\underline{X}\hat{A}+\hat{B}=\{x:\hat{A}(x)\subseteq \mathrm{X} \text{ or } x:\hat{B}(x)\subseteq \mathrm{X}\} \quad ---- \quad (7)$$

Step 2: Construct the Optimistic Multi-granulation rough set based upper approximation for the given dataset

$$\overline{X}\,\widehat{A}+\hat{B}=\sim\underline{(\sim X)}\,\widehat{A}+\hat{B} \quad ----- \quad (8)$$

Step 3: Generate the certain rules using Optimistic Multi-granulation rough set based lower approximation.
Step 4: Generate the possible rules using Optimistic Multi-granulation rough set based upper approximation.

The decision rules generated using proposed algorithm for the example presented in Table 1 is given in Table 3.

Table 3: Example for proposed work

A sample of the data set is an example 1 in order to extract the rules.
Input: Conditional attributes A1 and A2. Decision attribute X.
***Output*:** Generate decision rules
Step 1: Construct the OMGRS lower approximation given data in table 1.

$$\underline{X}\widehat{A1}+\widehat{A2}=\{e6,e7\}$$

Step 2: Construct the OMGRS upper approximation given data in table 1.

$$\overline{X}\,\widehat{A1}+\widehat{A2}=\{\text{e2, e3, e5, e6, e7}\}$$

Step 3: Generate certain rules using OMGRS lower approximation
If A1= 3 and A2 = 2 = > D = 1, If A1= 2 and A2 = 3 = > D = 1
Step 4: Generate possible rules using OMGRS upper approximation
If A1= 2 and A2 = 1 =>D =1, If A1= 2 and A2 = 2 = > D = 0, If A1= 2 and A2= 2 = > D = 0, If A1= 3 and A2 = 2 = > D = 1, If A1= 2 and A2 = 3 = > D = 1

5. Data Description

Datasets are taken from UCI- repository. Five types of diseases, such as breast cancer, liver disorder, Pima indian diabetes, heart diseases (echocardiogram) and hepatitis were considered for the analysis of proposed classification approach [24, 25]. The breast cancer contains 562 instances, 32 attributes and two classes. Such as benign and malignant.The attributes are retrieved from a digitized image of a breast form. They refer tophysical appearance of the cell nuclei present in the image. The liver disorder contains 345 instances, 7 attributes and two classes (yes and no). The first five variables are all blood tests which are understood to be sensitive to liver illnesses that might arise from excessive alcohol ingestion. The hepatitis contains 155 instances, 19 attributes and two classes (die and live). Pima indian diabetes contains 768 instances, 7 attributes and two classes (negative and positive). Quite a few constraints were located on the selection of these objects from a larger database. In specific, all patients here are females at least 21 years old of Pima Indian heritage. ADAP is an adaptive learning routine that generates and executes digital analogs of perceptron-like devices. The problem-solving, binary-valued variable examined is whether the patient shows symptoms of diabetes according to World Health Organization criteria (i.e., if the 2 hour post-load plasma glucose was at least 200 mg/dl at any survey examination or if found during routine medical care) [28]. Echocardiogram data set contains 132 instances, 13 attributes, and two classes (dead and alive). All the patients affected heart problems at some point in the earlier days. Some are still alive and some are not. The continued existence and still-alive people, when taken together, point to whether a patient stay alive for at least one year following the heart problems[28]. These five datasets are applied for medical diagnosis using multi-granulation rough set based classification.

6. Experimental Analysis

Classification [4, 11, 13, 14] of complex measurements is essential in an analysis process. Accurate classification of measurements may in fact be the most critical part of the diagnostic process. Several classification measures are available in the pattern recognition techniques. In this paper, seven classification measures such as precision, recall, F-measure, Folke-mallows, Kulcznski, rand and Russel-rao indexes were applied for evaluating the accuracy of classification [1, 24]. Precision, recall and F-measure are external measures in classification analysis and other four measures are internal measures in classification analysis. Most of the researchers have applied only external measures. In this paper, external measures along with four internal measures are applied to validate the proposed approaches. The various validation measures are applied toappraise the accuracy of proposed classification approach for diagnostic process. Table 5, 6, 7, 8 and 9 presents different algorithms for the detection of different diseases and effectiveness of those algorithms using various validation measures. Table 4 depicts various validation measures used in this work.

