International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 2 (2020), pp. 355-363
 https://dx.doi.org/10.37624/IJERT/13.2.2020.355-363

An Intelligent Mining Model for Medical Diagnosis of Heart Disease Based on Rough Set Data Analysis

Hossam A. Nabwey[1, 2]

[1]*Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj11942, Saudi Arabia.*

[2]*Department of Basic Engineering Science, Faculty of Engineering, Menoufia University, Shebin El-Kom,32511, Egypt.*

https://orcid.org/0000-0002-7167-3822

Abstract

Medical databases have accumulated large quantities of information about patients and their medical conditions. The classification of a set of objects into predefined homogenous groups is a problem with major practical interest in many fields, in particular, in medical sciences. It is well established fact that right decision at right time provides an advantage in medical diagnosis. Therefore, most important challenge is to retrieve data pattern from the accumulated voluminous data and dealing with the incomplete and vague information in classification and data analysis. Thus, the ultimate goal of this work is to present an intelligent model for mining and generating classification rules for medical diagnosis of heart disease based on rough sets theory. Rough sets with Boolean reasoning discretization algorithm is introduced to discretize the data, then the rough set reduction technique is applied to find all reducts. Finally, a set of generalized rules for heart diagnosis was extracted. The proposed model shows a higher overall accuracy rates and generate more compact rules.

Keywords: Medical diagnosis; heart disease; classifications; Rough set theory; feature selection.

1) INTRODUCTION

Medical databases have accumulated large quantities of information about patients and their medical conditions. Relationships and patterns within these data could provide new medical knowledge [1, 2]. 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 of data analysis are mainly based on quite strong assumptions (some knowledge about dependencies, probability distributions, large number of experiments), that are unable to derive conclusions from incomplete knowledge or cannot manage inconsistent pieces of information. The classification of a set of objects into predefined homogenous groups is a problem with major practical interest in many fields, in particular, in medical sciences [3].

Over the past two decades, several traditional multivariate statistical classification approaches, such as the linear discriminant analysis and the quadratic discriminant analysis, have been developed to address the classification problem. More advanced and intelligent techniques have been used in medical data analysis such as neural network, Bayesian classifier, genetic algorithms, decision trees, fuzzy theory, and rough set. Each one of these techniques has its own properties and features including their ability of finding important rules and information that could be useful for the medical field domain. Each of these techniques contributes a distinct methodology for addressing problems in its domain.

Rough set theory [4] is a fairly new intelligent technique that has been applied to the medical domain, and is used for the discovery of data dependencies, evaluates the importance of attributes, discovers the patterns of data, reduces all redundant objects and attributes, and seeks the minimum subset of attributes. Moreover, it is being used for the extraction of rules from databases. Many heuristic algorithms are proposed based on rough set theory, also numerous approached based on rough set theory and other theories are investigated to extract decision rules and reduce the dimensionality of dataset [5-18].One advantage of the rough set is the creation of readable if-then rules. Such rules have a potential to reveal new patterns in the data material. Thus, the ultimate goal of this work is to present an intelligent Model for mining and generating classification Rules for Medical Diagnosis of Heart Disease based on rough sets theory.

2) PROBLEM FORMULATION

Heart disease describes a range of conditions that affect the heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease; heart rhythm problems (arrhythmias); and heart defects you're born with (congenital heart defects), among others. In other words we can say that Heart disease is an umbrella term for any disorder that affects the structure and functions of the heart and circulation. The term "heart disease" is often used interchangeably with the term "cardiovascular disease." Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect the heart's muscle, valves or rhythm, also are considered forms of heart disease. There

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 2 (2020), pp. 355-363
 https://dx.doi.org/10.37624/IJERT/13.2.2020.355-363

