

Rough Set Approach for Generation of Classification Rules for Dengue

Sujogya Mishra

Research scholar, Utkal
University Bhubaneswar-751004,
India

Shakti Prasad Mohanty

Department of Mathematics,
College of Engineering and
Technology, Bhubaneswar-
751003, India

Radhanath Hota

Department of Computer Science,
College of Basic Science and
Humanities, OUAT,
Bhubaneswar-751003, India

J. Chandrakanta Badajena

Department of IT,
College of Engineering and Technology
Bhubaneswar-751003, India

ABSTRACT

In recent age medical science has improved to a significant height but lots of common known diseases need maximum number of medical tests that are not only expensive and but also inaccurate in proper diagnosis of the diseases in our paper we emphasized more on symptom rather than medical test. From a large domain of diseases we consider dengue fever. Every year millions of people died from dengue due to lack of proper diagnosis. We develop an algorithm using rough set concept, which emphasized more on symptom rather than medical test we divide the entire paper into three basic sections: 1st section about literature review, the 2nd section about the experiment on the data that are collected from different medical sources and the 3rd section is about the validation, conclusion and future work. The data collected from DR Pradeep Kumar Mishra MD

Keywords: *Rough Set Theory, Medical related data, Granular computing, Data mining.*

1. INTRODUCTION

The growth of the size of data and number of existing databases far exceeds the ability of human's brain to analyze these structures which creates both a need and an opportunity to extract knowledge from databases [1]. Medical databases have accumulated large quantities of information about patients and their medical conditions, relationships and patterns among these data could provide new medical knowledge. Analysis of medical data is often concerned with treatment of incomplete knowledge, with management of inconsistent pieces of information and with manipulation of various levels of representation of data. Existing intelligent techniques [2] of data analysis are mainly based on quite strong assumptions: knowledge about dependencies, probability distributions and large number of experiments are unable to derive conclusions from incomplete knowledge, or cannot manage inconsistent pieces of information. The standard intelligent techniques used in medical data analysis are neural network [3], Bayesian classifier [4], genetic algorithms [5], decision trees [6], fuzzy set [7]. Rough set theory and the basic concept was invented by Polish logician, Professor Z. Pawlak in early eighties [8]. The theory of rough sets is a mathematical tool for extracting knowledge from un-certain and incomplete data base

information. The theory assumes that we first have necessary information or knowledge of all the objects in the universe with which the objects can be divided into different groups. If we have exactly the same information of two objects then we say that they are indiscernible (similar), with known knowledge. The theory of Rough Set can be used to find dependence relationship among data, evaluate the importance of attributes, discover the patterns of data, learn common decision-making rules, reduce all redundant objects and attributes and seek the minimum subset of attributes so as to attain satisfying classification. More-over, the rough set reduction algorithms enable to approximate the decision classes using possibly large and simplified patterns [9]. This theory became very popular among scientists around the world and the rough set is now one of the most developed intelligent data analysis. Unlike other intelligent methods such as fuzzy set theory, Dempster-Shafer theory or statistical methods, rough set analysis requires no external parameters and uses only the information presented in the given data [10]. This paper discusses how rough set theory can be used to analyze medical data, and for generating classification rules from a set of observed samples of the medical data in this context we gathered all symptoms of dengue then apply the rough concept to generate rules for dengue.

2. PRILIMINARIES

2.1 Rough set Rough set theory as introduced by Z. Pawlak [8] is an extension of conventional set theory that support approximations in decision making.

2.1.2 Approximation Space: An Approximation space is a pair (U, R) where U is a non empty finite set called the universe R is an equivalence relation defined on U .

2.1.3 Information System: An information system is a pair $S = (U, A)$, where U is then non-empty finite set called the universe, A is the non-empty finite set of attributes

2.1.4 Decision Table: A decision table is a special case of information systems $S = (U, A = C \cup \{d\})$, where d is not in C . Attributes in C are called conditional attributes and d is a designated attribute called the decision attribute

2.1.5 Approximations of Sets: Let $S = (U, R)$ be an approximation space and X be a subset of U . The lower approximation of X by R in S is defined as $\underline{RX} = \{ e \in U \mid [e] \subseteq X \}$ and The upper approximation of X by R in S is defined as $\overline{RX} = \{ e \in U \mid [e] \cap X \neq \emptyset \}$ where $[e]$ denotes the equivalence class containing e . A subset X of U is said to be R -definable in S if and only if $\overline{RX} = \underline{RX}$. A set X is rough in S if its boundary set is nonempty.

