AKAMAS: Mining Association Rules Using a New Algorithm for the Assessment of the Risk of Coronary Heart Events

M. Karaolis, *Student Member, IEEE*, J.A. Moutiris, *FESC*, L. Papaconstantinou, C.S. Pattichis, *Senior Member, IEEE*

Abstract – Many algorithms have been developed for rule mining in large transaction databases. Discovery of some important association rules is a main database mining problem. The objective of this study was to develop a new data mining algorithm named AKAMAS using different measures to extract the most important association rules for the assessment of heart event related risk factors. The implemented measures were: support, confidence, p-value, chi square, coverage, prevalence, recall, specificity, accuracy, lift, leverage, added value, relative risk, odds ratio, and conviction. The AKAMAS algorithm is a variant of the Apriori algorithm, the main difference is that it does not use the iterative technique of k-itemset to build the (k + 1)-itemsets. It needs only one pass for extracting frequent itemsets. Although AKAMAS gave similar rules to Apriori it offers a wide selection of measures for filtering the best rules, including the computation of the chi square test, and its associated probability value (i.e. if a rule is statistically significant or not). Moreover, the rules are more comprehensively represented and are more easily to interpret.

Index Terms— *Data mining association rules, risk coronary heart events*

I. INTRODUCTION

One of the most important data mining problems is mining association rules. An association rule is a rule which implies certain association relationships among a set of objects (such as ‘occur together’ or ‘one implies the other’) in a database [1]. There are two steps to the data mining task for extracting association rules: i. finding all frequent itemsets, i.e. itemsets that occur in the database with a certain user-specified frequency, called minimum support, and ii. forming the rules among the frequent itemsets [2]. Given a set of transactions D , the problem of mining association rules is to generate all association rules that have support and confidence greater than the minimum support and minimum confidence specified by the user. If the user gives a small percentage for minimum support and minimum confidence, the result is a huge number of association rules. The problem is to find from this set of rules the best or better the strongest rules. On the other hand, if the user gives a high percentage of minimum support and confidence, most probably some important rules will not be extracted. There is also the case, that most of the rules have the same support or confidence. In this case, it is very difficult to decide which rules are important.

M. Karaolis, L. Papaconstantinou, and C. Pattichis, are with the Department of Computer Science, University of Cyprus, Nicosia, Cyprus (e-mail: karaolis@acm.org; pattichi@ucy.ac.cy).

J.A. Moutiris, is a cardiologist at the Department of Cardiology, Paphos General Hospital, Paphos, Cyprus and coordinator of the Paphos CHD Survey (email: moutiris@ucy.ac.cy).

The objective of this study was to develop a new data mining algorithm named AKAMAS for mining association rules, using more measures to filter out the best rules according to the users’ selection and needs. The results of the AKAMAS algorithm were compared with the Apriori algorithm (as implemented in the WEKA tool). Moreover, a database with cardiovascular disease patients was used for evaluating the AKAMAS algorithm.

Many studies were carried out examining the efficient mining of association rules and investigating Cardiovascular Heart Diseases and related risk factors. Data mining was employed in some of these studies, where different algorithms were used for rule extraction and evaluation like the Apriori, DHP, and variations of the Apriori.

The influential association rule mining algorithm, Apriori [3], has been developed for rule mining in large transaction databases. Another algorithm is the DHP [4], which is an extension of Apriori, that uses the hashing technique. Savasere *et al.* [5] described an efficient and fast algorithm for discovering association rules in large databases. Toivonen [6], [7], presented new algorithms that reduce the several passes over the database. The idea is to draw a random sample, to find using this sample all association rules that probably hold in the whole database, and then to verify the results with the rest of the database.

The rest of the paper is organized as follows. Section II describes the AKAMAS algorithm, Section III the Material and Methods, Section IV the Results and Discussion, and Section V the Conclusions.

II. AKAMAS ASSOCIATION RULE EXTRACTION ALGORITHM

AKAMAS is a variant of the Apriori algorithm, the difference being that it does not use the iterative technique of k-itemset to build the (k + 1)-itemsets. The advantage of AKAMAS is that it doesn’t need many passes through the database and is similar in this respect to the FP-Growth algorithm. The FP-Growth algorithm [8], [9] for mining the FP-Tree structure is a recursive procedure during which many sub FP-Trees and header tables are created. The process begins by examining each item in the header table, starting with the least frequent. For each entry, the support value for the item is produced by following the links connecting all occurrences of the current item in the FP-Tree. If the item is adequately supported, then for each leaf node a set of ancestor is produced. If the set of ancestor labels is not null a new tree is generated with the set of ancestor labels as the dataset, and the process is repeated.

