Rough-Mereology Framework for Making Medical Treatment Decisions Based on Granular Computing

Mohammed M. Eissa
Information Systems Department, Faculty of Business Administration and Information Systems
French University in Egypt, Cairo, Egypt
E-mail: Mohammed.mamdouh@gmail.com

Mohammed Elmogy
Information Technology Dept., Faculty of Computers and Information, Mansoura University, Mansoura, Egypt
E-mail: elmogy@gmail.com

Mohamed Hashem
Information Systems Dept., Faculty of Computers and Information, Ain Shams University, Cairo, Egypt
E-mail: mhashem100@yahoo.com

Keywords: rough – mereology, granular reflection, rough inclusion, hepatitis C virus (HCV), coronary heart disease (CHD)

Received: August 19, 2015

The medical field is considered as one of the most significant research resources. It receives a significant interest from researchers in the field of informatics and medical experts. It has a tremendous amount of data on various diseases and their symptoms that causes difficulty in diagnosing diseases. Therefore, several medical approaches based on knowledge discovery in the database have been proposed and developed. They include data mining techniques for cleaning, filtering, dimension reduction, and induction of rules. In this paper, a Rough-Mereology framework is proposed for classification of Hepatitis C Virus (HCV) and Coronary Heart Disease (CHD) medical datasets. The proposed system uses granular reflection mechanism based on rough inclusion to generate sets of granules at different radiuses. It selects the optimum radius based on accuracy to induce a set of rules that help medical experts to take Therapeutic Medical Decisions. During the experiments, the Rough based granular computing supplies a complimentary and comprehensive framework for the analysis of medical datasets.

Povzetek: *Prispevek uvaja novo metodo strojnega učenja na dveh domenah: hepatitis in srce.*

1 Introduction

Because of the vast amount of medical data in different forms, the medical field has become one of the richest areas of scientific research and informatics. It can improve medical decisions, reduce costs, and provide finest therapeutic service for patients. Therefore, several medical approaches based on knowledge discovery in the database have been proposed and developed.

In this section, we focus on two of the most currently common diseases in the world. They are Coronary Heart Disease (CHD) and Hepatitis C virus (HCV). CHD is one of most common disease in our modern life. It is the hardening of the vessels by greasy stores called plaque. The heart must get oxygen and supplements to functioning efficiently. Blood conveys the oxygen and supplements to the heart through arteries. As the plaque composed of the walls of the arteries, blood flow is decreased [1], [2], [3].

The second common disease is the HCV, which is a most popular liver disease. HCV currently infects millions of people in different countries. It results in intense illness, which is regularly turned into a deadly disease. It can prompt liver failure and liver cancer that leads to death. Infected blood that is transmitted between people is considered as the main reason for HCV infection [4].

Despite the significant progress in the medical field, the large number of medical indicators of both CHD and HCV make finding relationships between those indicators a difficult task. Consequently, the techniques of knowledge discovery in databases (KDD) are widely developed to determine relationships between medical indicators. It can predict the factors that affect treatment decisions using the historical cases, which are stored in medical datasets.

The organization of the rest of this paper will be as follows. In Section 2, overviews of the Rough-Mereology and Granular Computing frameworks are described. Section 3 discusses some current related work. The proposed framework is demonstrated in Section 4. Description of medical datasets, framework analysis, and experimental results are presented in Section 5. Research conclusion and future work are shown in Section 6.

2 Rough-mereology based granular computing: a brief overview

The information granularity concept, which is used in many areas, is proposed by Zadeh [5]. Most of the researchers characterized the main concepts of the granular computing from distinctive perspectives. Like an umbrella, Granular Computing covers themes in different fields even if they are considered disconnected. Granular Computing includes three main topics. These topics are Rough sets theory (RST) that is proposed by Pawlak [6], the theory of fuzzy sets that is introduced by Zadeh [7], and quotient space hypothesis that is presented by L. Zhang and B. Zhang [8]. One of the essential benefits of RST is its powerful data analysis and immediate application in classification. The main principals of RST will be discussed in the next subsections.

2.1 Knowledge representation

Granulation of a universe involves dividing the universe into subsets or grouping individual objects into a set of clusters. A granule is a subset of the universe. A family of granules that contains every object in the universe is called a granulation of the Universe. Let U is a non-empty finite set of objects, AT is a non-empty set of attributes, L is a language defined using attributes in AT, V_a is a non-empty set of values for $a \in AT$, and I_a is an information function. Then, the information table can be formulated as follows [9]:

$$S=(U, AT, L, \{V_a|a \in AT\}, \{I_a|a \in AT\}) \quad (1)$$

On the other hand, a generalized approximation space can be defined as $AS = (U, I, v)$, where I is the uncertainty function on U with values in the power set pow(U), which is the neighborhood of x) and v is the inclusion function defined on the *Cartesian product* $pow(u)\ x\ pow(u)$ with values [0,1] measuring the degree of inclusion of sets [10].

