

A FRAMEWORK FOR INTELLIGENT MEDICAL DIAGNOSIS USING ROUGH SET WITH FORMAL CONCEPT ANALYSIS

B. K. Tripathy¹, D. P. Acharjya¹ and V. Cynthia¹

¹School of Computing Science and Engineering, VIT University, Vellore, India
tripathybk@rediffmail.com, dpacharjya@gmail.com,
cynthia.isaacs1987@gmail.com

ABSTRACT

Medical diagnosis process vary in the degree to which they attempt to deal with different complicating aspects of diagnosis such as relative importance of symptoms, varied symptom pattern and the relation between diseases them selves. Based on decision theory, in the past many mathematical models such as crisp set, probability distribution, fuzzy set, intuitionistic fuzzy set were developed to deal with complicating aspects of diagnosis. But, many such models are failed to include important aspects of the expert decisions. Therefore, an effort has been made to process inconsistencies in data being considered by Pawlak with the introduction of rough set theory. Though rough set has major advantages over the other methods, but it generates too many rules that create many difficulties while taking decisions. Therefore, it is essential to minimize the decision rules. In this paper, we use two processes such as pre process and post process to mine suitable rules and to explore the relationship among the attributes. In pre process we use rough set theory to mine suitable rules, whereas in post process we use formal concept analysis from these suitable rules to explore better knowledge and most important factors affecting the decision making.

KEYWORDS

Rough Sets, Formal Concept Analysis (FCA), Information Table, Indiscernibility & Decision Rules

1. INTRODUCTION

Now-a-days Internet is the best example for distributed computing which involves dispersion of data geographically. Therefore, it is challenging for human being to retrieve information from the huge amount of data available geographically at different health research centres. Hence, it is very difficult to extract expert knowledge from the universe of medical dataset. The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians, and mathematicians. Recently it brings attention for computer scientists, particularly in the area of knowledge mining and artificial intelligence. There are many approaches to the problem of how to understand and manipulate imperfect knowledge. The fundamental one is the crisp set. However, it has been extended in many directions as far as modelling of real life situations are concerned. The earliest and most successful one is being the notion of fuzzy sets by L. A. Zadeh [1] that captures impreciseness in information. On the other hand rough sets of Z. Pawlak [2] is another attempt that capture indiscernibility among objects to model imperfect knowledge [3, 4, 5]. There were many other advanced methods such as rough set with similarity, fuzzy rough set, rough set on fuzzy approximation spaces, rough set intuitionistic fuzzy approximation spaces, dynamic rough set, covering based rough set were discussed by different authors to extract knowledge from the huge amount of data [6, 7, 8, 9, 10, 11]. Universe can be considered as a large collection of objects. Each object is associated with some information with it. In order to find knowledge about the universe we need to extract some information about these objects. We need sufficient amount of information to uniquely identify the objects which is not possible in case of all objects. Therefore, we require classification of these objects into similarity classes to characterize these objects in order to extract knowledge about the universe. Rough set is an approach to extract knowledge and association between data and values of data in recent years.

However, it generates too many rules that create many difficulties in taking decision for human being. Hence it is challenging for human being to extract expert knowledge. However, many researchers has analysed medical data by using data mining, fuzzy sets, and formal concept analysis for finding decision rules, and redundancies[12, 13].

In this paper, we use two processes such as pre process and post process to mine suitable rules and to explore the relationship among the attributes. In pre process we use rough set to mine suitable rules, whereas in post process we use formal concept analysis to explore better knowledge and most important characteristics affecting the decision making. The remainder of the paper is organized as follows: Section 2 presents the basics of rough set theory. Section 3 provides the basic idea of formal concept analysis. The proposed intelligent mining model is given in Sections 4. In Section 5, an empirical study on heart disease is presented. This further followed by a conclusion in Section 6.

2. FOUNDATIONS OF ROUGH SET THEORY

Convergence of information and communication technologies brought a radical change in the way medical diagnosis is carried out. It is well established fact that right decision at right time provides an advantage in medical diagnosis. But, the real challenge lies in converting voluminous data into knowledge and to use this knowledge to make proper medical diagnosis. Though present technologies help in creating large databases, but most of the information may not be relevant. Therefore, attribute reduction becomes an important aspect for handling such voluminous database by eliminating superfluous data. Rough set theory developed by Z. Pawlak [14] used to process uncertain and incomplete information is a tool to the above mentioned problem. However, it has many applications in all the fields of science and engineering. One of its strength is the attribute dependencies, their significance among inconsistent data. At the same time, it does not need any preliminary or additional information about the data. Therefore, it classifies imprecise, uncertain or incomplete information expressed in terms of data.

2.1. Rough Sets

In this section we recall the basic definitions of basic rough set theory developed by Z. Pawlak [14]. Let U be a finite nonempty set called the universe. Suppose $R \subseteq U \times U$ is an equivalence relation on U . The equivalence relation R partitions the set U into disjoint subsets. Elements of same equivalence class are said to be indistinguishable. Equivalence classes induced by R are called elementary concepts. Every union of elementary concepts is called a definable set. The empty set is considered to be a definable set, thus all the definable sets form a Boolean algebra and (U, R) is called an approximation space. Given a target set X , we can characterize X by a pair of lower and upper approximations. We associate two subsets $\underline{R}X$ and $\overline{R}X$ called the R – lower and R –upper approximations of X respectively and are given by

$$\underline{R}X = \bigcup \{Y \in U / R : Y \subseteq X\} \tag{1}$$

$$\text{and } \overline{R}X = \bigcup \{Y \in U / R : Y \cap X \neq \emptyset\} \tag{2}$$

The R –boundary of X , $BN_R(X)$ is given by $BN_R(X) = \overline{R}X - \underline{R}X$. We say X is rough with respect to R if and only if $\overline{R}X \neq \underline{R}X$, equivalently $BN_R(X) \neq \emptyset$. X is said to be R – definable if and only if $\overline{R}X = \underline{R}X$ or $BN_R(X) = \emptyset$. So, a set is rough with respect to R if and only if it is not R – definable.

2.2. Information System

As mentioned in the previous section rough set philosophy is based on the assumption that, in addition to crisp set theory, we have some additional information about the elements of a universe of discourse. Elements that exhibit the same information are indiscernible and form the

basic building blocks that can be considered as elementary concept of knowledge about the universe.

An information system is a table that provides a convenient way to describe a finite set of objects called the universe by a finite set of attributes thereby representing all available information and knowledge. The attribute sets along with the objects in an information system consists of the set of condition attributes and decision attributes. Therefore it is also named as decision table [15]. Let us denote the information system as $I = (U, A, V, f)$, where U is a finite non-empty set of objects called the universe and A is a non-empty finite set of attributes. For every $a \in A$, V_a is the set of values that attribute a may take. Also $V = \bigcup_{a \in A} V_a$. In addition, for every $a \in A$, $f_a : U \rightarrow V_a$ is the information function. Also $f : U \times A \rightarrow V$ [16].

Consider a medical information system Table 1 of ten patients $p_i; i=1, 2, \dots, 10$ as the set of objects of the universe with a set of attributes cough, vomiting, cold severing, nasal bleeding, temperature, delirium and fever. The attribute fever is considered as the decision attribute. Let us assume that $a_1, a_2, a_3, a_4, a_5, a_6, d$ as cough, vomiting, cold severing, nasal bleeding, temperature, delirium and fever respectively. In particular, patient (p_1) is characterized in the table by the attribute value set (cough, always), (vomiting, seldom), (cold severing, seldom), (nasal bleeding, never), (temperature, high), (delirium, never), and (fever, yes) which form the information about the patient.

Table 1. Classical view of the medical information system

| Patient | Cough | Vomiting | Cold Severing | Nasal bleeding | Temperature | Delirium | Fever |
|----------|--------|----------|---------------|----------------|-------------|----------|-------|
| p_1 | Always | Seldom | Seldom | Never | High | Never | Yes |
| p_2 | Seldom | Never | Always | Never | Normal | Never | No |
| p_3 | Never | Always | Seldom | Never | Very high | Never | Yes |
| p_4 | Always | Seldom | Seldom | Always | Normal | Never | No |
| p_5 | Never | Always | Seldom | Always | High | Never | No |
| p_6 | Never | Seldom | Always | Always | Normal | Never | Yes |
| p_7 | Always | Always | Seldom | Seldom | Normal | Never | No |
| p_8 | Seldom | Never | Always | Seldom | Very high | Never | Yes |
| p_9 | Always | Always | Never | Never | High | Never | Yes |
| p_{10} | Seldom | Seldom | Seldom | Seldom | High | Never | Yes |

To make our analysis simpler we have assigned values 1, 2, 3, 4, 5, 6 to the attribute values always, seldom, never, normal, high, and very high respectively. Also, we have assigned 1, and 2 for decision values yes, no respectively. However, these values are optional and do not affect the analysis. The reduced medical information system is presented in Table 2.

