



# A ROUGH SET BASED DATA MODEL FOR HEART DISEASE DIAGNOSTICS

Aaron Don M. Africa

Department of Electronics and Communications Engineering, Gokongwei College of Engineering, De La Salle University,  
Manila, Philippines

E-Mail: [aaron.africa@dlsu.edu.ph](mailto:aaron.africa@dlsu.edu.ph)

## ABSTRACT

Heart disease is one of the leading causes of death to human beings. This disease has taken numerous lives throughout human history. Heart disease describes a range of conditions that affects the heart. This disease refers to conditions that involve blocked blood vessels that can lead to a heart attack or stroke. Heart failure caused by damage to the heart that has developed over time cannot be cured. But it can be treated to improve its symptoms. In general, the earlier that a heart disease is detected the better options are available to diagnose it. This paper presented how Rough Set theory is applied to develop a data model to aid a physician to diagnose heart disease. In particular this research will utilize the data obtained from the Hungarian database UCI Machine Learning Repository. The results of the research showed that the rough set theory successfully reduced the dimensionality of the heart disease data set by approximately 49%. Empirical testing was used to validate the rules and gave a 100% result.

**Keywords:** rough set theory, heart disease, information systems, medical diagnostics, public health engineering.

## 1. INTRODUCTION

Heart Disease describes the problems that arise within the heart muscles and the arteries that provide blood to the heart muscles [1]. This type of disease, if untreated is dangerous to the person who has it [2]. Generally, heart disease has to be treated immediately, the earlier it is detected the better it can be diagnosed [3]. The problem is heart disease can be mistaken for other organ problems. Misdiagnosis may arise because some of its symptoms are the same with other organ failures [4]. Other diseases may be mistaken from heart disease so the physician may give a wrong treatment which in turn may make the disease worse rather than curing it [5]. There will also be instances where there will be incomplete information or data. A specific number of symptoms are equivalent to a specific possible cause [6]. The problem is not all those information will be available. The patient may not know the answer to that data or simply it cannot be verified [7]. This research tries to solve the problem of incomplete information by applying Rough Set theory to heart disease data. Rough Set theory can be used to minimize the rules in the dataset so only the essential symptoms will remain. This means that the physician will only need to gather the data on the essential symptoms to make a diagnosis. This research will save time in making the correct diagnosis and prevent misdiagnosis of the disease.

## 2. HEART DISEASE

Heart disease is a general term. These are the problems that arise within the heart muscles, valves within the heart or the arteries that supply blood to the heart muscles. Determining the differences between each disease of the heart can guide the diagnostics of heart disease [8].

The most common type of heart disease and a leading cause of death is coronary artery disease. This disease affects the arteries supplying blood to the heart

muscles. These arteries narrow and harden because of the buildup of fatty or waxy cholesterol substance referred to as plaque. Atherosclerosis is the name of this plaque buildup [9]. The increase in plaque buildup causes the coronary arteries to become narrower thus blood flow becomes restricted. This causes the amount of oxygen delivered to the heart muscles to be decreased. Decreasing the amount of oxygen delivered to the heart muscles causes chest pain or angina. This symptom causes a heart attack. Coronary artery disease weakens the heart muscle over time thus contributing to heart failures and abnormal heart rhythms called arrhythmias [10].

Another type of heart disease is cardiomyopathy. This disease affects the muscle of the heart. Cardiomyopathy can be caused by a viral infection or due to genetic factors. This type of disease can be classified as either primary or secondary. Primary cardiomyopathy is attributed to a specific cause like heart valve disease, congenital heart defects or hypertension. Secondary cardiomyopathy is attributed to specific causes like the diseases affecting other organs. There are three main types of cardiomyopathy [11]. One of them is called dilated cardiomyopathy which is the stretching and enlargement of the cardiac muscle. The second is called Hypertrophic cardiomyopathy which causes the thickening of the heart muscles. The third is called restrictive cardiomyopathy which causes the ventricles of the heart to become overloaded and rigid causing difficulties in the blood flow of the ventricles between heartbeats [12].

