Intelligent System for Predicting and Diagnosis of Breast Cancer

Saeed Khodary M. Hamouda, Reda. H. Abo El-Ezz, Mohammed E. Wahed

Abstract- Cancer is a disease in which cells in the body grow out of control. When cancer starts in the breast, it is called breast cancer. Breast cancer is one of the major death causing diseases of the women in the world. Every year more than million women are diagnosed with breast cancer more than half of them will die because of inaccuracies and delays in diagnosis of the disease. High accuracy in cancer prediction is important to improve the treatment quality and the survivability rate of patients. This paper, presents a Rough Set Theory (RST) as an efficient and intelligent technique, to analyze breast cancer dataset, approximate set classification, and improve the accuracy of diagnosis with limited attributes. (RST) estimates the risk of the breast cancer at the earlier stage, and may help researchers and consultants for predicting and diagnosis of breast cancer early.

Keywords- Breast Cancer, Diagnosis, MATLAB, Predictive, Rough Set Theory, Survivability.

I. INTRODUCTION

Cancer is a disease in which cells in the body grow out of control. When cancer starts in the breast, it is called breast cancer[2],[4]. Breast cancer is characterized by the uncontrolled growth of abnormal cells in the milk producing glands of the breast. There is no sure way to prevent from the breast cancer, but the women can reduce their breast cancer risk, breast cancer also develops in men. Earlier detection of cancer is curable and may increase the survivability rate of patients.

woman's must be checking a breasts for cancer before she has any symptoms , there are various tests are available for predict and diagnosis of breast cancer[1-4]**.** Large amounts of data about the patients with their medical conditions are presented in the medical databases. Analyzing all these databases is one of the difficult tasks in the medical environment. Thus, we need to extract knowledge about the patients from databases. Medical databases have a large quantity of information about patients and their medical conditions.

The rough set theory (RST) is a mathematical tool for extracting knowledge from uncertain and incomplete database information [11],[13]. Rough set methods utilize the comparison between elements, e.g.., discernibility, indiscernibility, and similarity. This paper is organized as follows; Section. 2 an overview of breast cancer, Section 3 describes fundamental concept of rough set theory, Section 4 describes the proposed approach and breast cancer diagnosis using rough set, and Section 5 the Conclusion.

Revised Version Manuscript Received on March 23, 2017.

Saeed Khodary M. Hamouda, Research Scholar, Computer Laboratories, Zagazig University, Zagazig, Egypt. E-mail: HamoudaSaeed@yahoo.com

Reda. H. Abo El-Ezz, Prof. at Computer& System Dept., Faculty of Engineering, Al-Azhar University, Cairo, Egypt. E-mail: drosprof.dr_Reda@yahoo.com

Mohammed E. Wahed, Prof. at Faculty Of Computers and Informatics, Suez Canal University, Ismailia, Egypt, E-mail: mewahed@yahoo.com

II. BREAST CANCER OVERVIEW

A. Breast Cancer

The **anatomy** of normal breast [1],[2],[4] shows in picture of "Fig.1"this picture shows the lobes and ducts inside the breast, it also shows lymph nodes near the breast.

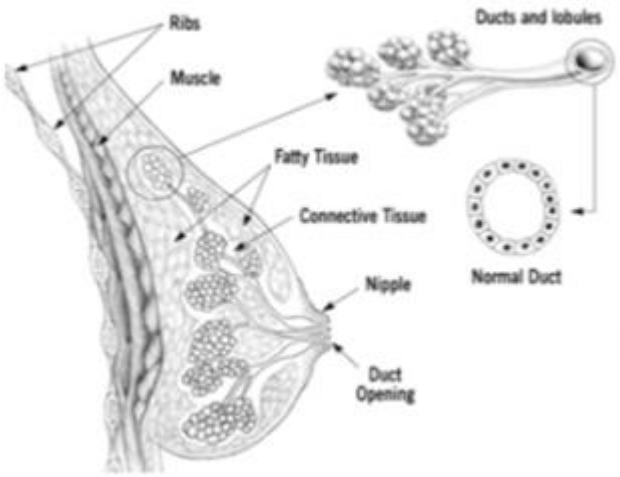

Figure 1. Normal Breast Anatomy

B. Breast Cancer Types: [1-4]

B.1 Ductal Carcinoma:

The first and most common type of breast cancer is ductal carcinoma "Fig.2" in this type the cancer begins in cells that line a breast duct, about 7 of every 10 women with breast cancer have ductal carcinoma.