Table 2. Reduced medical information system

| Patient | a_1 | a_2 | a_3 | a_4 | a_5 | a_6 | d |
|----------|-------|-------|-------|-------|-------|-------|-----|
| p_1 | 1 | 2 | 2 | 3 | 5 | 3 | 1 |
| p_2 | 2 | 3 | 1 | 3 | 4 | 3 | 2 |
| p_3 | 3 | 1 | 2 | 3 | 6 | 3 | 1 |
| p_4 | 1 | 2 | 2 | 1 | 4 | 3 | 2 |
| p_5 | 3 | 1 | 2 | 1 | 5 | 3 | 2 |
| p_6 | 3 | 2 | 1 | 1 | 4 | 3 | 1 |
| p_7 | 1 | 1 | 2 | 2 | 4 | 3 | 2 |
| p_8 | 2 | 3 | 1 | 2 | 6 | 3 | 1 |
| p_9 | 1 | 1 | 3 | 3 | 5 | 3 | 1 |
| p_{10} | 2 | 2 | 2 | 2 | 5 | 3 | 1 |

2.3. Indiscernibility Relation

Universe can be considered as a large collection of objects. Each object is associated with some information (data, knowledge) within it. In order to find knowledge about the universe we need to process these attribute values. Therefore, we require sufficient amount of information to uniquely identify, classify these objects into similar classes and to extract knowledge about the universe. The classification of the objects of the universe is done based on indiscernibility relation among these objects. It indicates that objects of a class cannot discern from one another based on available set of attributes of the objects [14, 17]. The indiscernibility relation generated in this way is the mathematical basis of rough set theory. Any set of all indiscernible objects is called an elementary concept, and form a basic granule (atom) of knowledge about the universe. Any union of the elementary sets is referred to be either crisp (precise) set or rough (imprecise) set. Let $P \subseteq A$ and $x_i, x_j \in U$. Then we say x_i and x_j are indiscernible by the set of attributes P in A if and only if the following (3) holds.

$$f(x_i, a) = f(x_j, a), \quad \forall a \in P \quad (3)$$

For example, given the attributes cough, vomiting, cold severing, and delirium $\{p_1, p_4\}$ are indiscernible. Similarly, the other indiscernible classes obtained are $\{p_2, p_8\}, \{p_3, p_5\}, \{p_6\}, \{p_7\}, \{p_9\}$, and $\{p_{10}\}$. In general each patient $p_i, i=1, 2, \dots, 10$ is compared with each other cell wise to find the indiscernibility in the attribute value. From the data set Table-1, on considering the attributes $A = \{\text{cough, vomiting, cold severing, delirium}\}$, we get the family of equivalence classes of A , i.e., the partition determined by set of attributes A , denoted by U/A or $I(A)$. Therefore,

$$U/A = \{\{p_1, p_4\}, \{p_2, p_8\}, \{p_3, p_5\}, \{p_6\}, \{p_7\}, \{p_9\}, \{p_{10}\}\}$$

On considering the target set $X = \{p_1, p_6, p_9, p_{10}\}$; patients p_1 and p_4 are the boundary-line objects, where

$$\underline{AX} = \bigcup \{Y \in U/A : Y \subseteq X\} = \{p_6, p_9, p_{10}\} \quad \text{and}$$

$$\overline{AX} = \bigcup \{Y \in U/A : Y \cap X \neq \emptyset\} = \{p_1, p_4, p_6, p_9, p_{10}\}$$

Furthermore, considering the attributes $P \subseteq A$, we can associate an index (i.e. $\alpha_A(X)$) called the accuracy of approximation for any set $X \subseteq U$ as follows:

$$\alpha_A(X) = \frac{\text{cardinality of } \underline{AX}}{\text{cardinality of } \overline{AX}} = \frac{|\underline{AX}|}{|\overline{AX}|}$$

For example, $\alpha_A(X) = \frac{3}{5}$, where $X = \{p_1, p_6, p_9, p_{10}\}$; and $A = \{\text{cough, vomiting, cold severing, delirium}\}$. From the information system given in Table 1, it is clear that patient p_1 has fever, whereas patient p_4 do not and they are indiscernible with respect to the attributes cough, vomiting, cold severing, and delirium. Hence, fever can not be characterized in terms of symptoms and symptoms-value pair (cough, always), (vomiting, seldom), (cold severing, seldom), and (delirium, never). Therefore, patients p_1 and p_4 are the boundary line objects and can not be classified in view of the available knowledge. On the other hand the patients p_6, p_9 , and p_{10} display symptoms with certainty as having fever. Therefore, the lower approximation of X is $\{p_6, p_9, p_{10}\}$, whereas the upper approximation is $\{p_1, p_4, p_6, p_9, p_{10}\}$.

2.4. Reduct and Rule Discovery

One of the important aspects of the rough set theory is the attribute reduction and core. In an information system, some attributes may be redundant and useless. If those redundant and useless attributes are removed without affecting the classification power of attributes, we can call them the superfluous attributes [18]. The core concept is commonly used in all reducts [19].

If the set of attributes is dependent, using the dependency properties of attributes, we are interested in finding all possible minimal subsets of attributes which have the same number of elementary sets without loss of the classification power of the reduced information system [20]. In order to express the above notions more clearly we need some auxiliary notations. Let $P \subseteq A$ and $a \in P$. We say that the attribute a is dispensable in P if the following condition (4) holds; otherwise a is indispensable in P .

$$I(P) = I(P - \{a\}) \tag{4}$$

Set P is independent if all its attributes are indispensable. Reduct (P') of P is a subset of attributes P such that the equivalence class induced by the reduced attribute set P' are the same as the equivalence class structure induced by the attribute set P . i.e., $I(P) = I(P')$. The core of the attribute set P is the set of all indispensable attributes of P . The important property connecting the notion of core and reducts is defined in (5), where $\text{Red}(P)$ is the set of all reducts of P .

$$\text{Core}(P) = \bigcap \text{Red}(P) \tag{5}$$

An information system is defined as $I = (U, A, V, f)$, where U is a non empty finite set of objects, $A = C \cup D$ is a non empty finite set of attributes, C denotes the set of condition attributes and D denotes the set of decision attributes, $C \cap D = \emptyset$. According to Pawlak [21], a decision rule S in information system is expressed as $\varphi \rightarrow \psi$, where φ and ψ are conditions and decisions of the decision rule S respectively. There are three measurements for decision rules. The first one is the accuracy of the decision rule. The second one is the support of a rule whereas the third measurement is the strength of the decision rule. The support of a rule is defined as $\text{Supp}_s(\varphi, \psi) = \text{Card}(\|\varphi \wedge \psi\|_s)$, whereas the strength of the decision rule $\varphi \rightarrow \psi$ is defined as:

$$\sigma(\varphi, \psi) = \frac{\text{Supp}_s(\varphi, \psi)}{\text{Card}(\|U\|_s)} \tag{6}$$

It implies that stronger rules cover more number of supporting objects and its strength can be calculated by using relation (6). The relative decision rules for Table 2 are given in Table 3, where we use $R_{i,d}$ for i^{th} rule in the decision class d .

Table 3. Relative decision rules

| Rule Number | Notation | Decision class | Meaning | Support | Strength rate (%) | Accuracy rate (%) |
|-------------|-----------|----------------|--|---------------|-------------------|-------------------|
| 1 | $R_{1,1}$ | 1 | If $(\alpha_3 = 2) \ \& \ (\alpha_1 = 5)$, then $d = 1$ | p_1, p_{10} | 33.33 | 100 |
| 2 | $R_{2,1}$ | 1 | If $(\alpha_3 = 6)$, then $d = 1$ | p_3, p_8 | 33.33 | 100 |
| 3 | $R_{3,1}$ | 1 | If $(\alpha_1 = 3)$, then $d = 1$ | p_6 | 16.17 | 100 |
| 4 | $R_{4,1}$ | 1 | If $(\alpha_3 = 3) \ \& \ (\alpha_1 = 1)$, then $d = 1$ | p_6 | 16.17 | 100 |
| 5 | $R_{5,2}$ | 2 | If $(\alpha_1 = 1) \ \& \ (\alpha_3 = 4)$, then $d = 2$ | p_4, p_7 | 50 | 100 |
| 6 | $R_{6,2}$ | 2 | If $(\alpha_3 = 1) \ \& \ (\alpha_1 = 3)$, then $d = 2$ | p_2 | 25 | 100 |
| 7 | $R_{7,2}$ | 2 | If $(\alpha_1 = 1) \ \& \ (\alpha_1 = 1)$, then $d = 2$ | p_3 | 25 | 100 |

3. FORMAL CONCEPT ANALYSIS AND BACKGROUND

Formal concept analysis (FCA) introduced by R. Wille [22] provides conceptual tools for the analysis of data, and it has applied to many quite different fields such as psychology, sociology, anthropology, medicine, biology, linguistics, computer sciences and industrial engineering. The main purpose of the method is to visualize the data in form of concept lattices and thereby to

make them more transparent and more easily discussible and criticisable. At the same time it allows knowledge acquisition from (or by) an expert by putting very precise questions, which either have to be confirmed or to be refuted by a counterexample.