There is a type of heart disease called Valvular heart disease. It is a disease that affects the valves of the heart. The function of the valves is to keep the blood going in the correct direction. Damage to the heart valves can be instigated by a number of conditions leading to regurgitation, prolapse or stenosis. This type of heart disease can be genetic and can be caused by certain infections like rheumatic fever or by radiation due to cancer [13].



A type of heart disease is pericardial disease. This type of heart disease is caused by the inflammation, fluid accumulation or stiffness of the pericardium. Pericardium disease can be caused by things such that those that occur after a heart attack [14].

A heart disease that developed before birth is called congenital heart disease. This disease usually involves the formation of the heart muscles, valves or chambers. One example is having ventricular septal defect or having holes in the heart that are formed before birth. This type of heart disease can be hereditary or caused by diseases by the mother while pregnant like German measles [15].

Another type of heart disease is heart failure. This disease is characterized by the heart's inability to effectively pump enough blood to the body's organs and tissues. When the body's organ does not receive enough blood it may lead to organ damage. It is important to remember that heart failure is congestive. It may result from other cardiovascular diseases or may develop over the years. It is important to diagnose heart failure as soon as possible [16].

### 3. ROUGH SET THEORY

Incomplete information for a long time has always been a problem. This challenge has bothered mathematicians for a long time [17]. It is because in reality you will not be able to obtain all the information that you need. There will be an instance that one set of information is not possible to obtain [18]. This issue has been critical for computer scientist, especially in the Artificial Intelligence field. The issue of incomplete information is also an issue in the medical field. There are instances when the physician could not obtain all the symptoms necessary for a correct diagnosis. One way to solve the problem of incomplete information is by using Rough Set Theory [19]. This theory is a formal approximation of a crisp set in terms of its lower and upper approximations of the original set [20]. This theory overlaps with many other theories like Boolean reasoning methods, evidence theory and fuzzy logic theory. This theory is based on the assumption that the user has additional data about the elements of a given set [21]. Available information is in the form of blocks and can be represented in the form of granules. Elementary concepts can be merged into compound concepts. With any merging of elementary sets a crisp set is formed [22]. Any other set that is formed with it is called a rough set. With every set  $X$ , two crisp sets can be associated with the lower and upper approximations. Sets are usually defined by employing the membership function. These membership functions can be used to determine incomplete information [23].

### 4. DATA AND RESULTS

#### 4.1 Information System of the Hungarian Institute of Cardiology database

The Hungarian database UCI Machine Learning Repository contains data sets from the Hungarian Institute of Cardiology [24]. This database contains 14 attribute

information about patients. Rough Set Data Explorer (ROSE) [25] was used to apply the Rough Set Theory Algorithm. The following are the attribute information and their explanations:

<b>A) Age:</b>	Age in years.
<b>B) Sex:</b>	Sex (1 = male; 0 = female).
<b>C) CP:</b>	Chest pain type.
<b>Value 1:</b>	Typical angina.
<b>Value 2:</b>	Atypical angina.
<b>Value 3:</b>	Non-anginal pain.
<b>Value 4:</b>	Asymptomatic.
<b>D) Trestbps:</b>	Resting blood pressure (in mm Hg on admission to the hospital).
<b>E) Chol:</b>	Serum cholesterol in mg/dl.
<b>F) FBS:</b>	(Fasting blood sugar > 120 mg/dl) (1 = true; 0 = false).
<b>G) Restecg:</b>	Resting electrocardiographic results.
<b>Value 0:</b>	Normal.
<b>Value 1:</b>	Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV).
<b>Value 2:</b>	Showing probable or definite left ventricular hypertrophy by Estes' criteria.
<b>H) Thalach:</b>	Maximum heart rate achieved.
<b>I) Exang:</b>	Exercise induced angina (1 = yes; 0 = no).
<b>J) Oldpeak=</b>	ST depression induced by exercise relative to rest.
<b>K) Slope:</b>	The slope of the peak exercise ST segment.
<b>Value 1:</b>	upsloping.
<b>Value 2:</b>	flat.
<b>Value 3:</b>	downsloping.
<b>L) CA:</b>	Number of major vessels (0-3) colored by fluoroscopy.
<b>M) HAL:</b>	3 = normal; 6 = fixed defect; 7 = reversible defect.
<b>N) ADS:</b>	(Angiographic Disease Status). Diagnosis of heart disease
<b>Value 0:</b>	no existence of ADS detected
<b>Value 1:</b>	minimal presence of ADS detected
<b>Value 2:</b>	moderate presence of ADS detected
<b>Value 3:</b>	high presence of ADS detected
<b>Value 4:</b>	critical presence of ADS detected