2.2 Dependency of Attributes

Let C and D be subsets of A . We say that D depends on C in a degree k ($0 \leq k \leq 1$) denoted by $C \rightarrow_k D$ if $K = \gamma(C, D) =$

$$\frac{|POS_C(D)|}{|U|} \text{ where } POS_C(D) = \bigcup_{x \in U} \underline{C}(x), \text{ is called positive}$$

region of the partition U/D with respect to C where $x \in u/d$, which is all elements of U that can be uniquely classified to the block of partition U/D . If $k = 1$ we say that D depends totally on C . If $k < 1$ we say that D depends partially (in a degree k) on C .

2.3 Dispensable and Indispensable Attributes-

Let $S = (U, A = C \vee D)$ be a decision table. Let c be an attribute in C . Attribute c is dispensable in S if $POS_C(D) = POS_{(C-\{c\})}(D)$ otherwise, c is indispensable. A decision table S is independent if all attributes in C are indispensable. Let $S = (U, A = C \vee D)$ be a decision table.

Rough Set Attribute Reduction (RSAR) provides a filter based tool by which knowledge may be extracted from a domain in a concise way; retaining the information content whilst reducing the amount of knowledge involved.

2.4 Reduct and Core Let $S = (U, A=C \cup D)$ be a decision table. A subset R of C is a reduct of C , if $POS_R(D) = POS_C(D)$ and $S' = (U, R \cup D)$ is independent, i.e., all attributes in R are indispensable in S' . Core of C is the set of attributes shared by all reducts of C . $CORE(C) = \bigcap RED(C)$ where, $RED(C)$ is the set of all reducts of C . The reduct is often used in the attribute selection process to eliminate redundant attributes towards decision making.

2.5 Correlation- Correlation define as a mutual relationship or connection between two or more things. The quantity r , called the *linear correlation coefficient*, measures the strength and the direction of a linear relationship between two variables. The linear correlation coefficient is sometimes referred to as the *Pearson product moment correlation coefficient* in honor of its developer Karl Pearson. The mathematical formula for its coefficient given by the formula

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

2.6 Goodness of fit-The **goodness of fit** of a statistical model describes how well it fits a set of observations. Measures of **goodness of fit** typically summarize the discrepancy between observed values and the values expected under the model in question.

2.7 Chi squared distribution- A **chi-squared test**, also referred to as χ^2 test, is any statistical hypothesis test in

which the sampling distribution of the test statistic is a chi squared distribution when the null hypothesis is true. Also considered a chi-squared test is a test in which this is *asymptotically* true, meaning that the sampling distribution (if the null hypothesis is true) can be made to approximate a chi-squared distribution as closely as desired by making the sample size large enough. The chi-square (I) test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories. Do the numbers of individuals or objects that fall in each category differ significantly from the number you would expect? Is this difference between the expected and observed due to sampling variation, or is it a real difference

2.8 Further analysis of chi square test- Basic properties of chi squared goodness fit is that it is non symmetric in nature. However if the degrees of freedom increased it appears to be to be more symmetrical. It is right tailed one sided test. All expectation in chi squared test is greater than 1. $E_i = np_i$ where n is the number samples considered p_i is the probability of i^{th} occurrence. Data selected at random there are two hypothesis null hypothesis and alternate hypothesis null hypothesis denoted by H_0 alternate hypothesis denoted by H_1 . H_0 is the claim does follow the hypothesis and H_1 is the claim does not follow the hypothesis here H_1 is called the alternate hypothesis to H_0 . If the test value found out to be K then K can be calculated by the formula $K = \sum (O_i - E_i)^2 / E_i$. Choice of significance level always satisfies type 1 error.

2.9 Different types of error-

- 1) Type 1 error-Rejecting a hypothesis even though it is true
- 2) Type 2 error-Accepting the hypothesis when it is false
- 3) Type 3 error-Rejecting a hypothesis correctly for wrong reason

3. BASIC IDEA

The basic idea for the proposed work is conceived from the general medical system. We initially consider 1000 samples, of Dengue cases and applying correlation technique the number reduced to 20. To start with and to apply rough concept on our gathered data find five symptom which are consider to be the very basic symptom of dengue we consider these attribute as our conditional attribute these are as follows such as Very high Fever, Headache, Muscle and joint pain, Pain behind your eye and Nausea and two decision attribute positive and negative. Values for these attributes are Low, Moderate and High consider as The Dengue cases are collected from Doctor Pradeep Kumar Mishra MD