The main objective is to support the user in analyzing and structuring a domain of interest based on mathematization of the concept and conceptual hierarchy. It activates mathematical thinking for conceptual data analysis and knowledge processing based on a formal understanding of a concept as a unit of thought. Concepts can be philosophically understood as the basic units of thought formed in dynamic processes within social and cultural environments. Therefore, by philosophical tradition a concept is comprised of extension and intension. The extension of a formal concept is formed by all objects to which the concept applies and the intension that consists of all attributes existing in those objects. The set of objects, attributes and the relations between an object and an attribute in a dataset form the basic conceptual structure of FCA known as formal context. Concepts can only lay on relationships with many other concepts where the subconcept-superconcept relation plays a vital role. A subconcept of a superconcept means that the extension of the subconcept is contained in the extension of the super concept. This is equivalent to the relationship that the intension of the subconcept contains the intension of the superconcept [23].

3.1. Formal Context and Formal Concept

In this section we recall the basic definitions and notations of formal concept analysis developed by R. Wille [22]. A formal context is defined as a set structure $K = (U, A, R)$ consists of two sets U and A while R is a binary relation between U and A , i.e. $R \subseteq U \times A$. The elements of U are called the objects and the elements of A are called the attributes of the context. The formal concept of the formal context (U, A, R) is defined with the help of derivation operators. The derivation operators are defined for arbitrary $X \subseteq U$ and $Y \subseteq A$ as follows:

$$X' = \{a \in A : uRa \quad \forall u \in X\} \tag{7}$$

$$Y' = \{u \in U : uRa \quad \forall a \in Y\} \tag{8}$$

A formal concept of a formal context $K = (U, A, R)$ is defined as a pair (X, Y) with $X \subseteq U$, $Y \subseteq A$, $X = Y'$, and $Y = X'$. The first member X , of the pair (X, Y) is called the extent whereas the second member Y is called the intent of the formal concept. Objects in X share all properties Y , and only properties Y are possessed by all objects in X . A basic result is that the formal concepts of a formal context are always forming the mathematical structure of a lattice with respect to the subconcept-superconcept relation. Therefore, the set of all formal concepts forms a complete lattice called a concept lattice [22]. The subconcept-superconcept relation can be depicted best by a lattice diagram and we can derive concepts, implication sets, and association rules based on the cross table. In such a diagram the name of each object is attached to its represented object concept and the name of each attribute is attached to its represented attribute concept. The subconcept-superconcept relation is transitive. It means that, a concept is the subconcept of any concept that can be reached by travelling upwards from it. Now we present the cross table of the decision rules for decision class 1 obtained in Table 3 is given in Table 4, where the rows are represented as objects and columns are represented as attributes. The relation between them is represented by a cross.

Table 4. Cross table of decision rules for decision class 1

| Rule Number | Notation | α_1 | | | α_2 | | | α_3 | | | α_4 | | | |
|-------------|-----------|------------|-----------|-----------|------------|-----------|-----------|------------|-----------|-----------|------------|-----------|-----------|-----------|
| | | $A_{1,1}$ | $A_{1,2}$ | $A_{1,3}$ | $A_{2,1}$ | $A_{2,2}$ | $A_{2,3}$ | $A_{3,1}$ | $A_{3,2}$ | $A_{3,3}$ | $A_{4,1}$ | $A_{4,2}$ | $A_{4,3}$ | $A_{4,4}$ |
| 1 | $R_{1,1}$ | | | | | × | | | | | | | × | |
| 2 | $R_{2,1}$ | | | | | | | | | | | | | × |
| 3 | $R_{3,1}$ | | | | | | | | | × | | | | |
| 4 | $R_{4,1}$ | | | × | | | | × | | | | | | |

The corresponding lattice diagram is presented in Figure 1, where the nodes represent formal concepts. In particular, one concept consists of object 1 and attributes A_{22} , A_{55} , is expressed as $(\{1\}, \{A_{22}, A_{55}\})$. Similarly, another concept consists of object 2 and attribute A_{56} is expressed as $(\{2\}, \{A_{56}\})$. It is also observed that, lines up give more general concepts whereas lines down give more specific concepts. A pair of set of objects and a set of attributes that is closed in this manner is called a formal concept [24].

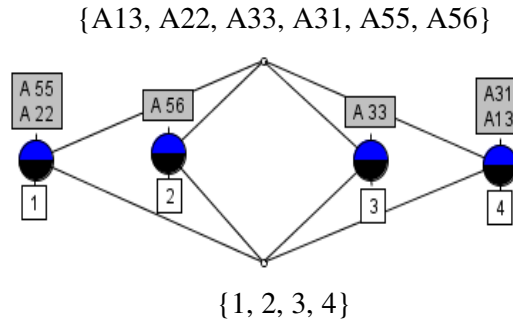


Figure 1. The lattice diagram of the decision rules for decision class 1

4. PROPOSED INTELLIGENT MINING MODEL

In this section, we propose our intelligent mining model that consists of problem definition, target data, preprocessed data, processed data, data partition, rule discovery, and formal concept analysis as shown in Figure 2.

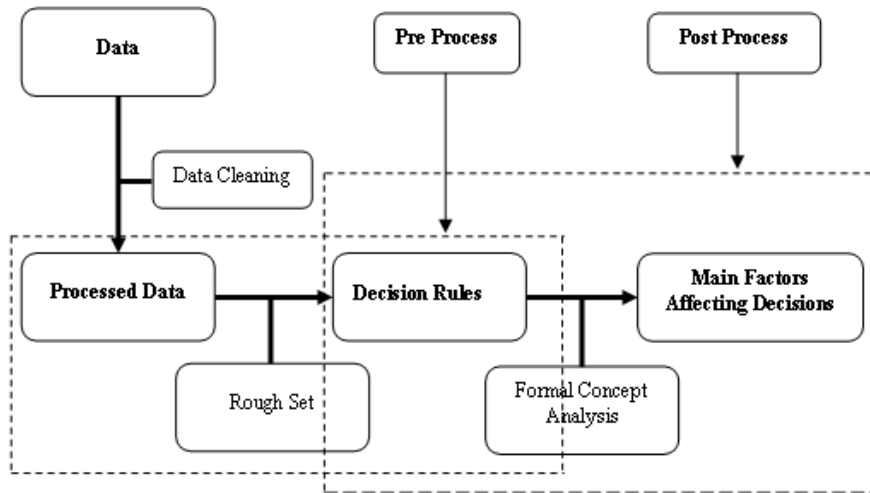


Figure 2. Proposed intelligent mining model

The fundamental step of any model in which we identify the right problem is the problem definition. Incorporation of prior knowledge is always associated with the problem definition however, the potential validity or usefulness of an individual data element or pattern of data elements may change dramatically from individual to individual, organization to organization, or task to task because of the acquisition of knowledge and reasoning may involve in vagueness and incompleteness. It is very difficult for human to find useful information that is hidden in the accumulated voluminous data. Therefore, most important challenge is to retrieve data pattern from the accumulated voluminous data. There is much need for dealing with the incomplete and vague information in classification, concept formulation, and data analysis. To this end here we

use two processes such as pre process and post process to mine suitable rules and to explore the relationship among the attributes. In pre process as shown in Figure 2 we use rough set theory for processing data, data classification after removal of noise and missing data to mine suitable rules. In post process we use formal concept analysis from these suitable rules to explore better knowledge for decision making.

The motivation behind this study is that the two theories aim at different goals and summarize different types of knowledge. Rough set theory is used for prediction whereas formal concept analysis is used for description. Therefore, the combination of both with domain intelligence leads to better knowledge.