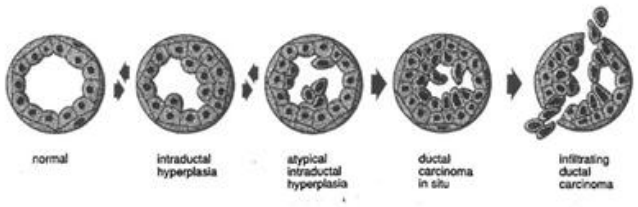

Figure 2. Ductal Carcinoma

B.2 Lobular Carcinoma:

Another type of breast cancer is lobular carcinoma. This cancer begins in a lobule of the breast. About 1 of every 10 women with breast cancer has lobular carcinoma.

B.3. Ductal and Lobular Carcinoma:

Other women have a mixture of ductal and lobular type or they have a less common type of breast cancer.

B.4 Inflammatory Breast Cancer:

The inflammatory cancer is other type of breast cancer. It occurs in about 1 of every 100 women with invasive breast cancer.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

C. Breast Cancer Stages:

The stage of breast cancer depends on the size of the tumour and determines whether it has spread and how far it has spread. Treated describe the stages of breast cancer to: stage 0, stage I, stage II, stage III, and stage IV. Stage I is early-stage breast cancer, and Stage IV is advanced-stage breast cancer that has spread to other parts of the body. The stage of a cancer is one of the most important factors in selecting treatment options, and it uses the *Tumors, Nodes and Metastasis* (TNM) system. When a patient's *T*, *N*, and *M* categories have been determined then this information is combined in a process known as stage grouping to determine a woman's disease stage[2],[4],[5],[6].

In *stage 1*, tumor size less than 2 cm and the cancer hasn't spread beyond the breast. *Stage II* tumor size less than or equal 5 cm and cancer may have spread to the lymph nodes. *Stage III* tumor size greater than 5 cm and tumors have spread to the lymph nodes, and possibly the chest wall. *Stage IV* means cancer has reached other, remote parts of the body.

D. Effect of Breast Cancer Stages on Survival

The effect of breast cancer stages on the survival [7],[8] vary depending on the stages of breast cancer,

Non-invasive and the early stages have a better chances of survival than that for the metastatic breast cancer (stage IV) which is the stage wherein the cancer has spread beyond the neighboring tissues. Table (1) shows the 5-year, and 10-year survivability rate of a cancer patient**.**

Table 1. Breast Cancer Survivability Rate

Stage	Description	5 – year survival	10 – year survival
Stage 0	T=0, LN=0, M=0	95%	90%
Stage I	T <= 2cm ,	85%	70%
	LN=1:3, M=0		
Stage II	T>2cm& <=5cm,	70%	50%
	L.N=3:9, M=0		
Stage III	T> 5cm, LN>9, M=0	55%	30%
Stage IV	T=any, LN=any, M=1	5%	2%
Metastasis			

E. E. Breast Cancer Risk Factors

Women who have one or more breast cancer risk factors never develop breast cancer, while many women with breast cancer have no known risk factors. Even when a woman with risk factors develops breast cancer, it's hard to know just how much these factors might have contributed.

Some risk factors [1],[2],[4],[9] can't be changed, like a person's age or race, other risk factors can changed, such as smoking, drinking, and diet. Some factors affect risk more than others, and the risk for breast cancer can change over time.

E.1 Risk Factors You Cannot Change

E.1.1 Getting Older: The risk of developing breast cancer increases with your age. Approximately, 75% of all breast cancers are diagnosed in women 50 years and over.

E.1.2 Genetic Risk Factors: About 5-10% of breast cancers are inherited, most commonly the genes BRCA1 and BRCA2.

E.1.3 Family History: If a woman has a personal or family history of breast cancer she is at increased risk of developing breast cancer.

E.1.4 HER2/neu Status: Cancers that are HER2-positive have too many copies of the HER2/neu gene, these cancers tend to grow and spread more aggressively than other breast cancers. All newly diagnosed invasive breast cancers should be tested for HER2/neu.

E.1.5 Gender: Main risk factor of breast cancer is being a woman. The disease is about 100 times more common amongst women than men.

E.2 Risk Factors You Can Change:

E.2.1 Lifestyle Factors: For example, being overweight or obese after the menopause, physical inactivity, smoking, a high fat diet, and high alcohol consumption can play an important role in the development of breast cancer.