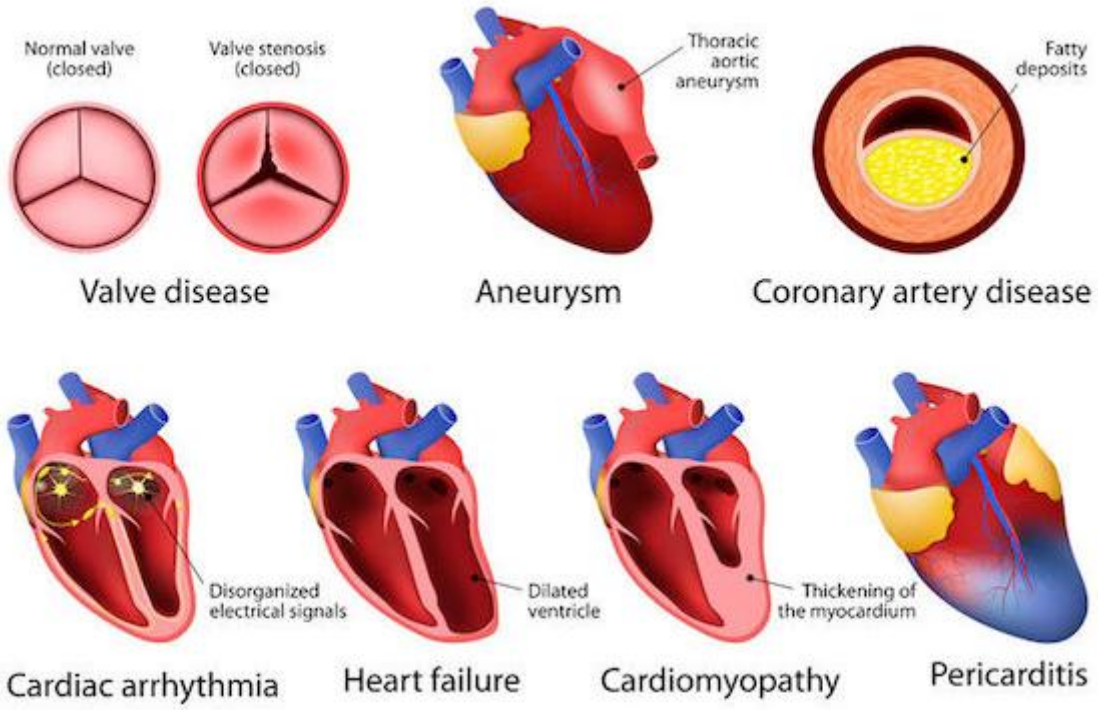

Fig. 1: Types of Heart Disease

are many types of heart disease, and each one has its own symptoms and treatment [19]. Fig. 1 shows the main types of Heart Disease.

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.

The ultimate goal of this work is to present an intelligent model for mining and generating classification rules for medical diagnosis of heart disease based on rough sets theory. Here numerical and literature values based on different symptoms were collected to the heart disease have collected from literature [20]. We consider the diagnosis decision of the patients as the decision variable. The attributes that play major role in heart disease are presented in the decision table shown in Table 1, where to write it in a simple form and to make our analysis simple we used the coding shown in table 1.

Table 1. Coding system for the Symptoms and decision attribute

Attribute	Attribute code	Attribute value	Code of the Attribute value
Chest pain (CP)	a1	Typical angina	1
		A typical angina	2
		Non-anginal pain	3
		Asymptomatic	4

Blood pressure (BP)	a2	Normal	1
		Medium	2
		High	3
		Very high	4
Cholesterol (CH)-LDL	a3	low	1
		medium	2
		high	3
		Very high	4
Fasting Blood sugar (FBS)	a4	normal	1
		high	2
Electrocardiography (ECG)	a5	normal	1
		ST-T abnormal	2
		hypertrophy	3
Maximum heart rate(MHR)	a6	medium	1
		normal	2
		high	3
Exercise(EX)	a7	false	1
		true	2
Old peak (OP)	a8	low	1
		risk	2
		temble	3
Thallium scan(TS)	a9	normal	1
		Fixed defect	2
		Reversible defect	3
Sex(SX)	a10	male	1
		female	2
Age	a11	young	1
		mild	2
		old	3
		Very old	4
Type of diagnosis (TD)	d	Hypertensive heart disease	1
		Coronary heart disease	2
		Heart failure	3
		Potential patient	4
		Cardiomyopathy	5