4.1. Pre process Architecture Design

In this section, we present our pre process architecture design that consists of problem definition, data preparation, data partition, rule generation, domain intelligence and rule validation as shown in Figure 3. Problem definition and incorporation of prior knowledge are the fundamental steps of any model in which we identify the right problem. Secondly, proper structuring the corresponding objectives and the associated attributes is done. Finally, a target dataset is created on which data mining is to be performed. Before further analysis, a sequence of data cleaning tasks such as consistency check, removing noise and data completeness is done to ensure that the data are as accurate as possible. Now, we discuss in detail the subsequent steps of the pre process architecture design of the proposed model.

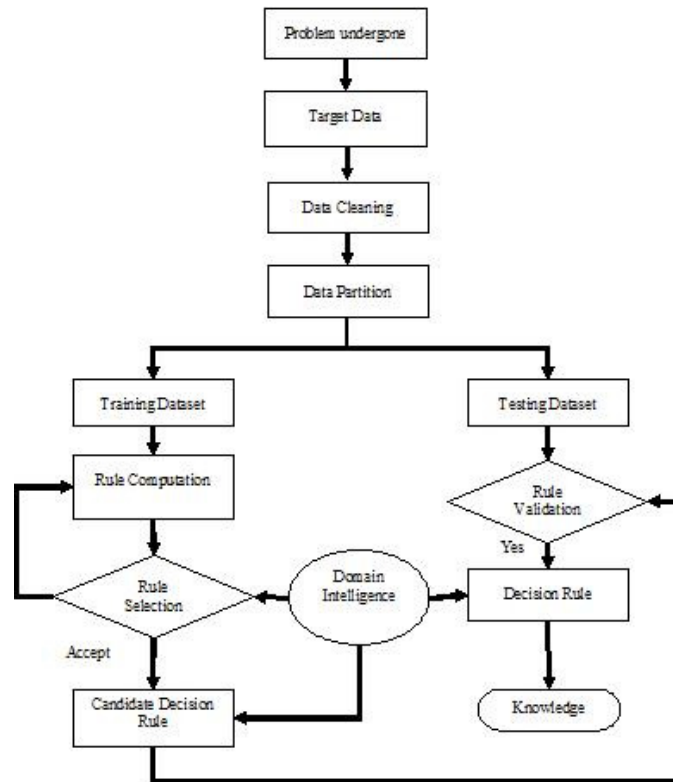


Figure 3. Pre process architecture design

4.1.1. Rule Computation Procedure

Now, we propose a rule generation algorithm to generate all the possible reducts by eliminating all dispensable attributes and derive the candidate decision rules from the training dataset as below:

1. Set object number $i = 1$

2. Choose object i from the training data set and compute a set of reducts for all the condition attributes.
3. Replace $i = i + 1$. If all objects have been chosen, then go to step 4; else go to step 2.
4. Compute the number of supporting objects for each reduct after combining the identical reducts.
5. Using the domain intelligent system, evaluate the generated reducts based and compute the number of supporting objects. If candidate decision rules are satisfied, then go to next step 6; else restore the objects associated with the unsatisfied rule and move to step 2.
6. Obtain the decision rules from the selected reducts.
7. Terminate the process and further validate the output results.

4.1.2 Rule-validation Procedure

To examine the objects in the testing dataset to estimate the validity of the rules derived from the rule composing algorithm, we apply the following steps in rule validation procedure.

1. Compare each decision rule obtained from the above algorithm with each new object from the testing data set. Compute the number of objects that support the rule.
2. Repeat step 1 for all decision rules obtained from the rule composing algorithm.
3. Calculate the accuracy of each rule by using the following equation (9).

$$Accuracy = \frac{\text{Total number of supported objects}}{\text{Sum of supported and non-supported objects}} \quad (9)$$

4. If the accuracy is greater than the predefined threshold, then go to step 5; else remove the rule.
5. Process terminates and writes the validated rules.

5. AN EMPIRICAL STUDY ON HEART DISEASES

In this empirical study we apply rough set theory for mining knowledge over the data set obtained from different research centers and finally we will find the major factor affecting the decision using formal concept analysis. However, we keep the identity confidential due to some specified official reason. In this study, we have collected historical data from different health research institutions. The 1487 patient's cases were checked for completeness and consistency. In order to avoid unnecessary complexity, we removed unrelated items in the data set. We have removed 553 patient details from that data set as they have no heart disease. In addition to this, 261 patient's data were also removed from the dataset, because none of them had sufficient support. Also, 117 patient's data were removed from the dataset, because of missing attribute values. In total, 931 patient's data were removed from the dataset. Also, we have discussed and attended several meeting with domain experts for getting the clear idea about the heart diseases. This process gave us an understanding of the historical data and essential attribute (symptom) knowledge about the heart disease. However, for completeness of the paper we state briefly about these symptoms.

Table 5. Symptoms notation representation table

| Symptoms | Abbreviation | Notation | Symptoms | Abbreviation | Notation |
|---------------------|--------------|------------|-------------------|--------------|---------------|
| Chest pain | CP | α_1 | Exercise | EX | α_7 |
| Blood pressure | BP | α_2 | Old peak | OP | α_8 |
| Cholesterol | CH | α_3 | Thallium scan | TS | α_9 |
| Blood sugar | BS | α_4 | Sex | SX | α_{10} |
| Electrocardiography | ECG | α_5 | Age | AG | α_{11} |
| Maximum heart rate | MHR | α_6 | Type of diagnosis | TD | d |

The most common symptom of heart disease is chest pain and it is of four types, viz. typical angina, atypical angina, non-anginal pain and asymptomatic. The other symptoms included are

blood pressure, cholesterol, blood sugar, electrocardiography, maximum heart rate, exercise, old peak, thallium scan, sex and age. Each patient's treatment is different and depends on several factors. Thus, it is essential to identify certain rules and the chief factors so that a patient can identify the disease at an early stage. It can also reduce the financial burden of a patient. Here the data is collected only to diagnosis the heart disease and then suggest the type of heart surgery required. Also, we analyze the historical data to provide the chief factors for each case of the diagnosis. Numerical and literature values based on different symptoms were collected and these parameters became our set of attributes. We consider the diagnosis decision of the patients as the decision variable. The attributes that play major role in heart disease and the notations that are used in our analysis is presented in Table 5. In particular to start with pre process, we randomly divide the 556 patients into the training data set that contained 306 patients (55%) and the testing data set that contained 250 patients (45%). The normalized information is given in Table 6. To make our analysis simpler we have assigned some value to each classification group and these are optional for analysis.

Table 6. Normalized information table

| Attribute | Normalized value | Classification |
|---------------------------|--|--|
| Chest pain (CP) | --- | Typical angina (1) Atypical angina (2) Non-anginal pain (3) Asymptomatic (4) |
| Blood pressure (BP) | 120 to 139 / 80 to 90 140 to 159 / 90 to 99 160 to 179 / 100 to 109 ≥ 180 / 110 | Normal (1) Medium (2) High (3) Very high (4) |
| Cholesterol (CH)-LDL | <160 160-190 190-250 >250 | Low (1) Medium (2) High (3) Very high (4) |
| Fasting Blood sugar (FBS) | <126 >126 | Normal (1) Very high (2) |
| Electrocardiography (ECG) | [-0.5,0.4] [2.45,1.8] [1.4,2.5] | Normal (1) ST-T abnormal (2) Hypertrophy (3) |
| Maximum heart rate (MHR) | <60 60-100 >100 | Medium (1) Normal (2) High (3) |
| Exercise (EX) | --- | False (1) True (2) |
| Old peak (OP) | <2 2-3 3> | Low (1) Risk (2) Terrible (3) |
| Thallium scan (TS) | 3 6 7 | Normal (1) Fixed defect (2) Reversible Defect (3) |
| Sex (SX) | --- | Male (1) Female (2) |
| Age (Age) | <35 36-60 60-75 >75 | Young (1) Mild (2) Old (3) Very old (4) |
| Type of diagnosis (TD) | --- | Hypertensive heart disease (1) Coronary heart disease (2) Heart failure (3) Potential patient (4) Cardiomyopathy (5) |

5.1. Pre process of Empirical Study

In this section, we discuss in detail the subsequent steps of the pre process architecture design of the empirical study taken under consideration. The training data set of 306 patients obtained is

further classified into five decision classes- decision class 1, 2, 3, 4, and 5- which expresses different diagnosis types: hypertensive heart disease, coronary heart disease, heart failure, potential patient, and cardiomyopathy respectively. The number of patients under each decision class is obtained as 71, 60, 49, 69, and 57 respectively. Here, decision class 1 is selected as an example in this section.