E.2.2 Hormone Replacement Therapy: Estrogen hormone therapy has been used to relieve symptoms of menopause and to help prevent osteoporosis but studies reveal that it also causes more risk of breast cancer.

E.2.3 A Late First Pregnancy: Women who have a late first pregnancy (after the age of 35) are more likely to develop breast cancer.

F. Breast Cancer Symptoms

Some people have no symptoms, other people may notice a change in the breast or doctor may find an unusual breast change during a physical examination. [2],[9]. The major symptoms of breast cancer are:

- New lump in the breast or underarm (armpit).
- Thickening or swelling of part of the breast.
- Irritation or dimpling of breast skin.
- Redness or flaky skin in the nipple area or the breast.
- Pulling in of the nipple or pain in the nipple area.
- Nipple discharge other than breast milk, including blood.
- Any change in the size or the shape of the breast.
- Pain in the breast.

III. ROUGH SET THEORY OVERVIEW

The theory of rough set theory [12] was developed by ZdzislawPawlakin the early 1980's, and other researchers. The main goal of the rough set analysis is induction of (learning) approximations of concepts, it is intelligent mathematical tool for managing uncertainty that is used for the discovery of data dependencies, to evaluate the importance of attributes, to discover patterns in data, to reduce redundancies, and to recognise and classify objects.

The theory has found many interesting applications. The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, decision support, engineering, environment, expert systems, banking, medicine and others. The main advantage of rough set

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

theory in data analysis is that it does not need any preliminary or additional information about data like probability in statistics. [12],[13]

A. Rough Set Basics Concept

RS [12],[13],[14] starts from *Information System(IS) table*, the columns of which are labeled by *attributes*, the rows by *objects* of interest and entries of the table are *attribute values.* An *Information System* is a pair S = *(U,A)* where *U* is a non-empty finite set of objects called the *universe* and *A* is a non-empty finite set of *attributes* such that $a : U \rightarrow Va$ for every $a \in A$. The set *Va* is called the *value set* of *a*, elements of *U* are called objects.

A *Decision System (DS)* is a special case of information systems, $S = (U, A \cup \{d\})$, where *d* (is not element of *A*) is the decision attribute. The elements of *A* are called conditional attributes or simply conditions.

A.1 Approximations Set

Let X is a subset of U, i.e. $X \subseteq U$.

A.1.1 Lower Approximation: Consists of all objects which *surely* belong to the set.

$$\underline{R}(X) = \{x \in U \mid [x]\ R \subseteq X\}$$

A.1.2 Upper Approximation: Contains all objects which *possibly* belong to the set.

$$\bar{R}(X) = \{x \in U \mid [x]\ R \cap X \neq \phi\}.$$

A.1.3 Boundary Region: The difference between the *upper* and the *lower* approximation constitutes the *boundary region* of the rough set.

Boundary positive and negative regions [15] are described as below.

$$BN_R(X) = |\ \bar{R}(X) - \underline{R}(X)\ |.$$

$$POS_R(X) = \underline{R}(X).$$

$$NEG_R(X) = U - \bar{R}(X).$$

A member of the negative region *NEG (X)* does not belong to *X*, a member of the positive region *POS*(*X*) belongs to *X* , and only one member of the boundary region *BND(X)* belongs to *X*, these approximation set and regions shown in the "Fig.3".

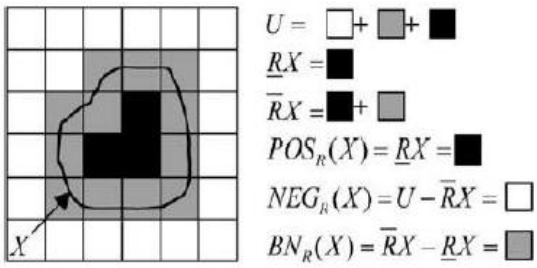

Figure 3. The approximations and regions of set *X*.

A.1.4 Approximation Accuracy: The accuracy of approximation (accuracy of roughness) of any subset $X \subseteq U$ with respect to $R \subseteq A$, denoted $\alpha_R(X)$ is measured by:

$$\alpha R(X) = |\underline{R}(X) / \bar{R}(X)\ |.$$

where v*X* vdenotes the cardinality of *X*. For empty set ϕ , we define $\alpha R(\phi) = 1$.

Obviously, $0 \leq \alpha R(X) \leq 1$.