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 2 (2020), pp. 355-363
 https://dx.doi.org/10.37624/IJERT/13.2.2020.355-363

Table 2. Decision table for heart disease diagnosis

U		Symptoms										Type of diagnosis (d)
	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	
X1	*	4	*	*	2	*	*	2	*	1	*	1
X2	2	3	*	*	*	3	1	*	*	*	*	1
X3	3	1	1	*	3	*	*	*	*	*	*	1
X4	2	*	1	*	*	*	*	*	2	*	*	1
X5	3	*	*	*	1	*	*	*	*	*	3	1
X6	*	*	*	*	*	2	2	*	*	*	1	1
X7	*	*	*	*	1	*	2	*	*	*	4	1
X8	*	4	4	*	*	*	*	*	3	*	2	1
X9	*	4	4	*	*	*	1	3	*	*	*	1
X10	*	2	4	*	*	*	*	1	*	*	*	2
X11	*	*	1	*	*	1	*	2	3	*	*	2
X12	2	4	*	*	*	1	1	*	2	*	*	2
X13	3	*	*	*	3	1	*	*	2	*	*	2
X14	*	4	*	*	2	*	1	*	*	1	1	2
X15	2	*	3	*	*	*	*	*	3	2	3	2
X16	2	2	*	*	*	*	*	*	3	*	4	2
X17	4	4	4	*	*	*	*	*	*	*	4	2
X18	*	*	1	*	3	*	*	*	*	*	1	2
X19	*	2	3	*	*	*	*	3	*	*	*	3
X20	3	*	*	*	*	3	2	*	*	*	3	3
X21	4	1	2	*	2	*	*	*	*	*	*	3
X22	*	4	*	*	*	*	*	*	*	2	1	3
X23	3	1	*	*	1	*	*	*	*	*	*	3
X24	*	*	*	*	*	*	1	2	3	*	1	3
X25	*	3	4	*	2	*	2	*	2	*	*	3
X26	*	*	*	*	2	*	*	*	*	*	2	3
X27	*	4	*	*	*	*	*	3	2	*	4	3
X28	2	2	*	*	*	*	*	1	*	2	3	3
X29	*	*	*	*	*	*	*	*	2	*	3	3
X30	*	*	1	1	3	*	*	3	*	*	1	3
X31	4	4	*	*	*	1	*	*	*	*	*	3
X32	1	*	4	*	*	*	*	3	2	*	4	3
X33	3	*	3	*	*	*	2	*	*	*	3	3
X34	3	*	1	*	*	*	*	2	*	2	*	4
X35	4	*	*	*	*	*	*	2	*	*	*	4
X36	*	*	*	*	2	*	*	1	1	*	2	4
X37	3	*	*	*	*	*	*	*	*	*	*	4
X38	2	2	2	*	*	*	1	*	*	*	3	4
X39	3	*	4	*	*	*	*	*	1	*	1	4

X40	2	3	*	*	*	*	*	1	*	*	*	4
X41	*	3	*	*	*	*	*	*	1	*	*	4
X42	*	4	2	*	3	*	*	*	*	*	3	4
X43	2	*	*	*	3	3	1	*	*	1	*	5
X44	4	*	1	*	*	*	*	1	*	1	*	5
X45	*	2	2	*	3	*	1	*	1	*	*	5
X46	*	*	2	*	*	*	*	1	*	*	2	5
X47	4	2	*	*	*	*	*	1	*	*	2	5
X48	3	*	4	*	*	*	*	*	1	*	*	5

***** Means do not care condition

3) ANALYSIS

In this section, we will discuss the proposed rough sets scheme to analyze, mining and generating classification rules for medical diagnosis of heart disease. The scheme used in this study consists of two main stages: preprocessing and processing. Preprocessing stage includes tasks such as data cleaning, completeness, correctness, attribute creation, attribute selection and discretization. Processing includes the generation of preliminary knowledge, such as computation of object reducts from data and derivation of rules from reducts. Fig. 2 shows the overall steps in the proposed rough sets data analysis scheme.