Table 4: Various classification measures

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

$$\text{F-Measure (Czekanowski-Dice index)} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Folkes-Mallows index} = \sqrt{\text{Precision} * \text{Recall}}$$

$$\text{Kulczynski index} = \frac{1}{2}(\text{Precision} + \text{Recall})$$

$$\text{Rand index} = \frac{(\text{True positive} + \text{False negative})}{(\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative})}$$

$$\text{Russel-Rao index} = \frac{\text{True positive}}{(\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative})}$$

Table 5 represents the accuracy of the proposed method (Optimistic Multi-granulation rough set based classification) of pattern separation which successfully diagnosed 73.18% for liver disorder medical data. Back propagation neural network, K^{th} nearest neighbor, support vector machine and Rough set based classifier algorithms provides accuracy of 69.50%, 61.09%, 64.52% and 71.75% for Liver disorder data set respectively.

Table 5: Results for the Liver disorder data set and detailed performance comparison of proposed method with other classifiers

Measures	OMGRS	RS	BPN	KNN	SVM
Precision	0.846	0.8278	0.697	0.6148	0.6444
Recall	0.668	0.669	0.6934	0.6092	0.6529
Czekanowski-Dice index	0.6876	0.6558	0.6947	0.6089	0.6385
Folkes-Mallows index	0.7267	0.7016	0.6949	0.6104	0.6436
Kulczynski index	0.7695	0.7534	0.6952	0.612	0.6487
Rand index	0.7420	0.7217	0.7027	0.6204	0.6431
Russel-Rao index	0.371	0.3358	0.3513	0.3102	0.3216

Table 6 represents the accuracy of the proposed method (Optimistic Multi-granulation rough set based classification) which is able to classify 97.11% correctly for breast cancer medical data. Other classifier algorithms Back propagation neural network, K^{th} Nearest Neighbor, support vector machine and Rough set provide an accuracy of 91.94%, 72.28%, 72.28% and 90.36% for Breast cancer data set respectively.

Table 6: Results for the breast cancer data set and detailed performance comparison between the proposed method and other classifiers

Measures	OMGRS	RS	BPN	K-NN	SVM
Precision	0.9634	0.9412	0.9194	0.7189	0.8054
Recall	0.9792	0.8734	0.9197	0.7268	0.8502
Czekanowski-Dice index	0.9708	0.8964	0.9193	0.7228	0.7965
Folkes-Mallows index	0.9709	0.9018	0.9194	0.7288	0.8120
Kulczynski index	0.9714	0.9073	0.9195	0.7310	0.8278
Rand index	0.9728	0.9127	0.9192	0.6144	0.8025
Russel-Rao index	0.4864	0.4563	0.4595	0.5028	0.4012

Table 7 demonstrates the accuracy of proposed method (Optimistic Multi-granulation rough set based classification) as 77.61% and a comparative analysis is made with Back propagation neural network, K^{th}Nearest Neighbor, support vector machine and Rough set algorithms. These methods provide accuracies of 64.73%, 45.96%, 50.23% and 63.78% for hepatitis data respectively.

Table 7: Results for the hepatitis's data set and detailed performance comparison between the proposed method and other classifiers

Measures	OMGRS	RS	BPN	KNN	SVM
Precision	0.84	0.7381	0.6537	0.4745	0.4872
Recall	0.7143	0.5470	0.6505	0.4874	0.5183
Czekanowski-Dice index	0.7741	0.6283	0.6377	0.4170	0.5014
Folkes-Mallows index	0.7369	0.4984	0.6461	0.4469	0.5000
Kulczynski index	0.7770	0.6426	0.6546	0.481	0.5264
Rand index	0.7420	0.5032	0.6400	0.4960	0.5012
Russel-Rao index	0.371	0.2516	0.3200	0.248	0.2673

The results for OMGRS, RS, KNN, SVM, and BPN are shown in the Table 8. Classification accuracy were 92.29% for OMGRS, 80.52% for BPN, 62.85% for KNN, 70.31% for SVM and 72.61% for Rough set when these approaches are applied to pima indian diabetes data set.

Table 8: Results for the Pima Indian diabetes data set and detailed performance comparison between the proposed method and other classifiers

Measures	OMGRS	RS	BPN	KNN	SVM
Precision	0.9512	0.8597	0.8423	0.6245	0.7024
Recall	0.8992	0.6362	0.7788	0.6348	0.7039
Czekanowski-Dice index	0.9184	0.6825	0.7942	0.6264	0.7031
Folkes-Mallows index	0.9218	0.6850	0.8023	0.628	0.7031
Kulczynski index	0.9253	0.7479	0.8106	0.6297	0.7032
Rand index	0.9297	0.7461	0.8218	0.6543	0.7321
Russel-Rao index	0.4648	0.3730	0.4109	0.3271	0.3660

Performances of Classification algorithms are presented in the following Table 9. The proposed method (Optimistic Multi-granulation rough set based classification) provides high accuracy of 91.75% for heart diseases (Echocardiogram) medical data. Other classifier algorithms Back propagation neural network, K^{th} nearest neighbor, support vector machine and Rough set based classifier algorithms provide accuracy of 83.40%, 82.75%, 75.36% and 89.15% heart diseases (Echocardiogram) data set.

