Analysis of a Dobutamine Stress Echocardiography Dataset Using Rough Sets

Kenneth R. Revett

University of Westminster, Harrow School of Computer Science, London, England HA1 3TP

Abstract. Stress echocardiography is an important functional diagnosis and prognostic tool that is now routinely applied to evaluate the risk of cardiovascular artery disease (CAD). A complete dataset containing data on 558 subjects undergoing a prospective longitudinal study is employed to investigate what attributes correlate with the final outcome. The dataset was examined using rough sets, which resulted in a series of decision rules that predict which attributes influence the outcomes measured clinically and recorded in the dataset. The results indicate that the ECG attribute was very informative. In addition, prehistory information has a significant impact on the classification accuracy.

1 Introduction

Heart disease remains the number one cause of mortality in the western world. Coronary arterial disease (CAD) is a primary cause of morbidity and mortality in patients with heart disease. The early detection of CAD was in part made possible in the late 1970s by the introduction of echocardiography a technique for measuring the physical properties of the heart using a variety of imaging techniques such as ultrasound, and doppler flow measurements. The purpose of these imaging studies is to identify structural malformations such as aneurysms and valvular deformities. Although useful, structural information may not provide the full clinical picture in the way that functional imaging techniques such as stress echocardiography (SE) may. This imaging technique is a versatile tool that allows clinicians to diagnosis patients with CAD efficiently and accurately. In addition, it provides information concerning the prognosis of the patient which can be used to provide on-going clinical support to help reduce morbidity.

The underlying basis for SE is the induction of cardiovascular stress, which generates ischemia, resulting in wall motion abnormality (WMA) distal to the coronary lesion. In addition to detecting CAD, the technique is also routinely employed to measure the extent of valvular heart disease. Normally, the walls of the heart (in particular the left ventrical) change (move) in a typical fashion in response to stress (i.e. heavy exercise). A quantitative measure called the wall motion score is computed and its magnitude is directly related to the extent of the WMA score. The WMA provides a quantitative measure of how the heart responds to stress. Stress echocardiography (SE) was originally induced under conditions of strenuous exercise such as bike and treadmills. In

M. Kryszkiewicz et al. (Eds.): RSEISP 2007, LNAI 4585, pp. 756–763, 2007.

many cases though, patients are not able to exercise to the level required and pharmacological agents such as dobutamine or dipyridamole have been used to induce approximately the same level of stress on the heart as physical exercise. Dobutamine in particular emulates physical exercise effects on the cardiovascular system by increasing the heart rate and blood pressure and impacts cardiac contractility which drives cardiac oxygen demand [1]. A number of reports have indicated that though there are subtle differences between exercise and pharmacologically induced stress, they essentially provide the same stimulus to the heart and can therefore, in general, be used interchangeably [2], [3].

In this paper, we investigate the effectiveness of dobutamine stress echocardiography (DSE) by analysing the results of a large study of 558 patients undergoing DSE. The purpose is to determine which attributes collected in this study correlate most closely with the decision outcome. After a careful investigation of this dataset, a set of rules is presented that relates condition attributes to decision outcomes. This rule set is generated through the application of rough sets, a data mining technique developed by the late Professor Pawlak [4]. In the next section, we present an overview of the dataset, followed by a description of the pre-processing stages.

1.1 The Dataset

The data employed in this study was obtained from a prospective dobutamine stress echocardiography (DSE) study at the UCLA Adult Cardiac Imaging and Hemodynamics Laboratory held between 1991 and 1996. The patients were monitored during a five year period and then observed for a further twelve months to determine if the DSE results could predict patient outcome. The outcomes were categorised into the following cardiac events: cardiac death, myocardial infarction (MI), and revascularisation by percutaneous transluminal coronary angioplasty (PTCA) or coronary artery bypass graft surgery (CABG) [2], [3]. After normal exclusionary processes, the patient cohort consisted of 558 subjects (220 women and 338 men) with a median age of 67 (range 26-93). Dobutamine was administered intraveneously using a standard delivery system yielding a maximum dose of 40 g/kg/min. There were a total of 30 attributes collected in this study which are listed in Table 1. The attributes were a mixture of categorical and continuous values. The decision class used to evaluate this dataset was the outcomes as listed as listed above and in Table 1. As a preliminary evaluation of the dataset, the data was evaluated with respect to each of the four possible measured outcomes included in the decision table individually, excluding each of the other three possible outcomes. This process was repeated for each of the outcomes in the decision table. Next, the effect of the echocardiogram (ECG) was investigated. Reports indicate that this is a very informative attribute with respect to predicting the clinical outcome of a patient [3]. To evaluate the effect of ECG on the outcomes, the base case investigation (all four possible outcomes) was investigated with (base case) and without the ECG attribute. Lastly, we investigated whether any prehistory information would provide a correlation between the DSE and the outcome. There were a total of six different history

Table 1. The decision table attributes and their data types (continuous, ordinal, or discrete) employed in this study (see [2] for details). Note the range of correlation coefficients was 0.013 to 0.2476 (specific data not shown).