4. DATA REDUCTION

As the volume of data is increases with time it is difficult to know which data are useful and responsible for a particular application. The basic objective of data reduction is to find the relevant attributes that have all essential information of the data set. The process is illustrated by applying rough set concept on 20 dissimilar samples which we had gathered by correlation techniques. In this particular problem we consider five conditional attributes as Headache, Very high Fever, Muscle and joint pain, Pain behind your eye and Nausea and its values are defined as low, moderate and high decision attributes are positive, negative. For better understanding and good clarity of the paper we rename the conditional attributes, value of the conditional attributes and the decision attribute as

$(a_1, a_2, a_3, a_4, a_5), (b_1, b_2, b_3), (c_1, c_2)$ respectively . Application and analysis on the data set and rule generation being presented in the following tables . Table -1 is the initial table, and the process of analysis is present in the subsequent tables

Table-1:

E	a ₁	a ₂	a ₃	a ₄	a ₅	d
E ₁	b ₂	b ₂	b ₁	b ₁	b ₁	c ₂
E ₂	b ₂	b ₂	b ₁	b ₃	b ₃	c ₂
E ₃	b ₁	b ₂	b ₂	b ₃	b ₃	c ₁
E ₄	b ₁	b ₂	b ₂	b ₃	b ₃	c ₁
E ₅	b ₃	b ₃	b ₃	b ₃	b ₂	c ₂
E ₆	b ₁	b ₂	b ₂	b ₂	b ₂	c ₂
E ₇	b ₂	b ₂	b ₂	b ₂	b ₂	c ₁
E ₈	b ₁	b ₁	b ₁	b ₁	b ₁	c ₂
E ₉	b ₁	b ₂	b ₂	b ₃	b ₃	c ₁
E ₁₀	b ₁	b ₂	b ₂	b ₂	b ₂	c ₂
E ₁₁	b ₂	b ₃	b ₃	b ₃	b ₃	c ₁
E ₁₂	b ₁	b ₂	b ₃	b ₁	b ₂	c ₁
E ₁₃	b ₃	b ₂	b ₂	b ₂	b ₁	c ₂
E ₁₄	b ₃	b ₃	b ₃	b ₃	b ₃	c ₁
E ₁₅	b ₂	b ₁	b ₁	b ₁	b ₁	c ₂
E ₁₆	b ₁	b ₁	b ₁	b ₁	b ₁	c ₂
E ₁₇	b ₁	b ₃	b ₂	b ₂	b ₃	c ₁
E ₁₈	b ₁	b ₂	b ₂	b ₃	b ₂	c ₂
E ₁₉	b ₁	b ₃	b ₁	b ₃	b ₃	c ₂
E ₂₀	b ₁	b ₃	b ₁	b ₃	b ₃	c ₁

The decision table -1 , takes the initial values before finding the reduct looking at the data table it is found that entities E₁₉, E₂₀, ambiguous in nature and E₃, E₄ gives same result so both E₁₉, E₂₀ drop from the table and E₃ , E₄ gives the same result so E₃ , E₄ records we only keep only one record in the table-2. From table -1 we get the next table-2

Reduced Table-2 from table-1

E	a ₁	a ₂	a ₃	a ₄	a ₅	D
E ₁	b ₂	b ₂	b ₁	b ₁	b ₁	c ₂
E ₂	b ₂	b ₂	b ₁	b ₃	b ₃	c ₂
E ₄	b ₁	b ₂	b ₂	b ₃	b ₃	c ₁
E ₅	b ₃	b ₃	b ₃	b ₃	b ₂	c ₁

E ₆	b ₁	b ₂	b ₂	b ₂	b ₂	c ₂
E ₇	b ₂	b ₂	b ₂	b ₂	b ₂	c ₁
E ₈	b ₁	b ₁	b ₁	b ₁	b ₁	c ₂
E ₉	b ₁	b ₂	b ₂	b ₃	b ₃	c ₁
E ₁₀	b ₁	b ₂	b ₂	b ₂	b ₂	c ₂
E ₁₁	b ₂	b ₃	b ₃	b ₃	b ₃	c ₁
E ₁₂	b ₁	b ₂	b ₃	b ₁	b ₂	c ₁
E ₁₃	b ₃	b ₂	b ₂	b ₂	b ₁	c ₂
E ₁₄	b ₃	b ₃	b ₃	b ₃	b ₃	c ₁
E ₁₅	b ₂	b ₁	b ₁	b ₁	b ₁	c ₂
E ₁₆	b ₁	b ₁	b ₁	b ₁	b ₁	c ₂
E ₁₇	b ₁	b ₃	b ₂	b ₂	b ₃	c ₁
E ₁₈	b ₁	b ₂	b ₂	b ₃	b ₂	c ₂

Indiscernibility relation:

Indiscernibility Relation is the relation between two or more objects where all the values are identical in relation to a subset of considered attributes.