Table 9: Results for heart diseases (Echocardiogram) data set and detailed Performance comparison between the proposed method and other Classifiers

Measures	OMGRS	RS	BPN	KNN	SVM
Precision	0.9000	0.9386	0.8561	0.8173	0.7537
Recall	0.9400	0.8542	0.8148	0.8458	0.7536
Czekanowski-Dice index	0.9125	0.8819	0.8311	0.8192	0.7537
Folkes-Mallows index	0.9163	0.8861	0.8456	0.8248	0.7538
Kulczynski index	0.9200	0.8964	0.8369	0.8163	0.7537
Rand index	0.9189	0.9054	0.8402	0.8478	0.8095
Russel-Rao index	0.4595	0.4527	0.3986	0.3208	0.4048

7. Discussion

Optimistic multi-granulation rough set are far and wide and successfully used models for classification in this paper, predicting and problem solving approach. OMGRS is proposed to diagnosis diseases. The main goal of this database is to construct the proposed model, which will be performed the probable diagnosis of medical data. To evaluate the effectiveness of the Optimistic multi-granulation rough set technique, three standard medical data sets, Such as Breast Cancer, Hepatitis's, Liver DisordersHeart diseases (Echocardiogram), and Pima Indian Diabetes from the UCI Repository of Machine Learning datasets are handled. A number of valuablepresentation metrics in medical fields which include Precision, Recall, F-measure, Folke-mallows, Kulcznski, Rand and Russel-rao indices are work out. The outcomes are analysed and related with those from other methods available in the literature. The experimental results positively demonstrated that the Optimistic multi-granulation rough set classification method is effective in responsibilityof medical data classification tasks. More importantly, the multi-granulation rough set only is able to produced better classification outcome measures than single-granulation rough set. As a result, domain users (i.e., medical specialist's) are able to realize the prediction given by the multi-granulation rough set technique; hence its role as a useful medical decision making tool. Overall, the results indicate thsat multi-granulation rough set method performs better than all other methods. This multi-granulation rough set can be applied to a variety of medical data. Experimental analysis demonstrates that the proposed method performs better than rough set, BPN, KNN, and SVM. However, those methods that are specialized to specific applications can often achieve better performance by taking into account former information. Selection of an applicable method to a classification problem can therefore be a challengingproblem. In consequence, still there is much room for over current medical data classification tasks. Therefore, there is anexcessiveprospective for the use of data mining approaches for medical data classification problem, which has been fully examined and would be one of the interesting directions for future research.

8. **Conclusion**

In this paper five classification techniques in data mining to predict medical disease in patients are compared: rule based Optimistic multi-granulation rough set, Rough-set, BPN, KNN and SVM. These techniques are compared on basis of classification measures of True Positive Rate and False Positive Rate. Our studies showed that Optimistic multi-granulation rough set model turned out to be the best classifier for medical disease prediction. In future, we intend to improve performance of these basic classification techniques by creating meta model which will be used to predict medical disease in patients.

Acknowledgement

The second author would like to thank UGC, New Delhi for the financial support received under UGC Major Research Project No. F-41-650/2012 (SR).