The lower approximation (B_{lower}) and the upper approximation (B_{upper}) operations can be defined by equations (2) and (3), where $v(I(x), X)$ represents the Rough similarity relation, as shown in Fig. 1:

$$B_{lower} = \{x \in U: v(I(x), X) = 1\} \quad (2)$$

$$B_{upper} = \{x \in U: v(I(x), X) > 0\} \quad (3)$$

2.2 Rough mereology

Rough Mereology is proposed by Lesniewski [12] as the theory of concept. The primitive relation of Mereology is a part of the relation. According to Polkowski and Artiemjew [13], the Mereology relation is described in equation (4), where $\pi(u, w)$ is a partial relation (proper part) and $ing(u, w)$ is ingredient relation means informally an improper part [14].

$$ing(u, w) \leftrightarrow \pi(u, w) \text{ or } u = w \quad (4)$$

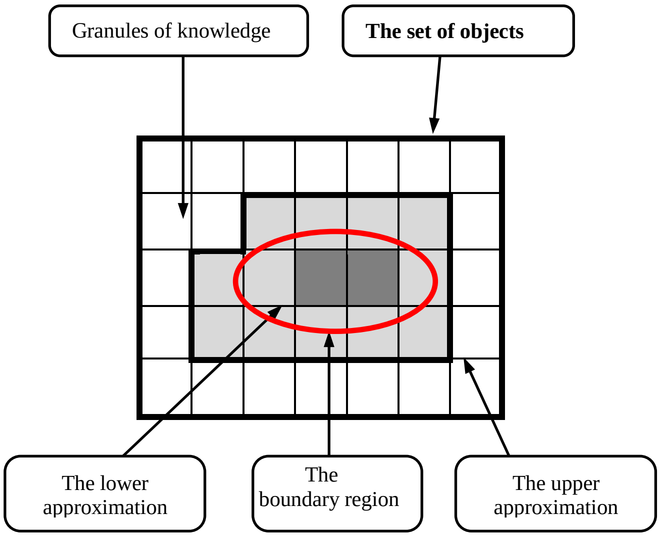

Figure 1: The definitions of approximations expressed regarding granules of knowledge [11].

The relation of the ingredient is a part order of things [14]. $\mu(X, Y, r)$ means rough Mereology relation x is part of y at least degree r, also described as shown in equation (5).

$$\mu(X, Y, r) = sim_{\delta}(X, Y, r) \leftrightarrow p(X, Y) \leq (1 - r) \quad (5)$$

Rough inclusion relation satisfies the similarity properties of Rough Mereology as mentioned in the next section.

2.3 Rough inclusion

The Rough inclusion function is the extension of Rough *indiscernibility* relation (IND) of traditional RST. However, Rough inclusion is less complexity time than *indiscernibility* relation in traditional RST. The Rough Inclusion Function $v: P(U)xP(U) \rightarrow [0,1]$ defines the degree of inclusion of X in Y. It means in Rough Mereology x is part of y at least degree . $X, Y \subseteq U$, $IND(X, Y) = a \in AT : a(x) = a(y)$, a is an attribute in an information system. Rough inclusion can be represented in terms of indescribability relation as shown in the following equation:

$$\mu(X, Y, r) = \frac{IND(X, Y)}{card(AT)} <=> r \quad (6)$$

There are three types of measures associated with granules. The first measures a single granule that indicates the relative size of the granule. The second measures the relationships among granules. Finally, the third measures the relationships between a granule and a family of granules, as indicated in following equations [15].

$$\text{Absolute Support } (x \Rightarrow y) = \frac{|m(x) \wedge m(y)|}{m(x)} \quad (7)$$

$$\text{Coverage } (x \Rightarrow y) = \frac{|m(x) \wedge m(y)|}{m(y)} \quad (8)$$

$$P(x|\,y) = \left(P(x_1|\,y) = \frac{|m(y) \land m(x_1)|}{m(y)}, \dots, P(\psi_n|\,y) = \frac{|m(y) \land m(x_n)|}{m(y)}\right) \tag{9}$$

The granule g of radius r can be defined as shown in equation (10), where μ is a Rough inclusion of an object $x\ and\ r \in [0,1]$.

$$g_r\, x = \mathrm{Cls}\{y : \mu(y, x, r)\}. \tag{10}$$

3 Related work

Granular Computing in the framework of RST and its extensions has been widely discussed by many researchers [16], [17], [18], [19], [20], [21], [22], [23], [24] and [25]. Rough-Mereology as an extension of RST is used in different knowledge discovery application areas. For example, Zheng and Zhan [26] explored Granular Computing technique based on Rough Mereology for rule generation. This research indicated that Granular Computing could improve the performance of reaching association rule based on approximation spaces. It is applied to a rule-based classifier. However, the model was not extended to handle more classifiers, and it is based on minimum length principle.

A survey of Rough-Mereology applications in knowledge discovery and intelligent systems is introduced by Polkowski [27]. On the other hand, Polkowski and Artiemjew [28] developed a model using granular reflections in the frame of Rough-Mereology for rules induction and classification. They made a comparative analysis of exhaustive RS classifier whose accuracy is less than their proposed model.

Many researchers used Granular Computing in the frame of RS and its extension Rough-Mereology for the classification of the medical datasets. For example, RS approach is developed by Zaki et al. [29] to transfer numeric attributes to discrete attributes and to induce HCV rules for classification. Although the approach did not use any reduction algorithm for attribute reduction, the classification accuracy was satisfied.

Badria et al. [30] proposed a Rough based Granular model by using the fundamental of RS. The model was used to discover attributes dependencies, data transformation, and the dynamic dimension reduction. Then, set of rules induced to make medical decisions.

Eissa et al. [31] introduced an HCV classifier based on a hybrid Rough genetic model. The model based on RS in the transformation of datasets, dimension reduction, and rule induction. A genetic algorithm is used in filtering and selecting the best rules that are encoded inside chromosomes to increase the accuracy.