With the aid of software called ROSETTA which is an RST analysis toolkit, rough sets with Boolean reasoning discretization algorithm is introduced to discretize the data, then the rough set reduction technique is applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification as shown un table 3. Finally, the rough sets dependency rules are generated directly from all generated reducts as shown in table 4.

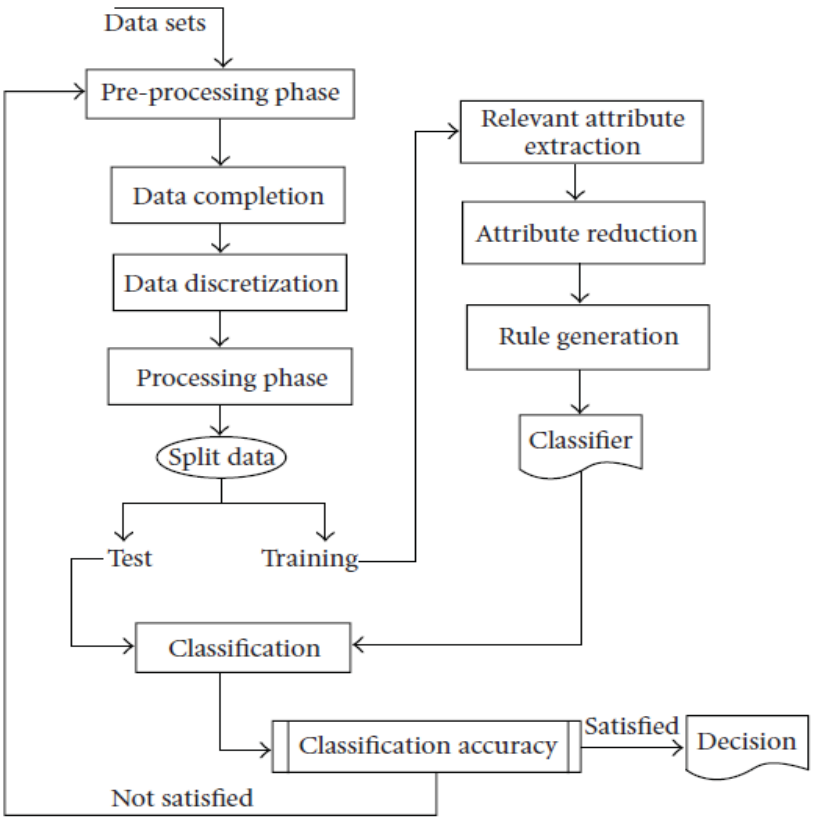

Fig. 2: the overall steps of the proposed intelligent model.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 2 (2020), pp. 355-363
 https://dx.doi.org/10.37624/IJERT/13.2.2020.355-363

Table 3. Reducts of Table 2.

	Reduct	Support	Length
1	{a2, a5, a8, a9, a11}	100	5
2	{a2, a3, a7, a8, a9, a11}	100	6
3	{a1, a5, a7, a8, a9, a11}	100	6
4	{a3, a5, a6, a7, a8, a11}	100	6
5	{a1, a2, a3, a5, a8, a11}	100	6
6	{a2, a3, a5, a7, a8, a10, a11}	100	7
7	{a2, a3, a4, a7, a9, a10, a11}	100	7
8	{a1, a3, a5, a8, a9, a10, a11}	100	7
9	{a1, a2, a3, a6, a7, a8, a11}	100	7
10	{a2, a3, a6, a8, a9, a10, a11}	100	7
11	{a1, a3, a4, a5, a6, a8, a11}	100	7
12	{a2, a3, a4, a6, a9, a10, a11}	100	7
13	{a1, a3, a5, a6, a8, a9, a11}	100	7
14	{a1, a2, a3, a4, a7, a9, a11}	100	7
15	{a3, a4, a5, a6, a8, a9, a10, a11}	100	8
16	{a1, a3, a4, a5, a6, a7, a9, a11}	100	8
17	{a1, a2, a3, a4, a5, a6, a9, a11}	100	8