Approximation:

The starting point of rough set theory is the indiscernibility relation, generated by information concerning objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge it is unable to discern some objects employing the available information Approximations is also other an important concept in Rough Sets Theory, being associated with the meaning of the approximations topological operations (Wu et al., 2004). The lower and the upper approximations of a set are interior and closure operations in a topology generated by the indiscernibility relation. Below is presented and described the types of approximations that are used in Rough Sets Theory.

a. Lower Approximation :

Lower Approximation is a description of the domain objects that are known with certainty to belong to the subset of interest. The Lower Approximation Set of a set X, with regard to R is the set of all objects, which can be classified with X regarding R, that is denoted as R_L .

b. Upper Approximation :

Upper Approximation is a description of the objects that possibly belong to the subset of interest. The Upper Approximation Set of a set X regarding R is the set of all of objects which can be possibly classified with X regarding R. Denoted as R_U

c. Boundary Region (BR) :

Boundary Region is description of the objects that of a set X regarding R is the set of all the objects, which cannot be classified neither as X nor -X regarding R. If the boundary region $X = \emptyset$ then the set is considered "Crisp", that is, exact in relation to R; otherwise, if the boundary region is a set $X \neq \emptyset$ the set X "Rough" is considered. In that the boundary region is $BR = R_U - R_L$.

The lower and the upper approximations of a set are interior and closure operations in a topology generated by a indiscernibility relation. In discernibility according to decision attributes in this case has divided in to two groups one group consist of positive case and another group consists of negative cases

$$E_{\text{positive}} = \{E_4, E_5, E_7, E_9, E_{11}, E_{12}, E_{14}\} \dots\dots\dots(1)$$

$$E_{\text{negative}} = \{E_1, E_2, E_6, E_8, E_{10}, E_{13}, E_{15}, E_{16}, E_{18}\} \dots\dots\dots(2)$$

$$E(a_1)_{\text{low}} = \{E_4, E_6, E_8, E_9, E_{10}, E_{12}, E_{16}, E_{17}, E_{18}\} \dots\dots\dots(3)$$

$$E(a_1)_{\text{moderate}} = \{E_1, E_2, E_7, E_{11}, E_{15}\} \dots\dots\dots(4)$$

$$E(a_1)_{\text{high}} = \{E_5, E_{13}, E_{14}\} \dots\dots\dots(5)$$

The above result when compared with the positive cases $E(a_1)_{\text{high}}$ strength[11]

Found to be nil where as for negative cases of high $E(a_1)$ strength[11] 3/3 cent percent similarly for negative cases of moderate $E(a_1)$ strength[11] gives rise to be 1/5 about 20% , positive cases of low $E(a_1)$ strength[11] 6/9 about 66% basic observation gives rise is that a_1 attribute does not give any significant result similarly

$$E(a_2)_{\text{low}} = \{E_8, E_{15}, E_{16}\} \dots\dots\dots(6)$$

$$E(a_2)_{\text{moderate}} = \{E_1, E_2, E_4, E_6, E_7, E_9, E_{10}, E_{12}, E_{13}, E_{18}\} \dots\dots\dots(7)$$

$$E(a_2)_{\text{high}} = \{E_5, E_{13}, E_{14}\} \dots\dots\dots(8)$$

Similar analysis strength[11] positive high a_2 will be 4/4=1 about cent percent

And for strength[11] negative for low a_2 will be also 3/3=1 about cent percent

That is we have a conclusion that if a person having high a_2 symptom shows the sign of malaria and in contrast a low symptom of a_2 shows a negative case