5.1.1. Rule Generation and Selection

We employed the training dataset to derive the reducts and select the final rules with the help of domain intelligence experts. We also removed the unusual outcomes. In addition to this, the identical objects in the training dataset were reduced to only one case in order to avoid unnecessary analysis. Based on the rule generation algorithm, the reducts were determined and presented in Table7. We have generated and summarized the candidate rules based on the number of supporting patients and domain intelligent experts. This is presented in Table 8 for further validation.

Table 7. Rules generated for the decision class 1

| Rules | a_1 | a_2 | a_3 | a_4 | a_5 | a_6 | a_7 | a_8 | a_9 | a_{10} | a_{11} | d | Number of supporting patients |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|-----|-------------------------------|
| [1] | x | 4 | x | x | 2 | x | x | 2 | x | 1 | x | 1 | 8 |
| [2] | 2 | 3 | x | x | x | 3 | 1 | x | x | x | x | 1 | 2 |
| [3] | 3 | 1 | 1 | x | 3 | x | x | x | x | x | x | 1 | 6 |
| [4] | x | x | x | x | 1 | x | 2 | x | x | x | 4 | 1 | 3 |
| [5] | 2 | x | 1 | x | x | x | x | x | 2 | x | x | 1 | 3 |
| [6] | x | x | x | 2 | x | x | x | x | x | x | x | 1 | 2 |
| [7] | x | 4 | 4 | x | x | x | x | x | 3 | x | 2 | 1 | 6 |
| [8] | 3 | x | x | x | x | x | x | 1 | x | x | x | 1 | 3 |
| [9] | x | 1 | 4 | x | x | x | 1 | 3 | x | x | x | 1 | 2 |
| [10] | 2 | 3 | x | x | x | x | x | x | x | x | 1 | 1 | 3 |
| [11] | x | x | 2 | x | x | x | x | x | 1 | 2 | x | 1 | 2 |
| [12] | 1 | x | x | x | 1 | 3 | x | x | x | x | x | 1 | 7 |
| [13] | 3 | 3 | x | x | x | 1 | x | x | x | x | 3 | 1 | 10 |
| [14] | 2 | x | 2 | x | 2 | x | x | x | x | x | x | 1 | 4 |
| [15] | x | x | x | x | 1 | 2 | x | x | x | x | x | 1 | 2 |
| [16] | x | x | x | x | x | 2 | 2 | x | x | x | 1 | 1 | 4 |
| [17] | x | 1 | 3 | x | x | x | 1 | 3 | x | x | x | 1 | 2 |
| [18] | x | 4 | 4 | x | x | x | 1 | 3 | x | x | x | 1 | 2 |

Table 8. candidate rules for hypertensive heart disease

| Rules selected | a_1 | a_2 | a_3 | a_4 | a_5 | a_6 | a_7 | a_8 | a_9 | a_{10} | a_{11} | d | Number of supporting patients |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|-----|-------------------------------|
| [1] | x | 4 | x | x | 2 | x | x | 2 | x | 1 | x | 1 | 8 |
| [2] | 2 | 3 | x | x | x | 3 | 1 | x | x | x | x | 1 | 2 |
| [3] | 3 | 1 | 1 | x | 3 | x | x | x | x | x | x | 1 | 6 |
| [4] | x | x | x | x | 1 | x | 2 | x | x | x | 4 | 1 | 3 |
| [5] | 2 | x | 1 | x | x | x | x | x | 2 | x | x | 1 | 3 |
| [7] | x | 4 | 4 | x | x | x | x | x | 3 | x | 2 | 1 | 7 |
| [9] | x | 1 | 4 | x | x | x | 1 | 3 | x | x | x | 1 | 2 |
| [10] | 2 | 3 | x | x | x | x | x | x | x | x | 1 | 1 | 3 |
| [12] | 1 | x | x | x | 1 | 3 | x | x | x | x | x | 1 | 7 |
| [13] | 3 | 3 | x | x | x | 1 | x | x | x | x | 3 | 1 | 9 |
| [14] | 2 | x | 2 | x | 2 | x | x | x | x | x | x | 1 | 4 |
| [17] | x | 1 | 3 | x | x | x | 1 | 3 | x | x | x | 1 | 2 |
| [18] | x | 4 | 4 | x | x | x | 1 | 3 | x | x | x | 1 | 2 |

5.1.2 The Rule Validation

We used the testing data set to examine the accuracy of each candidate rule generated, to estimate the corresponding validity. Unlike training data set, the testing data set of 250 patients

is classified into five decision classes- decision class 1, 2, 3, 4, and 5- which expressed different diagnosis types: hypertensive heart disease, coronary heart disease, heart failure, potential patient, and cardiomyopathy respectively for further validation. The number of patients under each decision class is obtained as 63, 51, 38, 55, and 43 respectively. Finally, we expressed these candidate rules in Table 9. The total number of supporting, non-supporting patients and the accuracy of each candidate rule are mentioned in the right hand column of the Table 9. Here, decision class 1 is selected as an example in this section.

Since the accuracy of candidate rules 4, and 10 is less than the predefined domain intelligence threshold 60%, these two rules are discarded while validation. Thus the rules selected from pre process are 1, 2, 3, 5, 7, 9, 12, 13, 14, 17, 18 is the input to the post process for identifying the chief factors. For example, rule 1 can be stated as: IF blood pressure is very high, electrocardiography (ECG) is ST-T abnormal, maximum heart rate (MHR) is normal, and sex is male THEN we can infer that the heart disease is hypertensive heart disease. Similarly, the other rules can also be obtained from the given Table 9.

Table 9. Candidate rule validation

| Rule | Description | | Support | Non-support |
|------|-------------|--|----------------|-------------|
| [1] | IF | $a_1 = 4, a_2 = 2, a_3 = 2, \text{ and } a_{10} = 1$ | 8 | 1 |
| | THEN | we can infer that the heart disease is hypertensive heart disease. ($d = 1$) | Accuracy: 88% | |
| [2] | IF | $a_1 = 2, a_2 = 3, a_3 = 3, \text{ and } a_4 = 1$ | 3 | 1 |
| | THEN | $d = 1$ | Accuracy: 75% | |
| [3] | IF | $a_1 = 3, a_2 = 1, a_3 = 1, \text{ and } a_4 = 3$ | 6 | 0 |
| | THEN | $d = 1$ | Accuracy: 100% | |
| [4] | IF | $a_1 = 1, a_2 = 2, \text{ and } a_{11} = 4$ | 0 | 2 |
| | THEN | $d = 1$ | Accuracy: 0% | |
| [5] | IF | $a_1 = 2, a_2 = 1, \text{ and } a_3 = 2$ | 3 | 1 |
| | THEN | $d = 1$ | Accuracy: 75% | |
| [7] | IF | $a_1 = 4, a_2 = 4, a_3 = 3, \text{ and } a_{11} = 2$ | 6 | 1 |
| | THEN | $d = 1$ | Accuracy: 85% | |
| [9] | IF | $a_1 = 1, a_2 = 4, a_3 = 1, \text{ and } a_4 = 3$ | 2 | 1 |
| | THEN | $d = 1$ | Accuracy: 66% | |
| [10] | IF | $a_1 = 2, a_2 = 3, \text{ and } a_{11} = 1$ | 1 | 2 |
| | THEN | $d = 1$ | Accuracy: 33% | |
| [12] | IF | $a_1 = 1, a_2 = 1, \text{ and } a_3 = 3$ | 6 | 1 |
| | THEN | $d = 1$ | Accuracy: 85% | |
| [13] | IF | $a_1 = 3, a_2 = 3, a_3 = 1, \text{ and } a_{11} = 3$ | 5 | 1 |
| | THEN | $d = 1$ | Accuracy: 83% | |
| [14] | IF | $a_1 = 2, a_2 = 2, \text{ and } a_3 = 2$ | 3 | 2 |
| | THEN | $d = 1$ | Accuracy: 60% | |
| [17] | IF | $a_1 = 1, a_2 = 3, a_3 = 1, \text{ and } a_4 = 3$ | 2 | 1 |
| | THEN | $d = 1$ | Accuracy: 66% | |
| [18] | IF | $a_1 = 4, a_2 = 4, a_3 = 1, \text{ and } a_4 = 3$ | 3 | 1 |
| | THEN | $d = 1$ | Accuracy: 75% | |