The predicted attribute value is the Angiographic Disease Status (ADS). A value of 0 means no existence of ADS is detected, 1 means moderate presence of ADS is detected, 3 means high presence of ADS is detected and a value of 4 means critical presence of ADS is detected. Variables with values of X means it is unknown or cannot be obtained.

Knowing the ADS will help the physician to know if a Heart disease is present so he can immediately treat it.

**Table-1.**Information system of the Hungarian Institute of Cardiology database.

#	A	B	C	D	E	F	G
1	40	1	2	140	289	0	0
2	49	0	3	160	180	0	0
3	37	1	2	130	283	0	1
4	48	0	4	138	214	0	0
5	54	1	3	150	X	0	0
6	39	1	3	120	339	0	0
7	45	0	2	130	237	0	0
8	54	1	2	110	208	0	0
9	37	1	4	140	207	0	0
10	48	0	2	120	284	0	0
11	37	0	3	130	211	0	0
12	58	1	2	136	164	0	1
13	39	1	2	120	204	0	0
14	49	1	4	140	234	0	0
15	42	0	3	115	211	0	1
16	54	0	2	120	273	0	0
17	38	1	4	110	196	0	0
18	43	0	2	120	201	0	0
19	60	1	4	100	248	0	0
20	36	1	2	120	267	0	0
21	43	0	1	100	223	0	0
22	44	1	2	120	184	0	0
23	49	0	2	124	201	0	0
24	44	1	2	150	288	0	0
25	40	1	3	130	215	0	0
26	36	1	3	130	209	0	0
27	53	1	4	124	260	0	1
28	52	1	2	120	284	0	0
29	53	0	2	113	468	X	0
30	51	1	2	125	188	0	0
31	53	1	3	145	518	0	0
32	56	1	3	130	X	0	0
33	54	1	4	125	224	0	0
34	41	1	4	130	172	0	1
35	43	0	2	150	186	0	0
36	32	1	2	125	254	0	0
37	65	1	4	140	306	1	0
38	41	0	2	110	250	0	1
39	48	0	2	120	X	1	1

40	48	0	4	150	227	0	0
41	54	0	2	150	230	0	0
42	54	0	3	130	294	0	1
43	35	1	2	150	264	0	0
44	52	1	3	140	259	0	1
45	43	1	4	120	175	0	0
46	59	1	3	130	318	0	0
47	37	1	4	120	223	0	0
48	50	1	2	140	216	0	0
49	36	1	3	112	340	0	0
50	41	1	4	110	289	0	0
51	50	1	4	130	233	0	0
52	47	0	4	120	205	0	0
53	45	1	2	140	224	1	0
54	41	0	2	130	245	0	0
55	52	0	4	130	180	0	0
56	51	0	2	160	194	0	0
57	31	1	4	120	270	0	0
58	58	1	3	130	213	0	1
59	54	1	4	150	365	0	1
60	52	1	4	112	342	0	1
61	49	1	2	100	253	0	0
62	43	0	3	150	X	0	0
63	45	1	4	140	224	0	0
64	46	1	4	120	277	0	0
65	50	0	2	110	202	0	0
66	37	0	2	120	260	0	0
67	45	0	4	132	297	0	0
68	32	1	2	110	225	0	0
69	52	1	4	160	246	0	1
70	44	1	4	150	412	0	0
71	57	1	2	140	265	0	1
72	44	1	2	130	215	0	0
73	52	1	4	120	182	0	0
74	44	0	4	120	218	0	1
75	55	1	4	140	268	0	0
76	46	1	3	150	163	X	0
77	32	1	4	118	529	0	0
78	35	0	4	140	167	0	0
79	52	1	2	140	100	0	0
80	49	1	4	130	206	0	0
81	55	1	3	110	277	0	0