Polkowski and Artiemjew [32] developed a granular classifier for coronary disease. The proposed model concentrated on dealing with missing values in a pre-processing phase, creating new information table and applying the granular classifier to discover absence or presence of coronary disease.

In this work, Rough-Mereology framework based on Granular Computing model is introduced to classify HCV at some stage in testing a new drug for HCV treatment. In addition, Coronary Artery Disease dataset is used to determine the number of blocked vessels in coronary. Section 4 provides a brief description and analysis of the proposed framework.

4 The proposed framework based on rough-mereology model

Although the great development of technologies of storage, data retrieval, and the vast amount of different forms of medical data, the research in the medical field has become one of the richest areas in informatics. They are made available to the medical research community. Forecasting a disease will remain one of the most defy tasks for researchers to provide prediction models for enhancing treatment decisions [33].

The proposed framework uses extended RS based on Granular Computing methodology to identify the most relevant attributes. It induces medical treatment rules from diverse medical datasets. The development processes of the proposed framework demonstrated in Fig. 2. It requires pre-processing of the medical dataset to remove redundancy, inconsistency, and convert continues data to discretized one to be more suitable for processing. In addition, attribute reduction is needed to find the optimum attributes that represent the datasets without losing the value of the data. Whereas, Rough Mereology and Rough inclusion techniques are utilized to clustering the datasets into sets of granules with different radius. They select the granules with optimum radius for inducing sets of medical rules, which are required for treatment decisions. The main steps of the proposed model will be illustrated in the next subsections.

4.1 The pre-processing phase:

4.1.1 Construction of information table

For classification tasks, it is assumed that each object in the information table is associated with a unique class label. Objects can be divided into classes, which form a granulation of the universe. Without loss of generality, we assume that there is a unique attribute class, which takes class labels as its values. The set of attributes is expressed as $AT = D \cup \{class\}$, where D is the set of attributes that is used to describe the objects, also called the set of conditional attributes. A granule is a definable granule if it is associated with at least one formula, i.e. X=m(φ), where $\varphi \in L$. The extended upper and lower approximations are calculated based on Rough inclusion after the construction of the medical information table and using Rough inclusion relation to finding the elementary granules. Its essential purpose is to seek for an approximation scheme that can efficiently solve a complex problem.

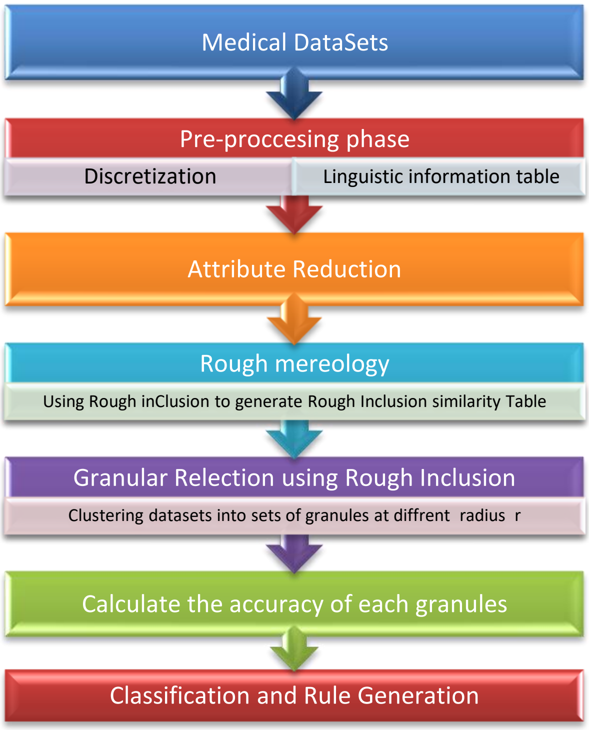

Figure 2: The proposed framework of the Rough-Mereology model.

4.1.2 Rough sets boolean reasoning discretization

The objective of discretization is to discover a set of cut points to partition the range into a small number of intervals that have good class coherence. It is usually measured by an evaluation function. In addition to the maximization of interdependence between class labels and attribute values, a perfect discretization technique ought to have an optional objective to speed up learning process and classification accuracy [34].

In this paper, the Rough sets Boolean Reasoning (RSBR) is used as a classification based discretization algorithm. It converts the continuous medical attributes into discrete ones. RSBR algorithm is described in Fig. 3, in which (i) define a set of Boolean variables BV(U), then a new decision table T^P is created using boolean variables BV(U), defined in the previous step (T^P is called P-discretization of decision table), then searching for a minimal subset of P that discerns all the objects in different decision classes. Finally, we obtain the prime implicants denoted by the discrenibility formula in disjunctive normal form DNF form. The minimal subset of P is chosen to preserve discrenibility [35].

Algorithm1: The RSBR Discretization Algorithm

Input

Information system Table (IS) with real-valued attributes A_{ij}

Output

Discretized Information Table

Begin procedure

(1) For $A_{ij} \in IS$ do

(2) Define a set of Boolean variables as follows:

$$BV(U) = \left\{ \sum_{i=1}^{n} C_{ai}, \sum_{i=1}^{n} C_{bi}, \sum_{i=1}^{n} C_{ci}, \ldots, \sum_{i=1}^{n} C_{ni} \right\}$$

Where $\sum_{i=1}^{n} C_{ai}$ a set of intervals Defined on the variables of attributes a.