If $\alpha R(X) = 1$. Thus, the set *X* is *crisp* with respect to *R*, otherwise

If $\alpha R(X) < 1$, *X* is *rough* with respect to *R*.

B. Reducts and Core:

B.1 Reduct: Reducts are minimal subsets of attributes which contain a necessaryportion of information fromthe set of all attributes[16],[17]. Some attributes in decision table may be superfluous (redundant), we reduce unnecessary attributes and only keep those that are required to tell two objects that have different values on some attribute in *A* apart all objects that have the same values on the attributes in *B* are indiscernible, or in the *IND***S**(*B*)-relation to each other.

In order to express the idea of reduct, let $\boldsymbol{B} \subseteq A$ and $a \in B$ in an information system *I = (U, A)* where *U* is the universe of objects, A is set of attributes, and *R(B)* is a binary relation.

- *a* is dispensable in *B* if *R(B) = R(B –{a})*; otherwise *a* is indispensable in *B.*
- Set *B* is independent if all its attributes are indispensable.
- $B' \subseteq B$ is a reduct of *B* if *B'* is independent and *R*(*B'*) = *R*(*B*).

B.2 Core: The core is the set of all indispensable attributes, i.e., it is the intersection of all reducts.

Let *Red (B)* is the set of all reducts of *B* in an information system *I = (U, A)* where $B \subseteq A$ then the core of *B* is defined as:

$$Core\ (B) = \cap\ Red\ (B)$$

The core is included in every reduct, i.e., each element of the core belongs to some reduct. Thus, the core is the most important subset of attributes, for none of its elements can be removed without affecting the classification.

IV. THE PROPOSED APPROACH

A. MATERIALS AND METHODS

Breast cancer data are often presented as a table, columns of which are labeled by attributes, rows by objects of interest and entries of the table are attribute values, in a table containing information about patients suffering from a certain disease objects. [18]

The data sets used in our experiments consists of 60 samples, each sample consists of ten of measurement features and decision attribute. All attributes have a data type value ranging from 0 to 5. Table (2) briefs the attributes of breast cancer dataset.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Table 2. Description of Breast Cancer Dataset.

Symbol	Attribute Name	Discrete Values					
		0	1	2	3	4	5
AG	Age Group	-	Age <=35	Age >35	-	-	-
FH	Family History	No	Yes	-	-	-	-
HER	HER2/neu Status	No	Yes	-	-	-	-
NS	Lymph Node Status	No Node	1< Node & <3	3< Node <9	Node >9	-	-
TS	Tumor Size (cm)	No Tumor	Tumor <2	Tumor >2& <=5	Tumor >5	Chest Wall	-
HT	Histological Type	No	Ductal	Lobular	Ductal-Lobular	-	-
HG	Histological Grade	No	Grade I	Grade II	Grade III	-	-
ERS	Estrogen Receptor Status	-	<20	20-49	50-100	>100	-
PRS	Progesterone Receptor Status	-	<5	5-30	31-50	51-100	>100
M	Metastasis	No	Yes	-	-	-	-
D	Decision (Stage)	Stage 0	Stage I	Stage II	Stage III	Stage IV	-

Information system table contains dataset about 60 patients suffering from breast cancer disease taken from NCI, Egypt, table (3) the breast cancer dataset identifies whether the patient's stage (stage 0, stage I, stage II, stage III, or stage IV).