Table 4. The rough sets dependency rules generated to diagnose the heart disease

	Rule	LHS Coverage	RHS Coverage
1	a2(4) AND a5(2) AND a8(2) AND a9(*) AND a11(*) => d(1)	0.020833	0.111111
2	a2(3) AND a5(*) AND a8(*) AND a9(*) AND a11(*) => d(1)	0.020833	0.111111
3	a2(1) AND a5(3) AND a8(*) AND a9(*) AND a11(*) => d(1)	0.020833	0.111111
4	a2(*) AND a5(*) AND a8(*) AND a9(2) AND a11(*) => d(1)	0.020833	0.111111
5	a2(*) AND a5(1) AND a8(*) AND a9(*) AND a11(3) => d(1)	0.020833	0.111111
6	a2(*) AND a5(*) AND a8(*) AND a9(*) AND a11(1) => d(1)	0.020833	0.111111
7	a2(*) AND a5(1) AND a8(*) AND a9(*) AND a11(4) => d(1)	0.020833	0.111111

 https://dx.doi.org/10.37624/IJERT/13.2.2020.355-363

8	a2(4) AND a5(*) AND a8(*) AND a9(3) AND a11(2) => d(1)	0.020833	0.111111
9	a2(4) AND a5(*) AND a8(3) AND a9(*) AND a11(*) => d(1)	0.020833	0.111111
10	a2(2) AND a5(*) AND a8(1) AND a9(*) AND a11(*) => d(2)	0.020833	0.111111
11	a2(*) AND a5(*) AND a8(2) AND a9(3) AND a11(*) => d(2)	0.020833	0.111111
12	a2(4) AND a5(*) AND a8(*) AND a9(2) AND a11(*) => d(2)	0.020833	0.111111
13	a2(*) AND a5(3) AND a8(*) AND a9(2) AND a11(*) => d(2)	0.020833	0.111111
14	a2(4) AND a5(2) AND a8(*) AND a9(*) AND a11(1) => d(2)	0.020833	0.111111
15	a2(*) AND a5(*) AND a8(*) AND a9(3) AND a11(3) => d(2)	0.020833	0.111111
16	a2(2) AND a5(*) AND a8(*) AND a9(3) AND a11(4) => d(2)	0.020833	0.111111
17	a2(4) AND a5(*) AND a8(*) AND a9(*) AND a11(4) => d(2)	0.020833	0.111111
18	a2(*) AND a5(3) AND a8(*) AND a9(*) AND a11(1) => d(2)	0.020833	0.111111
19	a2(2) AND a5(*) AND a8(3) AND a9(*) AND a11(*) => d(3)	0.020833	0.066667
20	a2(*) AND a5(*) AND a8(*) AND a9(*) AND a11(3) => d(3)	0.041667	0.133333
21	a2(1) AND a5(2) AND a8(*) AND a9(*) AND a11(*) => d(3)	0.020833	0.066667
22	a2(4) AND a5(*) AND a8(*) AND a9(*) AND a11(1) => d(3)	0.020833	0.066667
23	a2(1) AND a5(1) AND a8(*) AND a9(*) AND a11(*) => d(3)	0.020833	0.066667
24	a2(*) AND a5(*) AND a8(2) AND a9(3) AND a11(1) => d(3)	0.020833	0.066667
25	a2(3) AND a5(2) AND a8(*) AND a9(2) AND a11(*) => d(3)	0.020833	0.066667
26	a2(*) AND a5(2) AND a8(*) AND a9(*) AND a11(2) => d(3)	0.020833	0.066667
27	a2(4) AND a5(*) AND a8(3) AND a9(2) AND a11(4) => d(3)	0.020833	0.066667
28	a2(2) AND a5(*) AND a8(1) AND a9(*) AND a11(3) => d(3)	0.020833	0.066667
29	a2(*) AND a5(*) AND a8(*) AND a9(2) AND a11(3) => d(3)	0.020833	0.066667
30	a2(*) AND a5(3) AND a8(3) AND a9(*) AND a11(1) => d(3)	0.020833	0.066667
31	a2(4) AND a5(*) AND a8(*) AND a9(*) AND a11(*) => d(3)	0.020833	0.066667
32	a2(*) AND a5(*) AND a8(3) AND a9(2) AND a11(4) => d(3)	0.020833	0.066667
33	a2(*) AND a5(*) AND a8(2) AND a9(*) AND a11(*) => d(4)	0.041667	0.222222
34	a2(*) AND a5(2) AND a8(1) AND a9(1) AND a11(2) => d(4)	0.020833	0.111111
35	a2(*) AND a5(*) AND a8(*) AND a9(*) AND a11(*) => d(4)	0.020833	0.111111
36	a2(2) AND a5(*) AND a8(*) AND a9(*) AND a11(3) => d(4)	0.020833	0.111111