82	54	1	2	120	238	0	0
83	63	1	4	150	223	0	0
84	52	1	2	160	196	0	0
85	56	1	4	150	213	1	0
86	66	1	4	140	X	0	0
87	65	1	4	170	263	1	0
88	53	0	2	140	216	0	0
89	43	1	1	120	291	0	1
90	55	1	4	140	229	0	0
91	49	0	2	110	X	0	0
92	39	1	4	130	307	0	0
93	52	0	2	120	210	0	0
94	48	1	4	160	329	0	0
95	39	0	3	110	182	0	1
96	58	1	4	130	263	0	0
97	43	1	2	142	207	0	0
98	39	1	3	160	147	1	0
99	56	1	4	120	85	0	0
100	41	1	2	125	269	0	0
101	65	1	4	130	275	0	1
102	51	1	4	130	179	0	0
103	40	0	4	150	392	0	0
104	40	1	4	120	466	X	0
105	46	1	4	118	186	0	0
106	57	1	2	140	260	1	0
107	48	0	4	120	254	0	1
108	34	1	2	150	214	0	1
#	H	I	J	K	L	M	N
1	172	0	0	X	X	X	0
2	156	0	1	2	X	X	1
3	98	0	0	X	X	X	0
4	108	1	2	2	X	X	3
5	122	0	0	X	X	X	0
6	170	0	0	X	X	X	0
7	170	0	0	X	X	X	0
8	142	0	0	X	X	X	0
9	130	1	2	2	X	X	1
10	120	0	0	X	X	X	0
11	142	0	0	X	X	X	0
12	X	1	2	2	X	X	3
13	145	0	0	X	X	X	0
14	140	1	1	2	X	X	3

15	137	0	0	X	X	X	0
16	150	0	2	2	X	X	0
17	166	0	0	X	X	X	1
18	165	0	0	X	X	X	0
19	125	0	1	2	X	X	1
20	160	0	3	2	X	X	1
21	142	160	3	2	X	X	1
22	142	0	1	2	X	X	0
23	164	0	0	X	X	X	0
24	150	1	3	2	X	X	3
25	138	0	0	X	X	X	0
26	178	0	0	X	X	X	0
27	112	1	3	2	X	X	0
28	118	0	0	X	X	X	0
29	127	0	0	X	X	X	0
30	145	0	0	X	X	X	0
31	130	0	0	X	X	X	3
32	114	0	0	X	X	X	0
33	122	0	2	2	X	X	1
34	130	0	2	2	X	X	3
35	154	0	0	X	X	X	0
36	155	0	0	X	X	X	0
37	87	1	2	2	X	X	1
38	142	0	0	X	X	X	0
39	148	0	0	X	X	X	0
40	130	1	1	2	9	X	0
41	130	0	0	X	X	X	0
42	100	1	0	2	X	X	4
43	168	0	0	X	X	X	0
44	170	0	0	X	X	X	0
45	120	1	1	2	X	7	3
46	120	1	1	2	X	3	0
47	168	0	0	X	X	3	0
48	170	0	0	X	X	3	0
49	184	0	1	2	X	3	0
50	170	0	0	X	X	6	1
51	121	1	2	2	X	7	2
52	98	1	2	2	X	6	4
53	122	0	0	X	X	X	0
54	150	0	0	X	X	X	0
55	140	1	2	2	X	X	0
56	170	0	0	X	X	X	0