Table 3. Information System for Breast Cancer Dataset

Objects	*Condition Attribution*										*Decision (Stage)*
	AG	FH	HER	NS	TS	HT	HG	ERS	PRS	M	
1	2	1	0	0	3	1	1	2	3	0	2
2	2	1	1	1	4	1	2	3	5	1	4
3	1	0	0	0	0	0	0	1	1	0	0
4	1	0	0	0	1	1	1	2	2	0	1
5	1	0	0	0	2	2	0	3	2	0	2
6	2	0	0	1	1	1	1	2	2	0	2
7	2	1	1	1	1	1	1	3	3	0	2
8	2	1	1	2	4	1	2	4	5	1	4
9	2	0	1	2	4	2	3	4	4	1	4
10	2	1	0	3	1	1	1	3	5	0	3
11	2	1	1	3	1	3	2	4	4	0	3
12	2	1	0	2	1	1	1	3	2	0	2
13	2	1	1	0	3	1	2	4	5	1	4
14	2	1	1	1	4	2	2	4	5	1	4
15	2	1	1	3	2	1	1	4	3	0	3
16	2	1	1	3	3	1	2	3	4	0	3
17	2	0	1	3	4	1	3	4	4	1	4
18	2	0	1	3	3	1	2	4	5	1	4
19	2	1	1	1	1	3	3	4	5	0	2
20	2	1	1	1	2	2	2	4	3	0	2
21	2	1	1	1	3	1	2	4	5	0	3
22	1	0	0	0	0	0	0	1	2	0	0
23	1	0	0	0	0	0	1	2	1	0	0
24	1	0	0	0	1	1	1	1	2	0	1
25	2	1	0	1	2	1	2	2	3	0	2
26	1	1	0	1	3	3	1	2	2	0	3
27	2	1	1	3	2	1	2	4	4	0	3
28	2	1	1	3	3	2	2	3	3	0	3
29	1	1	0	1	2	1	1	2	3	0	2
30	2	1	0	2	3	1	3	3	4	0	3
31	2	1	0	2	3	3	2	4	3	0	3
32	2	1	1	0	4	1	3	4	3	1	4
33	1	0	0	0	0	0	0	2	2	0	0
34	1	0	0	0	1	1	1	1	2	0	1
35	1	1	0	1	2	1	2	2	3	0	2
36	1	1	0	1	1	1	1	2	2	0	2
37	1	1	0	1	2	3	1	2	3	0	2
38	2	1	0	2	3	1	1	3	4	0	3
39	2	1	0	2	3	1	2	4	3	0	3
40	1	1	0	2	1	1	3	3	4	0	3
41	1	0	0	0	2	2	1	3	2	0	2
42	2	0	1	2	2	1	2	4	4	0	3
43	1	0	0	0	1	2	1	3	3	0	2
44	2	1	1	2	1	1	2	3	4	0	3
45	1	1	1	2	3	1	3	3	3	0	3
46	1	0	0	1	1	1	1	2	3	0	2
47	1	0	0	0	0	0	0	1	2	0	0
48	1	0	0	0	2	1	1	2	2	0	2
49	2	1	1	2	3	1	3	4	5	1	4
50	2	1	1	2	2	1	2	3	4	0	3
51	2	1	0	0	3	3	1	2	2	0	2
52	1	1	0	0	3	2	0	3	3	0	2
53	2	0	0	0	3	1	1	2	2	0	2
54	2	1	1	1	4	3	1	4	5	0	3
55	2	1	1	3	1	1	3	4	5	1	4
56	1	0	0	0	0	0	1	1	1	0	0
57	2	1	0	1	2	1	1	2	2	0	2
58	2	0	0	1	1	1	0	3	3	0	2
59	2	1	0	2	2	1	2	3	4	0	3
60	2	1	0	2	2	2	1	4	4	0	3

All patients have been divided by experts into five classes (stages) corresponding to their health status. The stages are: stage 0, stage I, stage II, stage III, and stage IV.

The problem was to find the description of each class in terms of data available for each patient of this class, check whether the set of attributes is dependent or independent, find reducts for each class, and compute the core and accuracy of descriptions. We use the MATLAB programing to compute these processes. [18]

B. The Program

The program is written by MATLAB programming which**:**

(i) Computes lower, upper, and accuracy approximations.

(ii) Computes reduct and core of sets of attributes.

C. Approximations and Accuracy

Algorithm Procedure

%**input**: IS as a decision table T = (U,A,D,f)

%where U= $x_1,x_2,\ldots,x_m$, A= $a_1,a_2,\ldots,a_n$, D= $h_1,h_2,\ldots h_q$, f is information function

%**output**: NO Rules – set of if-then rules for T;

begin

Create matrix S ,size m x (n+1), from table T,

S={$s_1,s_2,\ldots,s_{m*(n+1)}$}

if any object s_x =ø then //($x=_{1,2,\ldots, m*(n+1)}$)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

for every object s_x do replace s_x by -1
if any vector X=[x_1,...,x_i]contain -1
then
//i=1,2,...m
delete x_i
end {if} ; **end** {for}; **end** {if}
for reduced table T
do compute I
// I= indiscernibility relations IND(A)
if IND(A) contain redundant values
then delete redundant values
end {for}
for T,I compute lower approximation $\underline{A}(X)$
if $x_i \in \underline{A}(X)$ then
create rule and insert it to Rules
end {if}
end {for}
for T,I compute upper approximation $\overline{A}(X)$
if $x_i \in \overline{A}(X)$ then
create rule and insert it to Rules
end {if}; **end** {for}
end {algorithm}
accuracy = $\underline{A}(X)/\overline{A}(X)$

D. Results of Approximations Set

After constructed Approximation MATLAB program we obtained the lower, upper, and accuracy attributes description of each class (stage) results. We see that stages 0, III, and IV are *crisp* describable in the system, and the remaining stages are *roughly* describable with the accuracy given in the last column. That is to say that data (symptoms) available from the patients characterize exactly classes 0, III, and IV only, and the remaining classes not are characterized exactly by these data. The accuracy of all stages (from stage 0 to stage IV) are: 100%, 60%, 92%, 100%, and 100% respectively, as shows in table (4).