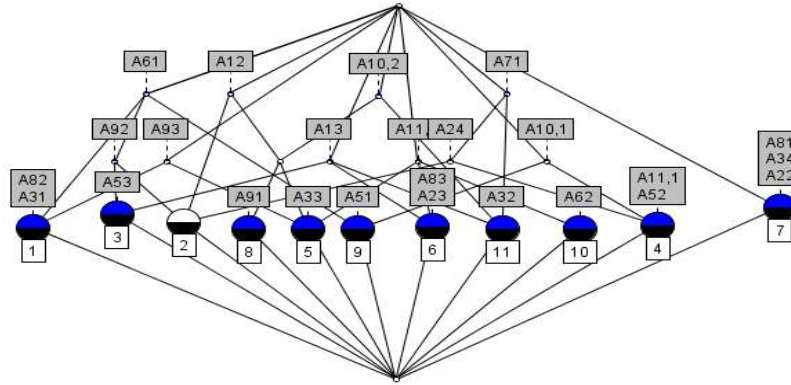


Figure 5 Lattice diagram of the decision class coronary heart disease

Table 14. Implication set table for decision class coronary heart disease

| Implication Sets | | |
|--|---|---|
| 1 < 1 > A13 A23 \implies A83 A113 | 20 < 1 > A83 \implies A13 A23 A113 | 39 < 1 > A92 A113 \implies A12 A33 A102 |
| 2 < 1 > A12 A24 \implies A61 A71 A92 | 21 < 1 > A13 A92 \implies A53 A61 | 40 < 1 > A102 A113 \implies A12 A33 A92 |
| 3 < 1 > A33 \implies A12 A92 A102 A113 | 22 < 1 > A24 A92 \implies A12 A61 A71 | 41 < 1 > A12 A114 \implies A22 A93 |
| 4 < 1 > A22 A34 \implies A81 | 23 < 1 > A53 A92 \implies A13 A61 | 42 < 1 > A22 A114 \implies A12 A93 |
| 5 < 1 > A41 \implies A61 A93 | 24 < 1 > A71 A92 \implies A12 A24 A61 | 43 < 1 > A23 A114 \implies A71 A93 A102 |
| 6 < 1 > A52 \implies A24 A71 A101 A111 | 25 < 1 > A31 A93 \implies A61 A82 | 44 < 1 > A24 A114 \implies A14 A34 |
| 7 < 1 > A13 A53 \implies A61 A92 | 26 < 1 > A101 \implies A24 A52 A71 A111 | 45 < 1 > A34 A114 \implies A14 A24 |
| 8 < 1 > A24 A53 \implies A34 A102 | 27 < 1 > A12 A102 \implies A33 A92 A113 | 46 < 1 > A71 A114 \implies A23 A93 A102 |
| 9 < 1 > A31 A53 \implies A111 | 28 < 1 > A24 A102 \implies A34 A53 | 47 < 1 > A102 A114 \implies A23 A71 A93 |
| 10 < 1 > A34 A53 \implies A24 A102 | 29 < 1 > A34 A102 \implies A24 A53 | 48 < 1 > A14 \implies A24 A34 A114 |
| 11 < 1 > A12 A61 \implies A24 A71 A92 | 30 < 1 > A53 A102 \implies A24 A34 | 49 < 1 > A12 A22 \implies A93 A114 |
| 12 < 1 > A13 A61 \implies A53 A92 | 31 < 1 > A92 A102 \implies A12 A33 A113 | 50 < 1 > A23 A71 \implies A93 A102 A114 |
| 13 < 1 > A24 A61 \implies A12 A71 A92 | 32 < 1 > A24 A111 \implies A52 A71 A101 | 51 < 1 > A12 A93 \implies A22 A114 |
| 14 < 1 > A31 A61 \implies A82 A93 | 33 < 1 > A31 A111 \implies A53 | 52 < 1 > A22 A93 \implies A12 A114 |
| 15 < 1 > A53 A61 \implies A13 A92 | 34 < 1 > A53 A111 \implies A31 | 53 < 1 > A23 A93 \implies A71 A102 A114 |
| 16 < 1 > A12 A71 \implies A24 A61 A92 | 35 < 1 > A71 A111 \implies A24 A52 A101 | 54 < 1 > A71 A93 \implies A23 A102 A114 |
| 17 < 1 > A61 A71 \implies A12 A24 A92 | 36 < 1 > A12 A113 \implies A33 A92 A102 | 55 < 1 > A23 A102 \implies A71 A93 A114 |
| 18 < 1 > A81 \implies A22 A34 | 37 < 1 > A13 A113 \implies A23 A83 | 56 < 1 > A71 A102 \implies A23 A93 A114 |
| 19 < 1 > A82 \implies A31 A61 A93 | 38 < 1 > A23 A113 \implies A13 A83 | 57 < 1 > A93 A102 \implies A23 A71 A114 |

Table 15. Implication relation table for decision class coronary heart disease

| Superconcept | a_{11} | a_{13} | a_{14} | a_{22} | a_{23} | a_{24} | a_{31} | a_{33} | a_{34} | a_{72} | a_{73} | a_{74} | | | |
|--------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Subconcept | a_{11} $a_{22}, a_{11,3}$ $a_{10,2}, a_{11,3}$ $a_{22}, a_{11,4}$ a_{22}, a_{29} a_{34}, a_{41} a_{41}, a_{71} a_{34}, a_{29} a_{71}, a_{29} $a_{29}, a_{10,2}$ | a_{13}, a_{41} a_{23} a_{13}, a_{29} $a_{13}, a_{11,3}$ | $a_{14}, a_{11,4}$ $a_{34}, a_{11,4}$ | $a_{22}, a_{11,4}$ a_{13}, a_{29} a_{21} | $a_{23}, a_{10,2}$ a_{21}, a_{29} $a_{21}, a_{11,4}$ $a_{10,2}, a_{11,4}$ a_{23} $a_{13}, a_{11,3}$ | a_{24} a_{34}, a_{13} a_{13}, a_{41} a_{13}, a_{71} a_{41}, a_{71} a_{21}, a_{29} $a_{10,1}$ $a_{13}, a_{10,2}$ $a_{23}, a_{10,2}$ $a_{13}, a_{11,4}$ a_{14} | a_{31}, a_{11} | $a_{33}, a_{11,3}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ $a_{13}, a_{11,3}$ | $a_{34}, a_{11,4}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{72}, a_{11} $a_{13}, a_{11,3}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{73}, a_{41} a_{13}, a_{29} $a_{13}, a_{10,2}$ $a_{13}, a_{11,3}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{74} a_{13}, a_{29} $a_{13}, a_{10,2}$ $a_{13}, a_{11,3}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | $a_{11,3}$ $a_{11,4}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | $a_{11,3}$ $a_{11,4}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | $a_{11,4}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ |
| Frequency | 19 | 7 | 4 | 5 | 13 | 21 | 2 | 10 | 10 | 5 | 10 | 18 | | | |
| Superconcept | a_{71} | a_{21} | a_{22} | a_{23} | a_{24} | a_{29} | a_{31} | a_{32} | a_{33} | a_{34} | a_{72} | a_{73} | | | |
| Subconcept | a_{13}, a_{24} a_{13} a_{13}, a_{41} a_{34}, a_{29} $a_{10,1}$ $a_{34}, a_{11,3}$ $a_{23}, a_{11,4}$ $a_{10,2}, a_{11,4}$ a_{23}, a_{29} $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{21}, a_{34} a_{21}, a_{29} | a_{22}, a_{41} a_{21}, a_{29} | a_{13}, a_{23} a_{13}, a_{13} a_{23}, a_{13} | a_{13}, a_{24} a_{13}, a_{23} a_{13}, a_{41} $a_{13}, a_{10,2}$ $a_{13}, a_{11,3}$ $a_{10,2}, a_{11,3}$ a_{23} | a_{41} a_{31}, a_{41} $a_{22}, a_{11,4}$ $a_{23}, a_{11,4}$ $a_{10,2}, a_{11,3}$ $a_{23}, a_{11,4}$ $a_{10,2}, a_{11,4}$ | a_{29} a_{31}, a_{41} $a_{22}, a_{11,4}$ $a_{23}, a_{11,4}$ $a_{10,2}, a_{11,3}$ $a_{23}, a_{11,4}$ $a_{10,2}, a_{11,4}$ a_{13}, a_{22} a_{23}, a_{71} $a_{23}, a_{10,2}$ $a_{21}, a_{10,2}$ | a_{31}, a_{11} $a_{13}, a_{11,3}$ $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{32}, a_{11} $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{33} a_{31} a_{13}, a_{23} a_{31} a_{13}, a_{23} $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{34} a_{23} a_{13}, a_{23} a_{31} a_{13}, a_{23} $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{72} $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{73} $a_{13}, a_{10,2}$ a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | a_{74} a_{13}, a_{22} a_{13}, a_{29} a_{13}, a_{71} $a_{29}, a_{10,2}$ | |
| Frequency | 22 | 2 | 4 | 6 | 15 | 22 | 5 | 15 | 3 | 8 | 19 | | | | |