37	a2(*) AND a5(*) AND a8(*) AND a9(1) AND a11(1) => d(4)	0.020833	0.111111
38	a2(3) AND a5(*) AND a8(1) AND a9(*) AND a11(*) => d(4)	0.020833	0.111111
39	a2(3) AND a5(*) AND a8(*) AND a9(1) AND a11(*) => d(4)	0.020833	0.111111
40	a2(4) AND a5(3) AND a8(*) AND a9(*) AND a11(3) => d(4)	0.020833	0.111111
41	a2(*) AND a5(3) AND a8(*) AND a9(*) AND a11(*) => d(5)	0.020833	0.166667
42	a2(*) AND a5(*) AND a8(1) AND a9(*) AND a11(*) => d(5)	0.020833	0.166667
43	a2(2) AND a5(3) AND a8(*) AND a9(1) AND a11(*) => d(5)	0.020833	0.166667
44	a2(*) AND a5(*) AND a8(1) AND a9(*) AND a11(2) => d(5)	0.020833	0.166667
45	a2(2) AND a5(*) AND a8(1) AND a9(*) AND a11(2) => d(5)	0.020833	0.166667
46	a2(*) AND a5(*) AND a8(*) AND a9(1) AND a11(*) => d(5)	0.020833	0.166667

CONCLUSION

In this paper, an intelligent data analysis approach based on rough sets theory for mining and generating classification rules for heart disease diagnosing. Further these suitable rules are explored to identify the chief characteristics affecting the relationship between heart disease and its attributes. This helps the decision maker a priori detection of the heart disease. The obtained results are in good agreement with previous studies. The technique has been simplified logic-based rules, reduces the time and resources required to building knowledge. an extension work of using rough sets with other intelligent systems like neural networks, genetic algorithms, fuzzy approaches, and so forth, will be considered in the future work.

ACKNOWLEDGMENTS

The author thank Prince Sattam bin Abdulaziz University, Deanship of Scientific Research at Prince Sattam bin Abdulaziz University for their continuous support and encouragement.

REFERENCES

[1] Zhong, N., 2001. A rough sets based knowledge discovery process. International Journal of Applied Mathematics and Computer Science, 11(3), pp.101-117.

[2] Abdelhafez, M.E. and Own, H.S., 2008. Rough Sets Data Analysis in Knowledge Discovery: A Case of Kuwaiti Diabetic Children Patients. Advances in Fuzzy Systems, 2008.

[3] Tsumoto, S., 2000. Automated discovery of positive and negative knowledge in clinical databases. IEEE Engineering in Medicine and Biology Magazine, 19(4), pp.56-62.