Table 4. The Lower, Upper, and Accuracy of Description of Each Stage

Stage Number	Lower Approx.	Upper Approx.	Accuracy
Stage 0	8	8	100%
Stage I	3	5	60%
Stage II	23	25	92%
Stage III	22	22	100%
Stage IV	10	10	100%

E. Reduction and Core of Attributes

According to indiscernibility and reduction Condition [17-19] in order to check whether the set of attributes is dependent or not to remove repeated condition attributes in the decision table based on indiscernible and reduction of rough set theory and method.

In this section we will show reducts of attributes, i.e. minimal sets of attributes necessary for the description of these classes (stages). I used a program written in MATLAB to achieve reduction of rough functions. Specific procedures are as follows:

QUICKREDUCT($\mathbb{C}$,$\mathbb{D}$).
$\mathbb{C}$, the set of all conditional attributes;
$\mathbb{D}$, the set of decision attributes.
(1) $R \leftarrow \{\}$
(2) **do**
(3) $T \leftarrow R$
(4) $\forall x \in (\mathbb{C} - R)$
(5) **if** $\gamma_{R \cup \{x\}}(\mathbb{D}) > \gamma_T(\mathbb{D})$
(6) $T \leftarrow R \cup \{x\}$
(7) $R \leftarrow T$
(8) **until** $\gamma_R(\mathbb{D}) == \gamma_{\mathbb{C}}(\mathbb{D})$
(9) **return** R

F. Reduction Results

After constructed reduction MATLAB program we obtained that attributes 1, 2, 6, and 7 are superfluous (redundants), attributes 3, 4, 5, 8, 9, and 10. (i.e. HER2/neu Status, Lymph Node Status, Tumor Size, Estrogen Receptor Status, Progesterone Receptor Status, and Metastasis) are the most effective symptoms in predict and accurate diagnosis of breast cancer stages, and the attributes 4, 5, and 10 (i.e. Lymph Node Status, Tumor Size, and Metastasis) are the core of these symptoms. The summary of these results are shown in table (5).

Table 5. Summary of Reduction Operations

Description	Attrib. No.	Attributes (Symptoms) Name
The most effective symptoms in predict & accurate diagnosis of breast cancer	3,4, 5, 8,9,10	HER2/neu Status Lymph Node Status, Tumor Size, Estrogen Receptor Status, Progesterone Receptor Status, and Metastasis.
The core of these symptoms	4, 5,10	Lymph Node Status, Tumor Size, and Metastasis.
The superfluous (redundants) attributes	1, 2, 6 ,7	Age Group, Family History, Histological Type, Histological Grade.

The results of reduction attributes and merge the same objects are simplified in the decision table as shown in table (6).

Table 6. The Decision Table After Reduction Algorithm

Objects	*Condition Attribution*						*Decision (Stage)*
	HER	NS	TS	ERS	PRS	M	
1	0	0	3	2	3	0	2
2	1	1	4	3	5	1	4
3,56	0	0	0	1	1	0	0
4	0	0	1	2	2	0	1
5,41	0	0	2	3	2	0	2
6,36	0	1	1	2	2	0	2
7	1	1	1	3	3	0	2
8	1	2	4	4	5	1	4
9	1	2	4	4	4	1	4
10	0	3	1	3	5	0	3
11	1	3	1	4	4	0	3
12	0	2	1	3	2	0	2
13	1	0	3	4	5	1	4
14	1	1	4	4	5	1	4
15	1	3	2	4	3	0	3