Attributename	*Attributetype*
bhr basal heart rate	Integer
basebp basal blood pressure	Integer
basedp basal double product (= bhr x basebp)	Integer
pkhr peak heart rate	Integer
sbp systolic blood pressure	Integer
dp double product (= pkhr x sbp)	Integer
dose dose of dobutamine given	Integer
maxhr maximum heart rate	Integer
mphr(b) % of maximum predicted heart rate	Integer
mbp maximum blood pressure	Integer
dpmaxdo double product on maximum dobutamine dose	Integer
dobdose dobutamine dose at which maximum double product	Integer
age	Integer
gender (male = 0)	Level (2)
baseef baseline cardiac ejection fraction	Integer
dobef ejection fraction on dobutamine	Integer
chestpain (0 experienced chest pain)	Integer
posecg signs of heart attack on ecg (0 = yes)	Integer
equivecg ecg is equivocal (0 = yes)	Integer
restwma wall motion anamoly on echocardiogram (0 = yes)	Integer
posse stress echocardiogram was positive (0 = yes)	Integer
newMI new myocardial infarction, or heart attack (0 = yes)	Integer
newPTCA recent angioplasty (0 = yes)	Level (2)
newCABG recent bypass surgery (0 = yes)	Level (2)
death died (0 = yes)	Level (2)
hxofht history of hypertension (0 = yes)	Level (2)
hxofptca history of angioplasty (0 = yes)	Level (2)
hxofcabg history of bypass surgery (0 = yes)	Level (2)
hxofdm history of diabetes (0 = yes)	Level (2)
hxofMI history of heart attack (0 = yes)	Level (2)

attributes (see Table 1) that were tested to determine if each in isolation had a positive correlation with the outcomes. In the next section, we describe the experiments that were performed using rough sets (RSES 2.2.1).

2 Results

In the first experiment, each outcome was used as the sole decision attribute. The four outcomes were: new Myocardial Infarction (MI) (28 cases), death (24 cases), newPTCA (27 cases), and newCABG (33 cases). All continuous attributes were discretised using the MDL algorithm within Rosetta ([9]). Note there were no missing values in the dataset. A 10-fold cross validation was performed using

Table 2. Confusion matrices for the base cases of the four different outcomes. The label A corresponds to death, B to MI, C to new PTCA, and D to newCABG. Note the overall accuracy is placed at the lower right hand corner of each subtable (large bold).

A	0	1		B	0	1	
0	204	7	0.97	0	205	4	0.98
1	2	10	0.80	1	0	14	1.0
	0.95	0.22	0.92		0.94	0	0.92
C	0	1		D	0	1	
0	207	9	0.96	0	191	25	0.88
1	1	6	0.14	1	0	7	1.0
	0.97	0.10	0.93		0.96	0.0	0.93

decision rules and dynamic reducts. The results reported here are the average values from 10 executions under identical conditions. Without any filtering of the reducts or rules, Table 2 presents randomly selected confusion matrices that were generated for each of the decision outcomes for the base case. The number of rules was quite large and initially no filtering was performed to reduce either the number of reducts nor the number of rules. The number of reducts for panels A D in Table 2 were: 104, 159, 245, and 122 respectively. On average, the length of the reducts ranged from 5-9, out of a total of 27 attributes (minus the 3 other outcome decision classes). The number of rules (all of which were deterministic) was quite large, with a range of 23,356-45,330 for the cases listed in table 2. Filtering was performed on both reducts (based on support) and rule coverage in order to reduce the cardinality of the decision rules. The resulting decision rule set were reduced to a range of 314-1,197. The corresponding accuracy was reduced by approximately 4% (range 3- 6%). Filtering can be performed on a variety of conditions, such as LHS support, coverage, RHS support. For a discussion of rule filtering, please consult [5], [6], [8] for a comprehensive discussions of this topic.