(3) End for

(4) Create new information TableT^P by using a set of intervals C_{ai}

(5) Find minimal subset of C_{ai} that discern all the objects in decision class D using the following formula:

$$\Phi^U = \wedge\{\psi\ (i,j)\ : d(x_i) \neq d(x_j)\}$$

Where ψ (i,j) is the number of minimal cut that must be used to distinguish two different instances x_i and x_j in the information Table.

End procedure

Figure 3: The RSBR discretization algorithm.

4.2 Attribute reduction algorithm

One of the important phases of the RST is the attribute reduction. In an information system, some attributes may be redundant and useless. If those redundant and useless attributes are removed without influencing the classification, they will be considered as superfluous attributes [36].

The core concept is commonly used in all reducts [37]. The attribute reduction concept can improve the performance of the generated rule systems. It accelerates the rule induction process by finding all minimal subsets of attributes. These attributes have the same number of elementary sets without the loss of the classification power of the reduced information system [38].

In this phase, discernibility matrix-based algorithm is utilized for reduction of superfluous attributes in medical datasets. Let $IS = (U, A \cup D, V, F)$ be a decision table. we denote nxn matrix called discernibility matrix M such that:

$$C_{ij} = \begin{cases} \emptyset, & f_D(X_i) = f_D(X_j) \\ \{a \in A : a(X_i) \neq a(X_j)\}, & f_D(X_i) \neq f_D(X_j) \end{cases} \quad (11)$$

Using discernibility matrix [39], attributes reduction can be constructed as shown in Fig. 4.

Algorithm 2: Attribute Reduction Algorithm

Input
Let M be the discenibility matrix of an information system
Output
Reduct (R) of Information system
Begin procedure
(1) $R \leftarrow \emptyset$
(2) Compute the core Co from M
(3) Reduct ← Co
(4) Construct matrix M_R from M
(5) For any $a \in A$ *Compute significant of attributes SFG*
$SFG(a, R, A) = \gamma_{R \cup \{a\}}(A) - \gamma_R(A)$
Where $\gamma_R(A) = \frac{|POS_R(A)|}{|U|}$ from M_R
(6) Let $SFG(a', R, A) = max_{a \in A-R}\{FG(a, R, A)\}$
(7) $R \leftarrow R \cup \{a'\}$
End procedure

Figure 4: The attribute reduction algorithm.

4.3 Classification phase

In this phase, Rough Mereology uses Rough inclusion relation as a similarity technique in the classification of medical datasets. It produces sets of granules at different radiuses that draw the outlines for inducing decision rules. First, Rough Mereology using Rough inclusion to construct Rough inclusion similarity table that describes the similarity of each object in each attribute in the medical datasets. Second, the clustering of the medical datasets into sets of granules with different radius are done by reproducing Rough inclusion table with different granule radius table. It reflects the degree of similarity between the medical indicators. Finally, the granular reflection of inclusion into the set of granules is applied using voting by training object algorithm to select the most optimized granules by selecting granule radius that achieves the best accuracy. The primary steps in the steps of the classification algorithm are provided in Fig. 5.

5 The experimental results and Discussion

5.1 The description of the medical datasets

In this paper, the historical medical data are collected from different medical research resources (Badria and Attia [40], Barakat et al. [41] and Amir et al. [42]). In addition, several meetings with medical experts had made to discuss and understand the contents of the medical datasets and to get a clear idea about the diseases. The computations of rules have been only done on the training dataset. The computations' results of the rules were applied to the classification of the granules from the tested dataset.

Algorithm 3: The Rough–Mereology Classification Algorithm

Input
Reduct Information Table (RIT)
Output
Decision Rules Granule Set (DRGS)
Begin procedure
1. **Initialize** Set of Rule Sets (SRS)) = ϕ
2. **For Each** Granular Radius $\in \left\{\frac{0}{card|A|}, \frac{1}{card|A|}, \dots, \frac{card|A|}{card|A|}\right\}$ **do**
3. **Initialize** Rough Inclusion set (RIS) = ϕ
4. **For Each** Attribute in (RIT) **do**
5. **Compute** Rough inclusion
6. **Add** Rough inclusion values in **(RIS)**
7. **End For**
8. **Initialize** Decision Rules Granule Set (DRGS) = ϕ
9. **For Each** row in (RIS) **Do**
10. **Induce** decision rule
11. **Remove** redundant rules
12. **Add** rule to(DRGS)
13. **Output DRGS**
14. **End For**
15. **Save** DRGS in SRS
16. **For Each rule in SRS do**
17. **Compute** Accuracy measure and store it in **AccSet**

$$Accuracy = \frac{card\,(C \wedge D)}{card(C)} * 100$$

18. **End For**
19. **Initialize** bestACC ← first element in **AccSet**
20. **For Each Element in AccSet do**
21. **If** element *e* **is** bestACC
22. bestACC= *e*
23. **End If**
24. **End For**
25. **Get the corresponding set of rules of** bestACC and store it in **DRGS**
26. **Apply voting by training object** equations to refine **DRGS**

End procedure

Figure 5: The Rough–Mereology classification algorithm.

Hepatitis C Virus (HCV) Dataset
These data were gathered from clinical trials at some stages in testing a new drug for HCV, which is recently developed and patented by Badria and Attia in 2007 [40]. It comprised of 119 HCV cases. Each case is portrayed by 28 medical indicators: 23 numerical indicators and 5 categorical indicators. The intention of the dataset is to forecast the presence or absence of the HCV. The attributes description of the HCV exists in Table 1. For each HCV record, patient data out of 27 condition attributes and the decision attribute describe the presence or absence of HCV-related to the proposed medication. All these data were gathered from the treatment of HCV and were divided with the splitting factor of 0.25 into training and test sets.