57	153	1	2	2	X	X	1
58	140	0	0	X	X	6	1
59	134	0	1	1	X	X	0
60	96	1	1	2	X	X	3
61	174	0	0	X	X	X	0
62	175	0	0	X	X	3	0
63	144	0	0	X	X	X	0
64	125	1	1	2	X	X	1
65	145	0	0	X	X	X	1
66	130	0	0	X	X	X	0
67	144	0	0	X	X	X	0
68	184	0	0	X	X	X	0
69	82	1	4	2	X	X	3
70	170	0	0	X	X	X	0
71	145	1	1	2	X	X	1
72	135	0	0	X	X	X	0
73	150	0	0	X	X	X	2
74	115	0	0	X	X	X	0
75	128	1	2	2	X	X	2
76	116	0	0	X	X	X	0
77	130	0	0	X	X	X	2
78	150	0	0	X	X	X	0
79	138	1	0	X	X	X	0
80	170	0	0	X	X	X	1
81	160	0	0	X	X	X	0
82	154	0	0	X	X	X	0
83	115	0	0	X	X	X	1
84	165	0	0	X	X	X	0
85	125	1	1	2	X	X	2
86	94	1	1	2	X	9	3
87	112	1	2	2	X	X	3

88	142	1	2	2	X	X	0
89	155	0	0	X	X	X	2
90	110	1	1	2	X	X	0
91	160	0	0	X	X	X	0
92	140	0	0	X	X	X	0
93	148	0	0	X	X	X	0
94	92	1	2	2	X	X	2
95	180	0	0	X	X	X	0
96	140	1	2	2	X	X	2
97	138	0	0	X	X	X	0
98	160	0	0	X	X	X	0
99	140	0	0	X	X	X	0
100	144	0	0	X	X	X	0
101	115	1	1	2	X	X	4
102	100	0	0	X	X	7	0
103	130	0	2	2	X	6	2
104	152	1	1	2	X	6	1
105	124	0	0	X	X	7	1
106	140	0	0	X	X	6	0
107	110	0	0	X	X	X	0
108	168	0	0	X	X	X	0

Table-1 shows the Information System of the Hungarian Institute of Cardiology database. The Information System contains a total of 108 rules.

#### 4.2 Reduction of the Hungarian Institute of Cardiology database

The Rough Set Theory Algorithm, with the aid of the Rough Set Data Explorer (ROSE), was used to reduce the rules of the database. Empirical testing was used to verify its accuracy [26]. Empirical Testing was performed by comparing the rules produced by the Rough Set Theory with the original Information System to see if they matched.