References

1. Bernard Desgraupes, “Clustering Indices”,*University of Paris Ouest - Lab Modal'X*, 1-34,2013.
2. Didier Dubois, Henri Prade, “Rough Fuzzy sets and fuzzy rough sets”, *International Journal of General Systems*, 17(2-3) (1990) 191-209.
3. Ming Zhang, WeiyanXu, Xibei Yangand Zhenmin Tang, “Incomplete Variable Multigranulation Rough Sets Decision”, An *International Journal Applied Mathematics & Information Sciences*”, 8(3) (2014) 1159-1166.
4. ManjeevanSeera, CheePeng Lim, “A hybrid intelligent system for medical data classification”,*Expert Systems with applications*41(5) (2014) 2239-2249.
5. Z.Pawlak, "Rough sets", (1982), *International Journal of Computer Information. Science*, 11(1982) 341–356.
6. Z.Pawlak Z, A.Skowron, (2007), "Rough sets: some extensions", *Information Science.* ,177 (2007) 28–40.
7. Z.Pawlak, A.Skowron, "Rough sets and Boolean reasoning", *Information Science*, 177(2007) 41–73.
8. R. Raghavan, “Validation Over Basic Set Operations Of Internal Structure of MultiGranular Rough Sets”,*International Journal of Latest Research In Engineering and Computing (IJLREC),* 1(2) (2013) 34-42.
9. R. Raghavan and B. K. Tripathy, “On Some Topological Properties of Multigranular Rough Sets”, *Advances in Applied Science Research*, 2 (3) (2011) 536-543.
10. R. Raghavan and B. K. Tripathy, “On Some Comparison Properties of Rough Sets Based on Multigranulations and Types of Multigranular Approximations of Classifications” , *International journal. Intelligent Systems and Applications,* 06 (2013) 70-77.
11. S.Senthilkumar , H.HannahInbarani, S.Udhayakumar, “Modified Soft Rough set for Multiclass Classification”, *Advances in Intelligent Systems and Computing*, 246 (2014) 379-384.
12. B. K. Tripathy, G. K. Panda, A. Mitra, “Incomplete Multigranulation Based on Rough Intuitionistic Fuzzy Sets”,*UNIASCIT,* 2 (1) (2012) 118-124.
13. S.Udhaya Kumar, H.HannahInbarani, S.Senthil Kumar, “Bijective soft set based classification of Medical data*”, Pattern Recognition, Informatics and Medical Engineering (PRIME), International Conference*, (2013)517 – 521.
14. S.Udhaya Kumar, H.HannahInbarani, S.Senthil Kumar “Improved Bijective-Soft-Set-Based Classification for Gene Expression Data”, *Advances in Intelligent Systems and Computing,* 246 (2014) 127-132.
15. Wu Chen, Xu Wei, Yang Xibei , Wang Lijuan, “A Variable Precision Rough Set Model Based on Multi-granulation and Tolerance”, *International Journal of Engineering and Innovative Technology*, 3(7) (2014) 1-6.
16. WeihuaXu, Xiantao Zhang, and Qiaorong Wang, “A Generalized Multi-granulation Rough Set Approach”, *ICIC 2011*, (2012) 681–689.
17. Xibei Yang, Xiaoning Song, Huili Dou, Jingyu Yang, “Multi-granulation rough set: from crisp to fuzzy case”, *Annals of Fuzzy Mathematics and Informatics*, 1(1) (2011) 55- 70.
18. YuhuaQian, Jiye Liang, Yiyu Yao, Chuangyin Dang, “MGRS: A multi-granulation rough set”, *Information Sciences* 180 (2010) 949–970.
19. Eric C.C. Tsang , Chen Degang, Daniel S. Yeung, “Approximations and reducts with covering generalized rough sets”, *Computers & Mathematics with Applications*56(1), (2008),279–289.
20. Bing-yuan Cao, Ji-hui Yang, “Advances in Fuzzy Geometric Programming”,*Fuzzy Information and Engineering, Advances in Soft Computing* , 40, (2007), 497-502.
21. Silvia Rissino, Germano Lambert-Torres, “Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications”, *Data Mining and Knowledge Discovery in Real Life Applications*, 2009, 35-60.
22. YuhuaQian a, Hu Zhangb, YanliSangb, Jiye Liang, “Multigranulation decision-theoretic rough sets”, *International Journal of Approximate Reasoning* 55 (2014) 225–237.

23. Jiye Liang, YuhuaQian, Chengyuan Chu, Deyu Li, Junhong Wang, "Rough Set Approximation Based on Dynamic Granulation", *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Lecture Notes in Computer Science* , 3641, (2005), 701-708.
24. Xin Yao, Yong Liu, "Neural networks for breast cancer diagnosis", *Evolutionary Computation, CEC 99. Proceedings of the 1999 Congress on*, 3, 1999.
25. Mohammed Abdul Khaleel, Sateesh Kumar Pradhan, G.N.Dash, F. A. Mazarbhuiya, "A Survey of Data Mining Techniques on Medical Data for Finding Temporally Frequent Diseases",*International Journal of Advanced Research in Computer and Communication Engineering*, 2(12), (2013), 4821-4824.
26. YuhuaQian, Shunyong Li, Jiye Liang, Zhongzhi Shi, FengWang, "Pessimistic rough set based decisions: a multigranulation fusion strategy*", International Journal of Information Sciences*, 264, (2014),196-210 .
27. YuhuaQian,JiyeLiang,Yiyu Yao, Chuangyin Dang, "MGRS: A multi-granulation rough set", *Information Sciences***,** 180(6), (2010), 949–970.
28. http://archive.ics.uci.edu/ml/datasets.html.