Coronary Heart Disease (CHD) Dataset
CHD dataset was collected from Cardiology Department, Faculty of Medicine, Mansoura University, Egypt. It consists of 215 Coronary patients that included condition attributes like age, sex, family history, smoking, and other medical measures. In addition, it shows the decision label that indicates the number of the vessels that may be injected (no vessel, single, two, and multiple). Most of the attributes are binary attributes and age attribute is the only numerical one, as shown in Table 2.

For training and testing purposes, with a splitting factor of 25 %, CHD dataset is divided into training and testing sets [41], [42].

Table 1: The HCV dataset.

ID	Medical Indicators	Indicator Description
1	Sex	Male or female
2	Source	Source of HCV: blood transfusion, non-sterile tools by dentist or surgery
3	S.G.P.T (ALT)	normal 0 to 40 U/L
4	S.G.O.T (AST)	normal 0 to 45 U/L
5	Serum Bilirubin (SB)	normal 0 to 1.1 mg/dL
6	Serum Albumen (SA)	Serum Albumin; 3.5 to 5.1 g/dL
7	Serum Ferritin	the normal 22 to 300
8	Ascites	No, Mild, and Ascites
9	Spleen	Normal, Absent, and Enlarged
10	Lesions	0,1 or 2
11	Portal vein (P.V)	Natural diameter is 12 mm
12	PCR	Quantitative analysis of the virus U/mL
13	PLT	Platelets normal 150 to 450 /cm m
14	WBC	White Blood Corpuscles normal 4 to 11/cm m
15	HGB Haemoglobin	male 12.5 to 17.5 g/dL female 11.5 to 16.5 g/dL
16	Headache	Yes or No
17	Blood Pressure	Yes or No
18	Nausea	Yes or No
19	Vertigo	Yes or No
20	Vomiting	Yes or No
21	Constipation	Yes or No
22	Diarrhea	Yes or No
23	Appetite	Yes or No
24	Gasp	Yes or No
25	Fatigue	Yes or No
26	Skin colour	Yes or No
27	Eye Colour	Yes or No
28	Decision	-1 absent, 1 present of HCV

Table 2: The CHD dataset.

ID	Medical Indicators	Indicator Description
1	Age	Continuous values between 35 and 62
2	sex	Male or female
3	smoking	Yes or no
4	diabetes mellitus (Dm)	Yes or no
5	Dyslipidemia (dyslipid)	Yes or no
6	Family history (family_h)	Yes or no
7	Left main coronary artery (lmca)	Normal or diseased
8	Left anterior descending artery (lad)	Normal or diseased
9	First diagonal artery (d1)	Normal or diseased
10	Second diagonal artery (d2)	Normal or diseased
11	Left circumflex artery (lcx)	Normal or diseased
12	Obtuse marginal artery1 (om1)	Normal or diseased
13	Obtuse marginal artery2 (om2)	Normal or diseased
14	Right coronary artery (rca)	Normal or diseased
15	Posterior descending coronary artery (pda)	Normal or diseased
16	Decision Label	number of vessels may be injected (no vessel, single, two, and multi)

5.2 Data pre-processing stage

As shown in Fig. 2, the data pre-processing phase includes constructing the medical information table using Rough inclusion relationship. It generates elementary granules and discretizes of continuous medical attributes. It converts the continuous values of an attribute to a set of intervals.

In this paper, a supervised discretization algorithm in light of a mix of RS and Boolean Reasoning are utilized. The RSBR combines discretization and classification, which is unlike some of the discretization methods. The RSBR algorithm shows the best results in the classification of the used medical datasets. Tables 3 and 4 illustrate the new HCV and CHD information table after applying the discretization stage.

Table 3: The discretized HCV dataset.

ID	Medical Indicators	Indicator Description
1	Sex	Male or female (0, 1)
2	Source	Source of HCV : blood (0),dentist(1)or surgery (2)
3	S.G.P.T (ALT)	>62 AND <62 [0] less [1] greater
4	S.G.O.T (AST)	>34 AND<34 [0] less [1] greater
5	(SB)Serum Bilirubin	<0.56 AND >0.56 [0] less [1] greater
6	(SA)Serum Albumen	Serum Albumin 3.5 to 5.1
7	Serum Ferritin	>80 AND <80 [0] less [1] greater
8	Ascites	No(0), Mild (1), and Ascites (2)
9	Spleen	Normal (0), Absent(1), and Enlarged (2)
10	Lesions	0,1 or 2
11	(P.V)Portal vein	>13 AND <13, [0] less,[1] greater
12	WBC	White Blood Corpuscles 4 to 11
13	PCR	Quantitative analysis of the virus
14	PLT	>161,<161 [0] less [1] greater
15	HGB	>13.1 AND <13.1 [0] less [1] greater
16	Headache	0,1
17	Blood Pressure	0,1
18	Nausea	0,1
19	Vertigo	0,1
20	Vomiting	0,1
21	Constipation	0,1
22	Diarrhea	0,1
23	Appetite	0,1
24	Gasp	0,1
25	Fatigue	0,1
26	Skin color	0,1
27	Eye Color	0,1
28	Decision	-1absent,1 present of HCV

Table 4: The discretized CHD dataset.