**Table-2.** Rough Set rules on the Hungarian Institute of Cardiology database.

Num.	Rule	Case	Value of Symptoms in Cases
1	$(C = 2) \& (J = 0) \Rightarrow (N = 0)$	$N = 0$	$C = 2, J = 0, N = 0$
2	$(C = 3) \& (D = 130) \& (G = 0) \Rightarrow (N = 0)$	$N = 0$	$C = 3, D = 130, G = 0, N = 0$
3	$(G = 0) \& (H = 140) \& (J = 0) \Rightarrow (N = 0)$	$N = 0$	$G = 0, H = 140, J = 0, N = 0$
4	$(E = 339) \Rightarrow (N = 0)$	$N = 0$	$E = 339, N = 0$
5	$(B = 0) \& (K = X) \Rightarrow (N = 0)$	$N = 0$	$B = 0, K = X, N = 0$
6	$(E = 412) \Rightarrow (N = 0)$	$N = 0$	$E = 412, N = 0$
7	$(E = 179) \Rightarrow (N = 0)$	$N = 0$	$E = 179, N = 0$
8	$(C = 3) \& (E = 277) \Rightarrow (N = 0)$	$N = 0$	$C = 3, E = 277, N = 0$
9	$(D = 140) \& (J = 0) \Rightarrow (N = 0)$	$N = 0$	$D = 140, J = 0, N = 0$
10	$(E = 216) \Rightarrow (N = 0)$	$N = 0$	$E = 216, N = 0$
11	$(C = 3) \& (D = 150) \Rightarrow (N = 0)$	$N = 0$	$C = 3, D = 150, N = 0$
12	$(E = 273) \Rightarrow (N = 0)$	$N = 0$	$E = 273, N = 0$
13	$(E = 184) \Rightarrow (N = 0)$	$N = 0$	$E = 184, N = 0$
14	$(E = 227) \Rightarrow (N = 0)$	$N = 0$	$E = 227, N = 0$
15	$(E = 365) \Rightarrow (N = 0)$	$N = 0$	$E = 365, N = 0$
16	$(M = 3) \Rightarrow (N = 0)$	$N = 0$	$M = 3, N = 0$
17	$(E = 229) \Rightarrow (N = 0)$	$N = 0$	$E = 229, N = 0$
18	$(D = 124) \Rightarrow (N = 0)$	$N = 0$	$D = 124, N = 0$
19	$(E = 147) \Rightarrow (N = 0)$	$N = 0$	$E = 147, N = 0$
20	$(E = 180) \& (H = 140) \Rightarrow (N = 0)$	$N = 0$	$E = 180, H = 140, N = 0$
21	$(A = 31) \Rightarrow (N = 1)$	$N = 1$	$A = 31, N = 1$
22	$(A = 38) \Rightarrow (N = 1)$	$N = 1$	$A = 38, N = 1$
23	$(E = 206) \Rightarrow (N = 1)$	$N = 1$	$E = 206, N = 1$
24	$(A = 63) \Rightarrow (N = 1)$	$N = 1$	$A = 63, N = 1$
25	$(F = 0) \& (H = 125) \Rightarrow (N = 1)$	$N = 1$	$F = 0, H = 125, N = 1$
26	$(E = 306) \Rightarrow (N = 1)$	$N = 1$	$E = 306, N = 1$
27	$(H = 124) \Rightarrow (N = 1)$	$N = 1$	$H = 124, N = 1$
28	$(C = 4) \& (E = 207) \Rightarrow (N = 1)$	$N = 1$	$C = 4, E = 207, N = 1$
29	$(E = 267) \Rightarrow (N = 1)$	$N = 1$	$E = 267, N = 1$
30	$(B = 1) \& (F = 0) \& (M = 6) \Rightarrow (N = 1)$	$N = 1$	$B = 1, F = 0, M = 6, N = 1$
31	$(E = 265) \Rightarrow (N = 1)$	$N = 1$	$E = 265, N = 1$
32	$(E = 466) \Rightarrow (N = 1)$	$N = 1$	$E = 466, N = 1$
33	$(H = 156) \Rightarrow (N = 1)$	$N = 1$	$H = 156, N = 1$
34	$(D = 125) \& (H = 122) \Rightarrow (N = 1)$	$N = 1$	$D = 125, H = 122, N = 1$
35	$(E = 268) \Rightarrow (N = 2)$	$N = 2$	$E = 268, N = 2$
36	$(E = 329) \Rightarrow (N = 2)$	$N = 2$	$E = 329, N = 2$
37	$(E = 529) \Rightarrow (N = 2)$	$N = 2$	$E = 529, N = 2$
38	$(B = 1) \& (D = 130) \& (I = 1) \& (J = 2) \Rightarrow (N = 2)$	$N = 2$	$B = 1, D = 130, I = 1, J = 2, N = 2$
39	$(C = 4) \& (D = 150) \& (K = 2) \& (L = X) \Rightarrow (N = 2)$	$N = 2$	$C = 4, D = 150, K = 2, L = X, N = 2$



40	$(E = 291) \Rightarrow (N = 2)$	$N = 2$	$E = 291, N = 2$
41	$(E = 182) \& (H = 150) \Rightarrow (N = 2)$	$N = 2$	$E = 182, H = 150, N = 2$
42	$(A = 66) \Rightarrow (N = 3)$	$N = 3$	$A = 66, N = 3$
43	$(A = 52) \& (G = 1) \& (I = 1) \Rightarrow (N = 3)$	$N = 3$	$A = 52, G = 1, I = 1, N = 3$
44	$(E = 234) \Rightarrow (N = 3)$	$N = 3$	$E = 234, N = 3$
45	$(D = 170) \Rightarrow (N = 3)$	$N = 3$	$D = 170, N = 3$
46	$(D = 136) \Rightarrow (N = 3)$	$N = 3$	$D = 136, N = 3$
47	$(D = 138) \Rightarrow (N = 3)$	$N = 3$	$D = 138, N = 3$
48	$(E = 288) \Rightarrow (N = 3)$	$N = 3$	$E = 288, N = 3$
49	$(D = 145) \Rightarrow (N = 3)$	$N = 3$	$D = 145, N = 3$
50	$(E = 172) \Rightarrow (N = 3)$	$N = 3$	$E = 172, N = 3$
51	$(E = 175) \Rightarrow (N = 3)$	$N = 4$	$(E = 175) \Rightarrow (N = 3)$
52	$(D = 130) \& (G = 1) \& (I = 1) \Rightarrow (N = 4)$	$N = 4$	$(D = 130) \& (G = 1) \& (I = 1) \Rightarrow (N = 4)$
53	$(A = 47) \Rightarrow (N = 4)$	$N = 4$	$(A = 47) \Rightarrow (N = 4)$