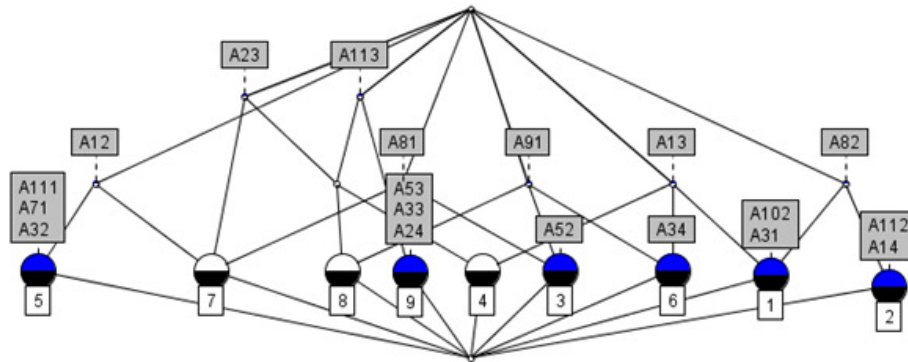


Figure 7 Lattice diagram of the decision class potential patient

From the higher frequency of the implication Table 21 it is observed that the age group old is the potential patients affecting the heart disease. The chief characteristics influencing the decision class is blood pressure-high and the next chief characteristic is chest pain-atypical angina.

Table 20. Implication set table for decision class potential patient

| Implication Sets | | |
|--------------------------------------|---------------------------------------|---------------------------------------|
| 1 < 1 > A14 \implies A82 A112; | 9 < 1 > A52 \implies A81 A91; | 17 < 1 > A81 A91 \implies A52; |
| 2 < 1 > A12 A23 \implies A81; | 10 < 1 > A53 \implies A24 A33 A113; | 18 < 1 > A102 \implies A13 A31 A82; |
| 3 < 1 > A13 A23 \implies A113; | 11 < 1 > A71 \implies A12 A32 A111; | 19 < 1 > A111 \implies A12 A32 A71; |
| 4 < 1 > A24 \implies A33 A53 A113; | 12 < 1 > A12 A81 \implies A23; | 20 < 1 > A112 \implies A14 A82; |
| 5 < 1 > A31 \implies A13 A82 A102; | 13 < 1 > A23 A81 \implies A12; | 21 < 1 > A13 A113 \implies A23; |
| 6 < 1 > A32 \implies A12 A71 A111; | 14 < 1 > A13 A82 \implies A31 A102; | 22 < 1 > A91 A113 \implies A23; |
| 7 < 1 > A33 \implies A24 A53 A113; | 15 < 1 > A13 A91 \implies A34; | |
| 8 < 1 > A34 \implies A13 A91; | 16 < 1 > A23 A91 \implies A113; | |

Table 21. Implication relation table for decision class potential patient

| Superconcept | a_{12} | a_{13} | a_{14} | a_{23} | a_{24} | a_{31} | a_{32} | a_{33} | a_{34} | a_{52} |
|--------------|--|------------------------------------|------------------------------|--|----------------------|--------------------------------|------------------------|----------------------|--|------------------|
| Subconcept | a_{23}, a_{81} $a_{11,1}$ a_{71} a_{32} | a_{31} a_{34} $a_{10,2}$ | $a_{11,2}$ | a_{12}, a_{81} $a_{13}, a_{11,3}$ $a_{91}, a_{11,3}$ | a_{33} a_{53} | a_{13}, a_{82} $a_{10,2}$ | a_{71} $a_{11,1}$ | a_{24} a_{53} | a_{13}, a_{91} | a_{81}, a_{91} |
| Frequency | 5 | 3 | 1 | 6 | 2 | 3 | 2 | 2 | 2 | 2 |
| Superconcept | a_{53} | a_{71} | a_{81} | a_{82} | a_{91} | $a_{10,2}$ | $a_{11,1}$ | $a_{11,2}$ | $a_{11,3}$ | |
| Subconcept | a_{24} a_{33} | a_{32} $a_{11,1}$ | a_{12}, a_{23} a_{52} | a_{14} a_{31} $a_{10,2}$ $a_{11,2}$ | a_{34} a_{52} | a_{31} a_{13}, a_{82} | a_{32} a_{71} | a_{14} | a_{24}, a_{33} a_{53} a_{13}, a_{23} a_{23}, a_{91} | |
| Frequency | 2 | 2 | 3 | 4 | 2 | 3 | 2 | 1 | 7 | |

5.2.5 Decision Class Cardiomyopathy

The number of rules generated using rule generation algorithm due to rough set theory is 17 and is further minimized to 15 by using domain intelligence. This is further validated and minimized to 14 numbers of rules using testing data set. Now we present the context table in Table 22,

Table 24. Implication relation table for decision class cardiomyopathy

| Superconcept | a_{12} | a_{13} | a_{14} | a_{22} | a_{23} | a_{24} | a_{31} | a_{32} | a_{33} | a_{34} | a_{41} | a_{52} | a_{53} |
|--------------|---|--|--|--|---|--|--|--|--------------------------------------|--|--|--|---|
| Subconcept | a_{81}, a_{53} a_{63}, a_{53} $a_{71}, a_{10,1}$ | a_{54} | a_{31} $a_{22}, a_{11,2}$ $a_{83}, a_{11,2}$ $a_{83}, a_{11,3}$ | a_{58}, a_{52} $a_{83}, a_{11,1}$ $a_{83}, a_{11,2}$ $a_{11,4}$ | $a_{91}, a_{1,3}$ $a_{52}, a_{11,2}$ $a_{10,2}, a_{61}$ | $a_{53},$ $a_{11,3},$ a_{33} | $a_{83},$ $a_{11,3}$ | $a_{63}, a_{11,3}$ $a_{10,1}, a_{11,3}$ | $a_{24},$ $a_{53},$ $a_{11,3}$ | a_{13} | a_{52}, a_{93} a_{52}, a_{63} a_{63}, a_{93} | $a_{10,2}$ $a_{23}, a_{11,2}$ a_{63}, a_{93} | $a_{71} \# 2$ a_{24}, a_{53} $a_{11,4}$ $a_{22}, a_{10,1}$ $a_{12}, a_{10,1}$ a_{12}, a_{63} |
| Frequency | 6 | 1 | 7 | 7 | 6 | 3 | 2 | 4 | 3 | 1 | 6 | 5 | 11 |
| Superconcept | a_{63} | a_{71} | a_{82} | a_{83} | a_{91} | a_{93} | $a_{10,1}$ | $a_{10,2}$ | $a_{11,1}$ | $a_{11,2}$ | $a_{11,3}$ | $a_{11,4}$ | |
| Subconcept | $a_{41}, a_{12},$ $a_{53}, a_{52},$ $a_{93}, a_{12},$ $a_{10,1}, a_{52},$ $a_{10,1}, a_{52},$ $a_{11,3}, a_{10,1},$ $a_{11,3}, a_{53},$ $a_{71}, a_{10,1}$ | a_{12}, a_{53} a_{12}, a_{63} $a_{12}, a_{10,1}$ a_{53}, a_{63} a_{53}, a_{91} | $a_{11,1}$ a_{22}, a_{93} | $a_{14} \# 2$ a_{31} a_{22}, a_{52} | $a_{23}, a_{11,3}$ a_{34} | a_{52}, a_{63} $a_{82}, a_{11,1}$ | a_{12}, a_{53} a_{12}, a_{63} a_{32}, a_{63} a_{53}, a_{63} $a_{32}, a_{11,3}$ $a_{63}, a_{11,3}$ $a_{11,4}$ a_{22}, a_{53} | $a_{23}, a_{11,2}$ $a_{52}, a_{11,2}$ | a_{82} a_{22}, a_{93} | $a_{10,2}, a_{23}$ a_{52}, a_{14} a_{22}, a_{83} | $a_{23},$ a_{91} | $a_{22},$ $a_{53},$ $a_{22},$ $a_{10,1}$ | |
| Frequency | 16 | 10 | 3 | 5 | 3 | 4 | 15 | 4 | 3 | 6 | 2 | 4 | |