ID	Medical Indicators	Indicator Description
1	Age	[*, 52), [52,59), [59,61) and [61, *)
2	sex	0,1
3	smoking	0,1
4	diabetes mellitus (Dm)	0,1
5	Dyslipidemia (dyslipid)	0,1
6	Family history (family_h)	0,1
7	Left main coronary artery (lmca)	0,1
8	Left anterior descending artery (lad)	0,1
9	First diagonal artery (d1)	0,1
10	Second diagonal artery (d2)	0,1
11	Left circumflex artery (lcx)	0,1
12	Obtuse marginal artery1 (om1)	0,1
13	Obtuse marginal artery2 (om2)	0,1
14	Right coronary artery (rca)	0,1
15	Posterior descending coronary artery (pda)	0,1
16	Decision	number of vessels may be injected (no vessel0, single1, two2, and multiple 3)

5.3 Medical datasets attribute reduction stage

In this stage, reduction algorithm that is based on discernibility matrix is utilized to reduce the number of attributes in medical data, as shown in Fig. 4. It successfully reduced CHD conditional attributes to 7 attributes. In addition, it reduces the HCV medical indicators from 27 to 9 indicators.

5.4 Classification of medical datasets stage

In this stage, Granular Computing in the frame of Rough Mereology formalized the idea of the granular reflection of the medical datasets. First, applying Rough Mereology concept in the frame of Rough inclusion to the medical datasets to induce Rough inclusion similarity table.

Second, a set of Rough inclusion tables is constructed by re-applying the first step with different radius r in the interval [0, 1] for clustering the datasets into sets of granules with different radius. In CHD dataset, 7 Rough inclusion tables are produced, as shown in Tables 5 and 6. In HCV dataset, 9 Rough inclusion tables are introduced, as shown in Tables 7 and 8.

After the granular reflection of the medical datasets, it reflects inclusion tables into a set of granules. Voting by training object is applied in two steps. First, it computes the accuracy measure for each Rough inclusion table with different radiuses r. Second, it selects the optimum radius with the highest accuracy that represents the optimized granules.

Results of the accuracy measure for CHD shows that the accuracy measure at r_3 is the highest accuracy, which equals to 96.2%, as shown in Tables 5 and 6. For HCV dataset, the granule radius r_5 is the optimal granule with accuracy equals to 96.6, as shown in Tables 7 and 8. Samples of decision rules of HCV and CHD are shown in Tables 9 and 10, respectively.

Table 5: The CHD granules accuracy.

Radius (r_{gran})	Accuracy Measure (acc_g)
r_0=0	90.0
r_1=0.166667	85.7
r_2=0.333333	92.1
r_3=0.5	**96.2**
r_4=0.666667	93
r_5=0.833333	0
r_6=1	0

Table 6: The CHD Confusion Matrix radius r3.

		Predicted				
Actual		Single	Two	Multi	no	
	Single	12	1	0	0	0.92
	Two	0	10	0	0	1
	Multi	0	0	14	0	1
	no	0	1	0	15	0.94
		1	0.83	1	1	0.96

Table 7: The HCV granules Accuracy.

Radius (r_{gran})	Accuracy Measure (acc_g)
r_0=0	88.9
r_1=0.111111	74.5
r_2=0.222222	94.7
r_3=0.333333	93.6
r_4=0.444444	95.4
r_5=0.555556	**96.6**
r_6=0.666667	90.2
r_7=0.777778	0
r_8=0.888889	0
r_9=1	0

Table 8: The HCV Confusion Matrix radius r5.

		Predicted		
Actual		absent	present	
	absent	4	0	1
	present	1	24	0.96
		0.8	1	0.97

Table 9: A sample of HCV treatment decision rules.

HCV Decision Rules	Accuracy
IF Sex(m) AND S#G#O#T (AST)3([34, *)) AND Ascites 3(no) AND Spleen3(normal) AND HGB 3([13.1, *)) **then** 1	0.9885
IF S#G#O#T (AST)3([34, *)) AND Seru Ferritin 3([86, *)) AND Ascites 3(no) AND Spleen3(normal) AND HGB 3([13.1, *)) **then** 1	0.9754
IF Sex(m) AND Seru# Bilrubin (SB)3([0.45, *)) AND Seru# Albuin (SA)3([*, 4.2)) AND Ascites 3(no) AND PLT 3([*, 161.000)) **then** 1	0.9611
IF Source(blood) AND Portal vien (P#V) 9([12, *)) AND PCR9([*, 23)) **then** 1	0.9601
IF Sex(m) AND Source(blood) AND Portal vien (P#V) 9([12, *)) AND PCR9([*, 23)) **then** -1	0.9552

A comparative analysis of the proposed model and other Knowledge Discovery models, such as RS, Back Propagation, and Genetic Algorithm, are conducted. In addition, some hybrid models like Rough Granular-Back-propagation and Rough Genetic are used to evaluate the proposed framework classification power and its impact related to the size of training datasets. Table 11 and Fig. 6 show the comparison between the classification accuracy of the different tested models based on HCV and CHD datasets.

Table 10: A sample of CHD treatment decision Rules.

CAD Decision Rules	Accuracy
IF age([46, 48)) AND sex(female) AND dyslipid(yes) **then**0	0.91
IF age([56, 59)) AND sex(female) AND dyslipid(yes) **then** 2	0.96
IF age([54, 56)) AND sex(female) AND dyslipid(no) **then** 1	0.89
IF age([38, 42)) AND sex(female) AND dyslipid(no) **then** 3	0.92
IF age([59, 61)) AND sex(female) **then** 2	0.9730

Table 11: The classification accuracy of different models based on HCV and CHD datasets.