Table-2 shows the rules on the Hungarian Institute of Cardiology database

#### 4.3 Analysis of data

The number of original rules taken from the Hungarian Database UCI Machine Learning Repository is 108. The Rough Set Theory algorithm presented was able to reduce it to 53 rules. The rules were reduced to approximately 49%. The reduced rule does not need to complete all the 13 symptoms to determine the correct value of the Angiographic Disease Status (ADS) but only the essential symptoms needed must remain. The percent validity of the system was computed using the formula  $pv = (b/a) \times 100$ . Where  $pv$  is the percent validity  $a$  and  $b$  has a value of 53 based on Tables 1 and 2. Substituting  $a$  and  $b$  to the equation, the percent validity is determined to be 100%. The results presented that Rough Set Theory can be used to determine the minimum symptoms essential to determine the condition of the ADS.

#### 5. CONCLUSIONS AND RECOMMENDATIONS

This research demonstrated that Rough Set Theory could be used to aid in diagnosing heart disease. The symptoms and possible cause from the Hungarian database UCI Machine Learning Repository were used to determine the Angiographic Disease Status (ADS). The database has 13 symptoms and 1 Possible Cause. Using Rough Set Theory, the dataset can be reduced where only the essential symptoms remain. This is valuable, especially because it is challenging to complete the symptoms in diagnosing heart disease. The patient or doctor may not know or may be unsure of what is the value of the symptom. The results were verified using the Empirical Testing and showed 100% accuracy. This research showed that ADS can still be determined even if there is incomplete information with the aid of Rough Set Theory.

For future research, it is recommended to try the methods in this research in other fields of Biomedical

Engineering. This research only tested it in the diagnostics of heart disease but in theory it can also work in other fields.

#### REFERENCES

- [1] Motimoto T., Sunagawa Y., Takaya T., Jita T., Shimatsu A., Fujita M. and Hasegawa K. 2008. Curcumin Reverses the Hypertension-induced Left Ventricular Concentric Remodeling in Rats. *Journal of Molecular and Cellular Cardiology*. 45(4): 17-18.
- [2] Hao K., Hanawa H., Ding L., Jita T., Ota Y., Yoshida K., Toba K., Ogura M., Ito H., Kodama M. and Aizawa Y. 2011. Free heme is a danger signal inducing expression of proinflammatory proteins in cultured cells derived from normal rat hearts. *Molecular Immunology*. 48(9): 1191-1292.
- [3] Rahman M., Spurier N., Mahmood M., Rahman M. H., Choudury S. and Leeder S. 2013. Rose Angina Questionnaire: Validation with cardiologists' diagnoses to detect coronary heart disease in Bangladesh. *Indian Heart Journal*. 65(1): 30-39.
- [4] Alghamadi M., Grosse-Wortmann L., Ahmad M., Mertens L. and Friedberg M. 2012. Can Simple Echocardiographic Measures Reduce the Number of Cardiac Magnetic Resonance Imaging Studies to Diagnose Right Ventricular Enlargement in Congenital Heart Disease? *Journal of the American Society of Echocardiography*? 25(5): 518-523.