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Objects	*Condition Attribution*						*Decision (Stage)*
	HER	NS	TS	ERS	PRS	M	
16	1	3	3	3	4	0	3
17	1	3	4	4	4	1	4
18	1	3	3	4	5	1	4
19	1	1	1	4	5	0	2
20	1	1	2	4	3	0	2
21	1	1	3	4	5	0	3
22,47	0	0	0	1	2	0	0
23	0	0	0	2	1	0	0
24,34	0	0	1	1	2	0	1
25,29,35,37	0	1	2	2	3	0	2
26	0	1	3	2	2	0	3
27	1	3	2	4	4	0	3
28	1	3	3	3	3	0	3
30,38	0	2	3	3	4	0	3
31,39	0	2	3	4	3	0	3
32	1	0	4	4	3	1	4
33	0	0	0	2	2	0	0
40	0	2	1	3	4	0	3
42	1	2	2	4	4	0	3
43	0	0	1	3	3	0	2
44	1	2	1	3	4	0	3
45	1	2	3	3	3	0	3
46	0	1	1	2	3	0	2
48	0	0	2	2	2	0	2
49	1	2	3	4	5	1	4
50	1	2	2	3	4	0	3
51,53	0	0	3	2	2	0	2
52	0	0	3	3	3	0	2
54	1	1	4	4	5	0	3
55	1	3	1	4	5	1	4
57	0	1	2	2	2	0	2
58	0	1	1	3	3	0	2
59	0	2	2	3	4	0	3
60	0	2	2	4	4	0	3

G. The Decision Rules

All decision rules which can be generated from objects represented in decision table are listed below:

Rule 1: *if* (TS=0) ∧ (NS=0) ∧ (M=0) ∧ (HER=0) ∧ (ERS=1 ∨ ERS=2) ∧ (PRS=1 ∨ PRS=2) *then* (Decision = Stage 0) → objects (*patients*) {3,22,23,33,47,56}

Rule 2: *if* (TS=1) ∧ (NS=0) ∧ (M=0) ∧ (HER=0) ∧ (ERS=1 ∨ ERS=2) ∧ (PRS=2) *then* (Decision = Stage 1) → objects (*patients*) {4,24,34,48}

Rule 3: *if* (TS=0 ∨ TS=1) ∧ (NS=1) ∧ (M=0) ∧ (HER=0 ∨ HER=1) ∧ (ERS=2 ∨ ERS=3) ∧ (PRS=2 ∨ PRS=3) *then* (Decision= Stage 2) → objects *(patients)* {6,7,19,36,46,58}

Rule 4: *if* (TS=2) ∧ (NS=0 ∨ NS=1) ∧ (M=0) ∧ (HER=0 ∨ HER=1) ∧ (ERS=2 ∨ ERS=3) ∧ (PRS=2 ∨ PRS=3) *then* (Decision = Stage2) → *objects* (*patients*) {5,20,25,29,35,37,41,48,57}

Rule 5: *if* (TS=3) ∧ (NS=0) ∧ (M=0) ∧ (HER=0 ∨ HER=1) ∧ (ERS=2 ∨ ERS=3) ∧ (PRS=2 ∨ PRS=3)) *then* (Decision = Stage 2) → objects (*patients*) {1,51,52,53}

Rule 6: *if* (TS=0 ∨ TS=1 ∨ TS=2) ∧ (NS=2) ∧ (M=0) ∧ (HER=0 ∨ HER=1) ∧ (ERS=3 ∨ ERS=4) ∧ (PRS=3 ∨ PRS=4 ∨ PRS=5) *then* (Decision = Stage 3) → objects (*patients*) {12,40,42,44,50,59,60}

Rule 7: *if* (TS=3) ∧ (NS=1 ∨ NS=2) ∧ (M=0) ∧ (HER=0 ∨HER=1) ∧ (ERS=3 ∨ ERS=4) ∧ (PRS=3 ∨ PRS=4 ∨ PRS=5) *then* (Decision = Stage 3) → objects (*patients*) {21,26,30,31,38,39,45}

Rule 8: *if* (TS=4) ∧ (NS=0 ∨ NS=1 ∨ NS=2 ∨ NS=3) ∧ (M=0) ∧ (HER=0 ∨ HER=1) ∧ (ERS=3 ∨ ERS=4) ∧ (PRS=3 ∨ PRS=4 ∨ PRS=5) *then* (Decision =Stage 3) → objects (*patients*) {54}

Rule 9: *if* (TS=0 ∨ TS=1 ∨ TS=2) ∧ (NS=3) ∧ (M=0) ∧ (HER=0 ∨ HER=1) ∧ (ERS=3 ∨ ERS=4) ∧ (PRS=3 ∨ PRS=4 ∨ PRS=5) *then* (Decision = Stage 3) → objects (*patients*) {10,11,15,16,27,28}