The number of rules was quite large and initially no filtering was performed to reduce either the number of reducts nor the number of rules. The number of reducts for panels A D in Table 2 were: 104, 159, 245, and 122 respectively. On average, the length of the reducts ranged from 5-9, out of a total of 27 attributes (minus the 3 other outcome decision classes). The number of rules (all of which were deterministic) was quite large, with a range of 23,356-45,330 for the cases listed in table 2. Filtering was performed on both reducts (based on support) and rule coverage in order to reduce the cardinality of the decision rules. This technique has been employed successfully in similar types of biomedical datasets (see [10], [11]. The resulting decision rule set were reduced to a range of 314-1,197. The corresponding accuracy was reduced by approximately 4% (range 3-6%).

In the next experiment, the correlation between the outcome and the ECG result was examined. It has been reported that the ECG, which is a standard cardiological test to measure functional activity of the heart, should be corre-

Table 3. Confusion matrices for the base cases without the inclusion of the ECG attribute for the four different outcomes (as in Table 2). The label A corresponds to death, B to MI, C to new PTCA, and D to newCABG. Note the overall accuracy is placed at the lower right hand corner of each subtable (large bold).

A	0	1		B	0	1	
0	206	5	0.98	0	209	0	1.0
1	2	10	0.80	1	0	14	1.0
	0.95	0.22	0.92		0.94	0	0.92
C	0	1		D	0	1	
0	207	9	0.96	0	191	25	0.91
1	1	6	0.86	1	0	7	1.0
	0.97	0.10	0.93		0.96	0.0	0.93

lated with the outcome [2]. We therefore repeated the experiment in Table 2, with the ECG attribute excluded (masked) from the decision table. The results are reported in Table 3. Lastly, we examined the effect of historical information that was collected and incorporated into the dataset (see Table 1). These historical attributes include: history of hypertension, diabetes, smoking, myocardial infarction, angioplasty, and coronary artery bypass surgery. We repeated the base set of experiments (including ECG) and withheld each of the historical attributes one at a time and report the results as a set of classification accuracies, listed in Table 4.

Table 4. The classification accuracy obtained from the classification using the exact same protocol for the table reported in Table 2 (note the ECG attribute was included in the decision table). The results are the average over the four different outcomes.

Attributename	*Classificationaccuracy*
History of hypertension	91.1%
History of diabetes	85.3%
History of smoking	86.3%
History of angioplasty	90.3%
History of coronary artery bypass surgery	82.7%

In addition to classification accuracy, rough sets provides a collection of decision rules in conjunctive normal form. These rules contain the attributes and their values that are antecedents in a rule base. Therefore, the decision rules provide a codification of the knowledge contained within the decision table. Examples of the resulting rule set for the base case, using MI as the decision attribute is presented in table 5.

Lastly, to further validate and compare the accuracy of the classification element of this study, two standard neural networks (radial basis function and feed forward multi-layer) were applied to this dataset. With both neural network validation experiments, the inputs were the discretised version employed

Table 5. Sample set of rules from the base case (+ ECG) with death as the outcome. The right hand column lists the support (LHS) for the corresponding rule. These rules were selected randomly from the rule set.

Decisionrule	*Support*
dp([20716, *)) AND dobdose(40) AND hxofDM(0) AND anyevent(0) = death(0)	19
dp([*, 13105)) AND dobdose(40) AND hxofDM(0) AND anyevent(0) = death(0)	18
basebp([*, 159)) AND sbp([115, 161)) AND dose(40) AND dobdose(40) AND dobEF([61, 71)) AND hxofDM(0) = death(0)	24
dp([*, 13105)) AND dobdose(35) AND dobEF([53, 61)) AND hxofDM(1) = death(1)	14
dp([20633, 20716)) AND dobdose(40) AND baseEF([56, 76)) AND hxofDM(0) AND anyevent(1) = death(1)	9
dp([*, 13105)) AND dobdose(30) AND hxofCABG(0) AND anyevent(1) AND ecg([*, 2)) = death(1)	12

in the rough sets analysis and the outputs were the aforementioned four decision classes. For the feed forward multi-layered network, training was applied using a batch mode back propagation algorithm, with a momentum value - 0.1 and a learning rate parameter = 0.2. The data was divided 70/30 (training/testing) and the error was measured over 10 trials and the results averaged. The classification accuracy for this neural network was 86.8%. The radial basis function network employed the same input/outputs and produced an overall accuracy (after 10 trials, 70/30 training/testing) of 89.9%.