[4] Pawlak, Z., 1984, December. On learning—a rough set approach. In Symposium on Computation Theory (pp. 197-227). Springer, Berlin, Heidelberg.

[5] Nabwey, Hossam A. "A Hybrid Approach for Extracting Classification Rules Based on Rough Set Methodology and Fuzzy Inference System and Its Application in Groundwater Quality Assessment." In Advances in Fuzzy Logic and Technology 2017, pp. 611-625. Springer, Cham, 2017.

[6] Nabwey, Hossam A., M. Modather, and M. Abdou. "Rough set theory based method for building knowledge for the rate of heat transfer on free convection over a vertical flat plate embedded in a porous medium." In 2015 International Conference on Computing, Communication and Security (ICCCS), pp. 1-8. IEEE, 2015.

[7] Nabwey, H.A.. An approach based on Rough Sets Theory and Grey System for Implementation of Rule-Based Control for Sustainability of Rotary Clinker Kiln. International Journal of Engineering Research and Technology, Volume 12, Number 12 (2019), pp. 2604-2610

[8] Shaaban, Shaaban M., and H. Nabwey. "A decision tree approach for steam turbine-generator fault diagnosis." International Journal of Advanced Science and Technology 51 (2013): 59-66.

[9] Shaaban, Shaaban M., and Hossam A. Nabwey. "A probabilistic rough set approach for water reservoirs site location decision making." In International Conference on Computational Science and Its Applications, pp. 358-372. Springer, Berlin, Heidelberg, 2012.

[10] Shaaban, Shaaban M., and Hossam A. Nabwey. "Rehabilitation and reconstruction of asphalts

pavement decision making based on rough set theory." In International Conference on Computational Science and Its Applications, pp. 316-330. Springer, Berlin, Heidelberg, 2012.

[11] Shaaban, M., and A. Nabwey. "Transformer fault diagnosis method based on rough set and generalized distribution table." Int J IntellEngSyst 5 (2012): 17-24.

[12] Mohamed, HossamAbdElmaksoud. "An Algorithm for Mining Decision Rules Based on Decision Network and Rough Set Theory." In International Conference on Ubiquitous Computing and Multimedia Applications, pp. 44-54. Springer, Berlin, Heidelberg, 2011.

[13] Zhao, Hong, Ping Wang, Qinghua Hu, and Pengfei Zhu. "Fuzzy Rough Set Based Feature Selection for Large-Scale Hierarchical Classification." IEEE Transactions on Fuzzy Systems 27, no. 10 (2019): 1891-1903.

[14] Nabwey, Hossam A., and Mahdy S. El-Paoumy. "An integrated methodology of rough set theory and grey system for extracting decision rules." International Journal of Hybrid Information Technology 6, no. 1 (2013): 57-65.

[15] Pathak, H.K., George, R., Nabwey, H.A., El-Paoumy, M.S. and Reshma, K.P., 2015. Some generalized fixed point results in ab-metric space and application to matrix equations. Fixed Point Theory and Applications, 2015(1), pp.1-17.

[16] George, R., Nabwey, H.A., Reshma, K.P. and Rajagopalan, R., 2015. Generalized cone b-metric spaces and contraction principles. Mat. Vesn, 67(4), pp.246-257.

[17] Nabwey, H.A., Boumazgour, M. and Rashad, A.M., 2017. Group method analysis of mixed convection stagnation-point flow of non-Newtonian nanofluid over a vertical stretching surface. Indian Journal of Physics, 91(7), pp.731-742.

[18] Hvidsten, T.R. and Komorowski, J., 2007. Rough sets in bioinformatics. In Transactions on rough sets VII (pp. 225-243). Springer, Berlin, Heidelberg.

[19] Mackay, J. and Mensah, G.A., 2004. The atlas of heart disease and stroke. World Health Organization.

[20] Tripathy, B.K., Acharjya, D.P. and Cynthya, V., 2013. A framework for intelligent medical diagnosis using rough set with formal concept analysis. arXiv preprint arXiv:1301.6011.