- [5] Fernandes S., Pearson D., Rzesut A., Mitchell S., Landzerg M. and Martin G. 2013. Adult Congenital Heart Disease Incidence and Consultation: A Survey of General Adult Cardiologists. *Journal of the American College of Cardiology*. 61(12): 1303-1304.
- [6] Liu Q. 2012. On redundant types and Bayesian formulation of incomplete information. *Journal of Economic theory*. 144(5): 2115-2145.
- [7] Otaki Y., Takahashi H., Watanabe T., Yamaura G., Funayama A., Takanori A., Shishihido T., Miyamoto T. and Kubota I. 2015. Heart-type fatty acid binding protein and high-sensitivity troponin T are myocardial damage markers that could predict adverse clinical outcomes in patients with peripheral artery disease. *BBA Clinical*. 4(1): 35-41.
- [8] Haugen K. 2016. Essay: What is Heart Disease? <http://onlineessays.com/essays/tech/what-is-heart-disease.php>
- [9] Zhou T., Ding J., Wang X. and Zheng X. 2016. Long noncoding RNAs and atherosclerosis. *Atherosclerosis*. 248(1): 51-61.
- [10] Zachariah G., Harikrishnan S., Krishnan MN., Mohanan PP., Sanjay G., Venugopal K. and Thankappan KR. 2013. Prevalence of coronary artery disease and coronary risk factors in Kerala, South India: A population survey 0-Design and methods. *Indian Heart Journal*. 65(3): 243-249.
- [11] Peters S. 2016. Papillary muscle or leaflet rupture of the tricuspid valve in arrhythmogenic cardiomyopathy. *Indian Journal of Cardiology*. 209(1).
- [12] Spina R., Baron D., Wong S., Hungerford S. and Muller D. 2016. Severe stress (takotsubo) cardiomyopathy following subarachnoid haemorrhage. *International Journal of Cardiology*. 210(1): 119-121.
- [13] Noseworthy P., Yao X., Shah N. and Gersh B. 2016. Comparative effectiveness and safety of non-vitamin K antagonist oral anticoagulants versus warfarin in patients with atrial fibrillation and valvular heart disease. *International Journal of Cardiology*. 209(1): 181-183.
- [14] Ghormade P., Kumar N., Tingne C. and Keoliya A. 2014. Distribution and diagnostic efficacy of cardiac markers CK-MB and LDH in pericardial fluid for postmortem diagnosis of ischemic heart disease. *Journal of Forensic and Legal Medicine*. 28(1): 42-46.
- [15] Offen S., Jackson D., Canniffe C. and Celermajer D. 2016. Dextrocardia in Adults with Congenital Heart Disease. *Heart Lung and Circulation*. 25(4): 352-357.
- [16] Young L., Barnason S. and Kupzyk K. 2016. Mechanism of engaging self-management behavior in rural heart failure patients. *Applied Nursing Research*. 30(1): 222-227.
- [17] Du W. and Hu B. 2016. Dominance-based rough set approach to incomplete ordered information systems. *Information Sciences*. 346(1): 106-129.
- [18] Meng L., Xu X. and Zhaor Y. 2015. Cooperative coalition for formation flight scheduling based on incomplete information. *Chinese Journal of Aeronautics*. 28(6): 1747-1757.
- [19] Palwak Z. 1982. Rough Sets. *International Journal of Computer and Information Sciences*. 11(5): 341-356.
- [20] Srivinas K., Rao G. and Govardhan A. 2014. Rough-Fuzzy Classifier: A System to Predict the Heart Disease by blending two different set theories. *Arab Journal of Science*. 39(1): 2857-2868.
- [21] Chen Y., Zhu Q. and Xu H. 2015. Finding rough set reducts with fish swarm algorithm. *Knowledge Based Systems*. 81(1): 22-29.
- [22] Palwak Z. and Skowron A. 2007. Rudiments of Rough Sets. *Information Sciences*. 177(1): 3-27.
- [23] Africa A. and Cabatuan M. 2015. A rough set based data model for breast cancer mammographic mass diagnostics. *International Journal of Biomedical Engineering and Technology*. 18(4): 359-369.
- [24] UCI Machine Learning Repository. 2013. <http://archive.ics.uci.edu/ml/datasets.html>
- [25] ROSE 2.0. 1999. <http://www.idss.cs.put.poznan.pl/rose>
- [26] Preece A. 2015. Evaluating Verification and Validation Methods in Knowledge Engineering. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.17.3766>