3 Conclusion

This dataset contained a complete set of attributes (30) that was a mixture of continuous and categorical data. The data was obtained from a prospective study of cardiovascular health obtained by professional medical personal (cardiographers). The attributes were obtained from patients undergoing stress echocardiography, a routine medical technique employed to diagnose cardiovascular artery disease. From the initial classification results, the specificity of the classification using rough sets was quite high (90+%) consistent with some literature reports [2]. The accuracy produced by rough sets was higher than that generated using neural networks such as multi-layer perceoptrons and a radial basis function. As can be seen in Table 2, the sensitivity of the test was quite low, resulting in a reduced classification accuracy. The effect of ECG, the attribute most correlated with the clinical outcome of CAD, was measured by masking this attribute. The results indicate that this attribute did not have a significant impact on the overall classification accuracy, but did manage to increase the sensitivity when it was excluded from the decision table. This is an interesting result that may require specific medical knowledge in order to interpret. The effect of patient history was examined, and the results (see Table 4) indicate that in general, relevant medical history did have a positive impact on the classification accuracy. This result was quantified by examining the classification accuracy when these 5 history factors were removed from the decision table (one at a time). The effect of their combination was not examined in this paper, which is left for future work. Lastly, the rule set that was produced yielded a consistently reduced set of attributes ranging from 4-9 attributes, greatly reducing the size of the dataset. As displayed in Table 5 - and generally across the rule set, the dp and dobdose attributes appear consistently (has a large support) within all decision outcomes (data not displayed). This type of analysis is a major product of the rough sets approach to data analysis extraction of knowledge from data.

This is a preliminary study that will be pursued in conjunction with a qualified cardiologist. The results generated so far are interesting and certainly consistent and in many cases superior to other studies [1],[3]. To this authors knowledge, this is the first report which examined the dobutamine SE literature using rough sets. Komorowski & Ohn have examined a similar dataset but the imaging technique and attributes selected were different from those used in the study investigated in this work [7]. It is hoped that close collaboration between medical experts and data mining engineers will provide the conditions necessary for a full extraction of knowledge from the data.

Acknowledgements

The author would like to acknowledge the source of the dataset used in this study:

http://www.stat.ucla.edu:16080/projects/datasets /cardiacexplanation.html

References

1. Armstrong, W.F.: Stress Echocardiography: Current Methodology and Clinical Applications. J. Am. Coll. Cardiology 45, 1739–1747 (2005)
2. Krivokapich, J., Child, J.S., Walter, D.O., Garfinkel, A.: Prognostic value of dobutamine stress echocardiography in predicting cardiac events in patients with known or suspected coronary artery disease. J. Am. Coll. Cardiology 33, 708–716 (1999)
3. Bergeron, S., Hillis, G., Haugen, E., Oh, J., Bailey, K., Pellikka, P.: Prognostic value of dobutamine stress echocardiography in patients with chronic kidney disease. American Heart Journal 153(3), 385–391 (2005)
4. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
5. Bazan, J., Szczuka, M.: The Rough Set Exploration System. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets III. LNCS, vol. 3400, pp. 37–56. Springer, Heidelberg (2005)
6. Komorowski, J., Pawlak, Z., Polkowski, L., Skowron, A.: Rough sets: A tutorial. In: Rough Fuzzy Hybridization – A New Trend in Decision Making, pp. 3–98. Springer, Heidelberg (1999)
7. Komorowski, J., Orhn, A.: Modelling prognostic power of cardiac tests using rough sets. Artificial Intelligence in Medicine 15, 167–191 (1999)

8. Slezak, D.: Approximate Entropy Reducts. Fundamenta Informaticae (2002)
9. Rosetta: Rosetta: `http://www.idi.ntnu.no/~aleks/rosetta`
10. Revett, K., Gorunescu, F., Gorunescu, M., El-Darzi, E., Gorunescu, S.: A Hybrid Breast Cancer Diagnosis System: A combined Approach Using Rough Sets and Probabilistic Neural Networks. In: Eurocon 2005, IEE Conference, Belgrade, Serbia and Montenegro, November 22-24, 2005, pp. 1124–1127 (2005)
11. Revett, K., Khan, A.: Datamining a Hepatitis dataset Using Rough Sets, The 5th International Conference on Artificial Intelligence and Digital Communications, Research Notes in Artificial Intelligence and Digital Communications (AIDC) 2005, Craiova, Romania, pp. 12–20 (2005)