Classification Model	HCV	CAD
Rough Set	95.5	92
Back-Propagation (Neural Network)	93	95.6
Genetic-Algorithm	92.1	93.7
Hybrid Rough-Genetic	95	95.1
Rough Granular Back-Propagation	94.8	94.4
Rough - Mereology based on Granular Computing	96.6	96.2

6 Conclusions

Rough-Mereology theory is an extension of the RST that replaces the indescribability relation with similarity relation Rough inclusion relation. Rough-Mereology is dedicated to the concept of granular computing in constructing elementary information granules. It finds the relationships between information granules and building granules network.

In this paper, new Rough-Mereology based on Granular Computing model is introduced. It helps medical experts to make relationships between varieties of medical indicators. It reduces the ratio of the false positive diagnosis that leads to taking accurate treatment decisions.

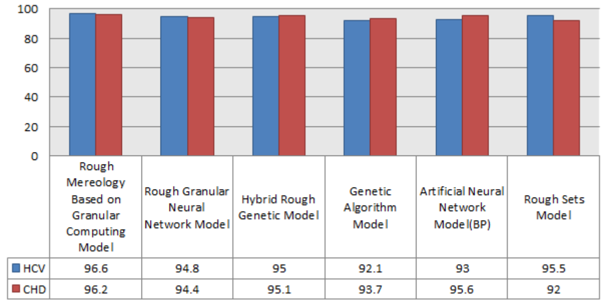

Figure 6: The comparative study of the different classification models.

The proposed framework is constructed by creating the medical decision table based on Rough-Mereology concepts. It uses Rough Sets Boolean Reasoning algorithm in the discretization of continues medical data sets. It makes attribute reduction using discernibility matrix based reduction algorithm to discover the characteristics insignificant arrangement of qualities to maintain the knowledge. Finally, it induces the medical decision rules from produced set of granules generated using Rough inclusion similarity tables with different radius r in the interval [0, 1]. Then, it computes granular reflection by computing accuracy rate of each similarity tables. It selects the optimal granules with higher accuracy. The most promising rules are discovered and chosen with the highest priority for classification purposes.

The experimental results showed that the proposed framework can classify multi-valued decision class (CHD dataset) with acceptable classification rate. According to the competitive analysis, the classification accuracy of the proposed model on the other hybrid RS models with Genetic Algorithm and Artificial Neural Network is improved. The proposed model achieves the best case for the accuracy rather than other models.

Although Rough-Mereology based on Granular Computing model achieve good classification accuracy, the proposed framework will be enhanced for classifying complex information systems.

References

[1] Truswell, A. S. (2002). Cereal grains and coronary heart disease. European journal of clinical nutrition, 56(1), 1-14.

[2] Marcus, F. I., McKenna, W. J., Sherrill, D., Basso, C., Bauce, B., Bluemke, D. A. & Fontaine, G. (2010). Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia. European heart journal, ehq025.

[3] Da Silva Gasparotto, G., Gasparotto, L. P. R., da Silva, M. P., Bontorin, M. S., Lissa, M., & de Campos, W. (2012). Cardiovascular risk factors in freshman physical education course: comparison between the sexes. ConScientiae Saúde, 11(3), 406.