6. Conclusion

In this study, the rule generation algorithm of rough set theory generates 91 rules. This is further minimized to 72 candidate rules with the help of domain intelligence and is further minimized to 65 rules by validation process and threshold value. Further these suitable rules are explored to identify the chief characteristics affecting the relationship between heart disease and its attributes by using formal concept analysis. This helps the decision maker a priori detection of the heart disease. The chief characteristics of hypertensive heart disease are old peak (a_{83}), exercise (a_{71}); the other characteristics include chest pain (a_{12}), blood pressure (a_{23}), and maximum heart rate (a_{63}). The chief characteristics of coronary heart disease are thallium scan (a_{93}), exercise (a_{71}); the other characteristics include chest pain (a_{12}), and blood pressure (a_{24}). The chief characteristics of heart failure are chest pain (a_{14}), exercise (a_{72}), and old peak (a_{83}); the other characteristics include chest pain (a_{13}), and thallium scan (a_{92}). The chief characteristics of patient that may lead to heart disease are blood pressure (a_{23}), and chest pain (a_{12}), where the targeted age group is old. Finally, the chief characteristics of cardiomyopathy are maximum heart rate (a_{63}), and ECG (a_{53}), where the targeted group in general is male. We believe that, formal concept analysis can be used to find furthermore information regardless of the type of rule based soft computing. We also believe that the proposed model is a useful method for decision makers.

REFERENCES

[1] Zadeh, L. A., (1965) "Fuzzy Sets", *Information and Control*, Vol. 8, pp338-353.
 [2] Pawlak, Z, (1982) "Rough Sets", *International Journal of Computer and Information Sciences*, Vol. 11, pp341-356.
 [3] Pawlak, Z. & Skowron, A., (2007) "Rudiments of rough sets", *Information Sciences, An International Journal, Elsevier*, Vol. 177, No. 1, pp 3–27.
 [4] Pawlak, Z. & Skowron, A., (2007) "Rough sets: Some extensions", *Information Sciences, An International Journal, Elsevier*, Vol. 177, No. 1, pp 28–40.

- [5] Pawlak, Z. & Skowron, A., (2007) "Rough sets and Boolean reasoning", *Information Sciences, An International Journal, Elsevier*, Vol. 177, No. 1, pp 41–73.
- [6] Slowinski, R. & Vanderpooten, D., (2000) "A Generalized Definition of Rough Approximations Based on Similarity", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 12, No. 2, pp331-336.
- [7] Dubois, D. & Prade, H., (1990) "Rough fuzzy sets and fuzzy rough sets", *International Journal of General System*, Vol. 17, pp191-209.
- [8] Acharjya, D.P. & Tripathy, B.K, (2008) "Rough Sets on Fuzzy Approximation Spaces and Applications to Distributed Knowledge Systems", *International Journal of Artificial Intelligence and Soft Computing*, Vol. 1, No. 1, pp1-14.
- [9] Li, Dong Ya., & Hu, Bao Qing, (2007) "A kind of Dynamic Rough Sets", *Fourth International Conference on Fuzzy Systems and Knowledge Discovery(FSKD)*, pp79-85.
- [10] Acharjya, D.P. & Tripathy, B.K, (2009) "Rough Sets on Intuitionistic Fuzzy Approximation Spaces and Knowledge Representation", *International Journal of Artificial Intelligence and Computational Research*, Vol. 1, No. 1, pp29-36.
- [11] Zhu, W & Wang, F.Y., (2007) "On three types of covering rough sets", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, No. 8, pp1131-1144.
- [12] Gupta, Anamika, Kumar Naveen & Bhatnagar Vasudha, (2005) "Analysis of Medical Data using Data Mining and Formal Concept Analysis", *World Academy of Science, Engineering and Technology*, Vol. 11, pp61-64.
- [13] Adeli, A. & Neshat, Mehdi, (2010) "A Fuzzy Expert System for Heart Disease Diagnosis", *Proceedings of International Multi Conference of Engineers and Computer Scientists, Hong Kong*, Vol. 1, pp134 -139.
- [14] Pawlak, Z, (1991) *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, Netherlands.
- [15] Pawlak, Z., (1981) "Information systems: theoretical foundations", *Information Systems*, Vol. 6, pp205-218.
- [16] Saleem Durai, M. A., Acharjya, D. P., Kannan, A. & Iyengar, N. Ch. S. N., (Submitted) "An intelligent knowledge mining model for kidney cancer using rough set theory", *International Journal of Bioinformatics Research and Applications*.
- [17] Kumawat Madan Lal & Acharjya, D. P., (2010) "Knowledge Granulation, Association Rules and Granular Computing", *Proceedings of Second National Conference on Advanced Technologies in Electrical Engineering, Virudhunagar, Tamilnadu, India*, pp43-46.
- [18] Pawlak, Z., (1984) "Rough Classification", *International Journal of Man-Machine Studies*, Vol. 20, No. 5, pp469-483.
- [19] Shyng, J. Y., Tzeng, G. H. and Wang, F. K., (2007) "Rough set theory in analyzing the attributes of combination values for insurance market", *Expert System with Applications*, Vol. 32, No. 1, pp. 56-64.
- [20] Walczak, B. & Massart, D. L., (1999) "Rough set theory", *Chemometrics and Intelligent Laboratory*, Vol. 47, No. 1, pp1-16.
- [21] Pawlak, Z., (2002) "Rough sets and intelligent data analysis", *Information Sciences*, Vol. 147, pp1-12.
- [22] Wille, R., (2005) "Formal Concept Analysis as Mathematical theory of concept and concept Hierarchies", *Lecture Notes in Artificial Intelligence*, Vol. 3626, pp1-33.
- [23] Wagner, H., (1973) "Begriff", In: H. Krings, H. M. Baumgartner, C. Wild (eds.): *Handbuch philosophischer Grundbegriffe, Kosel, Munchen*, pp191-209.
- [24] Wormuth, B. & Becker, P., (2004) "Introduction to format concept analysis", In 2nd International conference of Formal Concept Analysis, February 23-February 27, Sydney, Australia.

Authors



B K Tripathy, a senior professor in the school of computing sciences and engineering, VIT University, at Vellore, India. He has been awarded with Gold Medals both at graduate and post graduate levels of Berhampur University, India. Also, he has been awarded with the best post graduate of the Berhampur University. He has received national scholarship, UGC fellowship, SERC visiting fellowship and DOE (Govt. of India) scholarship at various levels of his career. He has published more than hundred technical papers in various international journals, conferences, and Springer book chapters. He has produced ten Ph. Ds under his supervision. He is associated with many professional bodies like IEEE, IRSS, WSEAS, AISTC, ISTP, CSI, AMS, and IMS. His name also appeared in the editorial board of several international journals like CTA, ITTA, AMMS, IJCTE, AISS, AIT, and IJPS. Also, he is a reviewer of international journals like Mathematical Reviews, Information Sciences, Analysis of Neural Networks, Journal of Knowledge Engineering, Mathematical Communications, IJET, IJPR and Journal of Analysis. His name has come out in Marquis who's who and in International Biographical Centre, London. His research interest includes fuzzy sets and systems, rough sets and knowledge engineering, data clustering, social network analysis, soft computing and granular computing.



D P Acharjya received the M. Tech. degree in computer science from Utkal University, India in 2002; M. Phil. from Berhampur University, India; and M. Sc. from NIT, Rourkela, India. He has been awarded with Gold Medal in M. Sc. He is a research student of Berhampur University, India. Currently, he is an Associate Professor in the school of computing sciences and engineering, VIT University, Vellore, India. His 15 years of experience in higher education include as a faculty member, reviewer, academic councillor, academic guide and member of board of studies. He has authored many national and international journal papers and three books entitled Fundamental Approach to Discrete Mathematics, 2nd Ed, New-Age International

Pvt. Ltd, New Delhi, India (2009); Computer Based on Mathematics, Laxmi Publications, New Delhi, India (2007); and Theory of Computation, MJP Publishers, Chennai, India (2010) to his credit. He is associated with many professional bodies CSI, ISTE, IMS, AMTI, ISIAM, OITS, IACSIT, CSTA, IEEE and IAENG. He was founder secretary of OITS Rourkela chapter. His current research interests include rough sets, formal concept analysis, knowledge representation, granular computing, data mining and business intelligence.



V. Cynthya, received the B. E. in Computer Science from Karunya University, Coimbatore, India in 2009. She is a M. Tech. (CSE) final year student of VIT University, Vellore, India. She has keen interest in teaching and applied research. Her research interests include rough sets, granular computing, formal concepts and knowledge mining.