Rule 10: *if* (TS=0 ∨ TS=1 ∨ TS=2 ∨ TS=3 ∨ TS=4) ∧ (NS=0 ∨ NS=1 ∨ NS=2 ∨ NS=3) ∧ (M=1) ∧ (HER=1) ∧ (ERS=3 ∨ ERS=4) ∧ (PRS=3 ∨ PRS=4 ∨ PRS=5) *then* (Decision = Stage4) → objects (*patients*) {2,8,9,13,14,17,18,32,49,55}

V. CONCLUSION

Predicting and diagnosis of breast cancer early is very important to improve the treatment quality and survivability rate of patients.

In this paper we presented a Rough Set Theory (RST) as an efficient technique for predicting and diagnosing breast cancer in early stages, where RST technique proved that it more efficiency in the field of computational biology because of the effective classification and high diagnostic capability.

REFERENCES

1. National Breast Cancer Foundation, "About Breast Cancer", (2016) available at: http://www.nationalbreastcancer.org/about-breast-cancer
2. Cancer.Net Editorial Board, "Breast Cancer", (2016), available at: http://www.cancer.net/cancer-types/breast-cancer
3. American Cancer Society, "Breast Cancer types", (2016) available at: http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-breast-cancer-types
4. National Cancer Institute, (2016) available at: http://www.cancer.gov/cancertopics/types/breast
5. American Cancer Society, California Cancer Registry (CCR), Cancer Stage at Diagnosis, (2016), available at: http://www.cdph.ca.gov or http://www.ccrcal.org
6. Cancer Staging, General Cancer Classification, Staging, And Grouping Systems, available at: http://stedmansonline.com/webFiles/Dict-Stedmans28/APP21.pdf
7. Carter et. Al. (January-1989), "Relation of Tumor Size, Lymph Node Status, and Survival in 24,740 Breast Cancer Cases", PROGNOSTIC FACTORS FOR BREAST CANCER, (1989) Vol. 63.
8. Abdelghani Bellaachia, Erhan Guven, (2005), "Predicting Breast Cancer Survivability using Data Mining Techniques", Department of Computer Science, The George Washington University Washington DC 20052.
9. Breast Cancer Organization, available at: http://www.breastcancer.org/ risk/factors/
10. Breast Cancer Organization, available at: http://www.breastcancer.org/symptoms/
11. American Cancer Society, California Department of Health Services, (2016), "A Woman's Guide To Breast Cancer Diagnosis And Treatment, Breast Cancer Early Detection Program", available at: http://www.ombc.ca.gov/forms_pubs/breast_cancer.pdf
12. Silvia Rissino and Germano Lambert-Torres, (2009), "Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications", Federal University of Rondonia, Itajuba Federal University Brazil.
13. Pawlak, Z., Grazymala-Busse, J. W., Slowinski, R., and Ziarko, W.,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

(1995), "Rough sets. Communications of the ACM", 38, 88-95.
14. Lech Polkowski, Shusaku Tsumoto, Tsau Y. Lin, (May-2000), "Rough Set Methods and Applications", Warsaw, Izumo City, San Jose.
15. Hirano S. and Tsumoto S., (2005), "Rough representation of a region of interest in medical images", International Journal of Approximate Reasoning, vol.40, pp.2334.
16. ZDZISJdAW PAWLAK , (1984), "Rough classification", Institute of Computer Science, Polish Academy of Sciences, Poland, Int. J. Man-Machine Studies, vol.20,p 469-483.
17. Zdzisław Pawlak, (2002), "Rough set theory and its applications", Journal of Telecommunications and Information Technology 44-000 Gliwice, Poland.
18. Jan G. Bazan1, Hung Son Nguyen, Sinh Hoa Nguyen Piotr Synak , Jakub Wr6blewski, "Rough Set Algorithms in Classification Problem", Institute of Mathematics, Warsaw University, 2, 02-097 Warsaw , Poland.
19. Stankov, J. E. Bargallo-Rocha, A. Namendys-Silva Silvio, M. T. Ramirez, K. Stankova-Ninova, and A.Meneses-Garcia, (2012), "Prognostic Factors and Recurrence in Breast Cancer", Experience at the National Cancer Institute of Mexico, International Scholarly Research Network ISRN Oncology, Article ID 825258, 7pages doi:10.5402/2012/825258.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.