[4] Wasik, S., Jackowiak, P., Krawczyk, J. B., Kedziora, P., Formanowicz, P., Figlerowicz, M., & Błażewicz, J. (2010). Towards prediction of HCV therapy efficiency. Computational and mathematical methods in medicine, 11(2), 185-199.
[5] Zadeh, L. A. (2011). A note on Z-numbers. Information Sciences, 181(14), 2923-2932.
[6] Pawlak, Z. (1982). Rough sets. International Journal of Computer & Information Sciences, 11(5), 341-356.
[7] Zadeh, L. A. (1996). Fuzzy logic= computing with words. IEEE transactions on fuzzy systems, 4(2), 103-111.
[8] Ling, Z., & Bo, Z. (2003). Theory of fuzzy quotient space (methods of fuzzy granular computing).
[9] Livi, L., Rizzi, A., & Sadeghian, A. (2015). Granular modeling and computing approaches for intelligent analysis of non-geometric data. Applied Soft Computing, 27, 567-574.
[10] Yao, Y. Y., & Yao, J. T. (2002, May). Granular computing as a basis for consistent classification problems. In Proceedings of PAKDD (Vol. 2, pp. 101-106).
[11] Yao, Y. (2011). Two semantic issues in a probabilistic rough set model.Fundamenta Informaticae, 108(3-4), 249-265.
[12] Eissa, M. M., Elmogy, M., & Hashem, M. (2014, December). Rough—Granular neural network model for making treatment decisions of Hepatitis C. In Informatics and Systems (INFOS), 2014 9th International Conference on (pp. DEKM-19). IEEE.
[13] Polkowski, L. (2007, June). Granulation of knowledge in decision systems: The approach based on rough inclusions. The method and its applications. In International Conference on Rough Sets and Intelligent Systems Paradigms(pp. 69-79). Springer Berlin Heidelberg.
[14] Polkowski, L., & Artiemjew, P. (2011). Granular computing in the frame of rough mereology. a case study: Classification of data into decision categories by means of granular reflections of data. International journal of intelligent systems, 26(6), 555-571.
[15] Yager, R. R. (2002, October). Using granular objects in multi-source data fusion. In International Conference on Rough Sets and Current Trends in Computing (pp. 324-330). Springer Berlin Heidelberg.
[16] Yao, Y. (2005, July). Perspectives of granular computing. In 2005 IEEE international conference on granular computing (Vol. 1, pp. 85-90). IEEE.
[17] Yao, Y. Y. (2001). Information granulation and rough set approximation.International Journal of Intelligent Systems, 16(1), 87-104.
[18] Guan, Y. Y., Wang, H. K., Wang, Y., & Yang, F. (2009). Attribute reduction and optimal decision rules acquisition for continuous valued information systems. Information Sciences, 179(17), 2974-2984.
[19] Skowron, A., & Wasilewski, P. (2011). Toward interactive rough-granular computing. Control and Cybernetics, 40, 213-235.
[20] Polkowski, L. (2013). Rough Mereology as a Tool for Knowledge Discovery and Reasoning in Intelligent Systems: A Survey.
[21] Skowron, A., Stepaniuk, J., & Swiniarski, R. (2012). Modeling rough granular computing based on approximation spaces. Information Sciences, 184(1), 20-43.
[22] Qian, Y., Liang, J., Wei-zhi, Z. W., & Dang, C. (2011). Information granularity in fuzzy binary GrC model. IEEE Transactions on Fuzzy Systems, 19(2), 253-264.
[23] Ganivada, A., Ray, S. S., & Pal, S. K. (2012). Fuzzy rough granular self-organizing map and fuzzy rough entropy. Theoretical Computer Science, 466, 37-63.
[24] Lin, T. Y., & Louie, E. (2003). Association rules with additional semantics modeled by binary relations. In Rough Set Theory and Granular Computing (pp. 147-156). Springer Berlin Heidelberg.
[25] Riza, L. S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Śle, D., & Benítez, J. M. (2014). Implementing algorithms of rough set theory and fuzzy rough set theory in the R package “roughsets.” Information Sciences, 287, 68-89.
[26] Zheng, H. Z., Chu, D. H., & Zhan, D. C. (2005). Association Rule Algorithm Based on Bitmap and Granular Computing. Artificial Intelligence and Machine Learning (AIML) Journal, 5(3), 51-54.
[27] Polkowski, L. (2013). Rough Mereology as a Tool for Knowledge Discovery and Reasoning in Intelligent Systems: A Survey.
[28] Polkowski, L., & Artiemjew, P. (2011). Granular computing in the frame of rough mereology. a case study: Classification of data into decision categories by means of granular reflections of data. International journal of intelligent systems, 26(6), 555-571.
[29] Zaki, A., Salama, M. A., Hefny, H., & Hassanien, A. E. (2012, December). Rough sets-based rules generation approach: A hepatitis c virus data sets. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 52-59). Springer Berlin Heidelberg.
[30] Badria, F. A., Eissa, M. M., Elmogy, M., & Hashem, M. (2013, October). Rough Based Granular Computing Approach for Making Treatment Decisions of Hepatitis C. In 23rd International Conference on Computer Theory and Applications ICCTA, Alexandria, Egypt (pp. 29-31).
[31] Eissa, M. M., Elmogy, M., Hashem, M., & Badria, F. A. (2014, April). Hybrid rough genetic algorithm model for making treatment decisions of hepatitis C. In Engineering and Technology (ICET), 2014 International Conference on (pp. 1-8). IEEE.
[32] Polkowski, L., & Artiemjew, P. (2007, August). Granular computing: Granular classifiers and

missing values. In 6th IEEE International Conference on Cognitive Informatics (pp. 186-194). IEEE.

[33] Ding, S., & Tong, C. (2011). Research and Comparison of Granularity Cluster Analysis Methods. International Journal of Advancements in Computing Technology, Advanced Institute of Convergence Information Technology,3(7), 154-159.

[34] Garcia, S., Luengo, J., Sáez, J. A., Lopez, V., & Herrera, F. (2013). A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. IEEE Transactions on Knowledge and Data Engineering, 25(4), 734-750.

[35] Ding, S., Chen, J., Xu, X., & Li, J. (2011). Rough neural networks: a review.J Comput Inf Syst, 7(7), 2338-2346.

[36] Jiang, F., & Sui, Y. (2015). A novel approach for discretization of continuous attributes in rough set theory. Knowledge-Based Systems, 73, 324-334.

[37] Zhang, D. B., & Wang, Y. N. (2006). Fuzzy-rough neural network and its application to vowel recognition. Control and Decision, 21(2), 221.

[38] Suguna, N., & Thanushkodi, K. G. (2011). An independent rough set approach hybrid with artificial bee colony algorithm for dimensionality reduction. American Journal of Applied Sciences, 8(3), 261.

[39] Wang, R., Miao, D., & Hu, G. (2006, December). Discernibility matrix based algorithm for reduction of attributes. In Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology (pp. 477-480). IEEE Computer Society.

[40] Badria, F. A., & Attia, H. A. (2007). Effect of selected natural products, thioproline and Pegasys on hepatic platelet activating factor (paf) In ccl4-induced hepatic fibrosis in rats. Saudi Pharmaceutical Journal, 15(2), 96-104.

[41] Barakat S. I., Eissa M. M., EL-Henawy I.(2009). Hybrid Rough Sets and Probabilistic Flow Graph Model In Coronary Artery Disease. In. Egyptian Computer Science Journal, 33.

[42] Amir, M. Z., Eissa, M. M., & EL-Henawy, I. (2010). Hybrid Rough Sets and Decision Tree Model In Coronary Heart Disease. Zagazig University Medical Journal January.