That is why we are not analyzing the moderate a_2 cases now similarly for a_3

$$E(a_3)_{\text{low}} = \{E_1, E_2, E_8, E_{15}, E_{16}\} \dots\dots\dots(9)$$

$$E(a_3)_{\text{moderate}} = \{E_4, E_6, E_7, E_9, E_{10}, E_{13}, E_{17}, E_{18}\} \dots\dots\dots(10)$$

$$E(a_3)_{\text{high}} = \{E_5, E_{11}, E_{12}, E_{14}\} \dots\dots\dots(11)$$

Finding the strength [11] high a_3 will be 4/4=1 that is about cent percent

And similarly for strength [11] for low a_3 cases will be gives negative result will be of 4/5 about 80% negative cases for low a_3 cases so we are not consider the moderate a_3 case now similarly for a_4 cases we consider the

$$E(a_4)_{\text{low}} = \{E_1, E_8, E_{12}, E_{15}, E_{16}\} \dots\dots\dots(12)$$

$$E(a_4)_{\text{moderate}} = \{E_6, E_7, E_{10}, E_{13}, E_{17}\} \dots\dots\dots(13)$$

$$E(a_4)_{\text{high}} = \{E_2, E_4, E_5, E_9, E_{11}, E_{14}, E_{18}\} \dots\dots\dots(14)$$

Analyzing a_4 $E(a_4)_{\text{low}}$ negative strength[11] will be 4/5 that is

about 80% similarly for $E(a_4)_{\text{high}}$ positive strength[11] cases will be about 5/7 about 70% now considering a_5

$$E(a_5)_{\text{low}} = \{E_1, E_8, E_{13}, E_{15}, E_{16}\} \dots\dots\dots(15)$$

$$E(a_5)_{\text{moderate}} = \{E_5, E_6, E_7, E_{10}, E_{12}, E_{18}\} \dots\dots\dots(16)$$

$$E(a_5)_{\text{high}} = \{E_2, E_4, E_9, E_{11}, E_{14}, E_{17}\} \dots\dots\dots(17)$$

$E(a_5)_{\text{low}}$ strength[11] for negative case will be 4/5 about 80% and $E(a_5)_{\text{high}}$ strength for positive 4/6 about 66% positive strength[11] moderate case given by 3/8 about 37% so after analyzing the above data by strength view point we ignore attribute a_1 and a_5 as in a_1 strength [11] for low negative case is 100% high positive strength[11] a_1 about 20% similar argument in case of a_5 it's strength[11] among a_2, a_3, a_4 it is the minimum one secondly the upper approximation positive case of a_5 in both moderate and high case will be $\{E_5, E_7, E_{12}, E_4, E_9, E_{11}, E_{14}, E_{17}\}$

And the positive lower approximation the definite positive cases will be

$E_{\text{positive}} = \{E_4, E_5, E_7, E_9, E_{11}, E_{12}, E_{14}\}$ so the boundary region for positive case will be only a single record E_{17} and there are lots of ambiguity that is in high a_5 there also of negative cases that is E_2 similarly for the moderate cases the negative cases also lies that is E_6, E_{10}, E_{18} that is some cases of a_5 we have also negative cases so in table-3 we drop a_1, a_5 so after dropping a_1 and a_5 from table 2 we have the new reduct table, named as table-3

Reduced Table-3 from Table-2

E	a_2	a_3	a_4	d
E_1	b_2	b_1	b_1	c_2
E_2	b_2	b_1	b_3	c_2
E_4	b_2	b_2	b_3	c_1
E_5	b_3	b_3	b_3	c_1
E_6	b_2	b_2	b_2	c_2
E_7	b_2	b_2	b_2	c_1
E_8	b_1	b_1	b_1	c_2
E_9	b_2	b_2	b_3	c_1
E_{10}	b_2	b_2	b_2	c_2
E_{11}	b_3	b_3	b_3	c_1
E_{12}	b_2	b_3	b_1	c_1
E_{13}	b_2	b_2	b_2	c_2
E_{14}	b_3	b_3	b_3	c_1
E_{15}	b_1	b_1	b_1	c_2
E_{16}	b_1	b_1	b_1	c_2
E_{17}	b_3	b_2	b_2	c_1
E_{18}	b_2	b_2	b_3	c_2

Upon analyzing table-3 we have the following result that is (E₅,E₁₁,E₁₄), (E₈,E₁₅,E₁₆), (E₁₀,E₁₃), forms group and (E₆,E₇) (E₉,E₉,E₁₈),ambiguous so we keep on record for each group and delete all records which give ambiguous result so we have the new table appears as table-4 given as follows

Reduced Table-4 from Table-3

E	a ₂	a ₃	a ₄	D
E ₁	b ₂	b ₁	b ₁	c ₂
E ₂	b ₂	b ₁	b ₃	c ₂
E ₅	b ₃	b ₃	b ₃	c ₁
E ₈	b ₁	b ₁	b ₁	c ₂
E ₁₀	b ₂	b ₂	b ₂	c ₂
E ₁₂	b ₂	b ₃	b ₁	c ₁
E ₁₇	b ₃	b ₂	b ₂	c ₁

further reduction Table -4 is not possible

From the table we are develop an algorithm is as follows

1. High (High fever), muscle and joint pain and pain behind your eye and leads to positive case of dengue
2. Moderate High fever , high muscle and joint pain and low pain behind the eye leads to positive case of dengue
3. Moderate fever, muscle and joint pain and pain behind the eye leads to negative case of dengue
4. Low High fever , muscle and joint pain and pain behind the eye leads to negative case of dengue
- 5.High High fever moderate muscle and joint pain and moderate pain behind the eye leads to positive case of dengue

This give the idea that Very High fever, Muscle and joint pain and Pain behind the eye in general symptom for dengue.

Time complexly analysis- For finding the reduct we are comparing each record with another suppose there are n records then time complexity will be

$n+(n-1)+(n-2)+(n-3)+\dots+1=n(n-1)/2$ that is of $\Theta(n^2)$ that is the worst case analysis average case analysis of breaking the table $\Theta(n \lg n)$ we are taking the average case analysis for dividing the table because on an average single table break down to half of it's size every time so the entire complexity will be $\Theta(n \lg n + n^2)$

Statistical validation- For validate our findings we basically depends upon

chi-square test for this purpose we consider we take a survey by taking data regarding the positive case and we are not focused on one medical centre to collect data we approached several hospital and the apply chi square test to validate our claim. Chi square test- Expected 15%,10%,15%,20%,30%,15% and the Observed samples are 25,14,34 45,62,20 so totaling these we have

total of 200 samples so expected numbers of samples per each day as follows 30,20,30,40,60,30 . We then apply chi square distribution to verify our result assuming that H₀ is our hypothesis that is correct H₁ as alternate hypothesis that is not correct , Then we expect sample in six cases as

chi squared estimation formula is $\sum(O_i - E_i)^2 / E_i$ where $i=0,1,2,3,4,5$ so the calculated as follows $X^2=(25-30)^2/20+(14-20)^2/20+(34-30)^2/30+(45-40)^2/40+(62-60)^2/60+(20-30)^2/30$

$$X^2=25/20+36/20+16/30+25/40+4/60+100/30=7.60$$

the tabular values we have with degree of freedom 5 we get result 11.04 the values of X² lies much below the tabular value which validate our result

Future work – Our work can be extended to the film industries , general market system ,small scale industries .

5. REFERENCES

- [1] Cios, K., W. Pedrycz and R. Swiniarski (1998). *Data Mining Methods for Knowledge Discovery*. Kluwer Academic
- [2] Wolf, S., H. Oliver, S. Herbert and M.Michael (2000). Intelligent data mining for medical quality management
- [3] Se-Ho, Ch., and P. Rockett (2002). The training of neural classifiers with condensed datasets. *SMCB*, **32**(2),202–206.
- [4] Setiono, R. (2000). Generating concise and accurate classification rules for breast cancer diagnosis. *Artificial Intelligence in Medicine*, **18**(3), 205–219
- [5] Cheeseman, P., and J. Stutz (1996). Bayesian classification (AutoClass): theory and results. In U.M.Fayyad
- [6] Grzymala–Busse, J., Z. Pawlak, R. Slowinski and W. Ziarko (1999). Rough sets. *Communications of the ACM*
- [7] Hassanien, A.E. (2003). Classification and feature selection of breast cancer data based on decision tree algorithm
- [8] Parido, A., and P. Bonelli (1993). A new approach to fuzzy classifier systems. In *Proceedings of the Fifth International Conference on Genetic Algorithms*. pp. 223–230
- [9] Lin, T.Y., and N. Cercone (1997). *Rough Sets and Data Mining*. Kluwer Academic Publishers. Ning, S., H. Xiaohua, W. Ziarko and N. Cercone (1994). A generalized rough sets model. In *Proceedings of the 3rd Pacific Rim International Conference on Artificial Intelligence*, Vol. 431. Beijing, China. Int. Acad.Publishers. pp. 437–443.
- [10] Pawlak, Z. (1991). *Rough Sets-Theoretical Aspect of Reasoning about Data*. Kluwer Academic Publishers.
- [11] Pawlak, Z., J. Grzymala–Busse, R. Slowinski, W. Ziarko (1995). Rough sets. *Communications of the ACM*
- [12] Renu Vashist Prof.M.L.Garg Rule Generation based on Reduct and Core :A rough set approach International Journal of Computer Application(0975-887) Vol 29 September -2011 Page 1-4