Input:

- D, transactional database.
- min_supp, minimum support.

Output: Association_Rules, all strong rules that satisfy the minimum support.

Method:

```

(1) L1 = find_frequent_1-itemsets(D);
(2) for(k = 2; Association_rulesk-1 ≠ 0; k++) {
(3)   for each itemset l1 ∈ L1 {
(4)     new_rule = l1
(5)     for each itemset l2 ∈ L1 && l2 < l1 {
(6)       new_rule = new_rule ∪ l2
(7)       if (new_rule.length = k) {
(8)         if(support(new_rule) ≥ min_supp
(9)         Association_rulek = Association_rulek ∪ new_rule
(10)        new_rule = new_rule - l2
(11)      }
(12)    }
(13)  }
(14) return Association_rules = Uk
Association_rulesk;

```

Fig. 1. The AKAMAS rule association algorithm

Input:

- R, Association_rules.
- thresh_measure, measure.

Output: Filtered_Association_Rules, all rules that satisfied the measure.

Method:

```

(1) for (each rule rk ∈ R; k++) {
(2)   if (measure(rulek) ≥ thresh_measure {
(3)     filtered_Association_rulek =
      filtered_Association_rulek ∪ rulek
(4)   }
(5) return filtered_Association_rules = Uk
Filtered_Association_Rulesk;

```

Fig. 2. AKAMAS rule filtering

The description of the pseudocode of the AKAMAS algorithm is as follows (see Fig. 1):

1. Initially the algorithm accepts a database of transactions. It also receives the minimum threshold of support from the user.
2. In the first iteration all the identified frequent itemsets with one feature are kept in L₁.
3. Then, for each iteration k (step 2), where k is the number of features of the rule, an iteration is carried out until there are no rules left with (k-1)-itemsets that satisfy the minimum support.
4. Each itemset that belongs to L₁ (step 3) will be associated with the k-1 itemsets that also belongs to L₁, in order to create a rule with k-itemsets. A new rule is created. Then every itemset L₂ that belongs to L₁, and L₂ is smaller than L₁, is added to the rule set (step 6) in order to make sure that similar rules will not be duplicated (step 5). The algorithm checks if the rule has k-itemsets (step 6) and if not, then it will continue with another L₂, otherwise it will

check whether the rule satisfies the minimum support (step 8) and it will be added to the set of rules. Then L₂ will be deducted from the rule in order to create other possible combinations (step 10).

5. Finally, the algorithm will return the association rules (step 14).

The description of the pseudocode of the AKAMAS filtering algorithm is as follows (see Fig. 2):

1. Initially, the algorithm accepts the association rules. It also receives the minimum threshold of the measure from the user.
2. Then, it checks if the rule satisfies the minimum threshold of the measure.
3. If the above is true, it will be added to the set of rules.
4. Finally, the algorithm will return the filtered association rules in tabular format.

III. MATERIAL AND METHODS

A. Data Collection

Data from 1200 consecutive CHD patients were collected, between the years 2003 – 2006 (300 patients each year) according to a pre-specified protocol, at the Paphos General Hospital of Cyprus. Patients had at least one of the following criteria on enrollment history of: myocardial infarction (MI), percutaneous coronary intervention (PCI), or coronary artery bypass graft surgery (CABG).

TABLE I
RISK FACTORS WITH THEIR CORRESPONDING CODINGS

	Risk Factor	Code 1	Code 2	Code 3	Code 4
Clinical factors					
1	Age (AGE)	1: 34-50	2: 51-60	3: 61-70	4: 71-85
2	Sex (SEX)	M: MALE	F: FEMALE		
3	Smoking Before (SMBEF)	Y: YES	N: NO		
4	Systolic Blood Pressure, (SBP)*	L<90	N: 90-120	H>120	
5	Diastolic Blood Pressure, (DBP)*	L<60	N: 60-80	H>80	
6	Family History (FH)	Y: YES	N: NO		
7	Hypertension (HT)	Y: YES	N: NO		
8	Diabetes (DM)	Y: YES	N: NO		
Biochemical factors					
9	Total cholesterol, (TC)**	L<200	N: 201–240	H>240	
10	High Density Lipoprotein, (HDL)**				
	Women	L<50	M: 50-60	H>60	
	Men	L<40	M: 40-60	H>60	
11	Low Density Lipoprotein, (LDL)**	N<130	H: 131-160	D>160	
12	Triglycerides, (TG)**	N<150	H: 151-200	D>200	
13	Glucose, (GLU)**	H>110	N<110		

L: Low, N: Normal, H: High, D: Dangerous

* in mmHg ** in mg/dL

B. Data Cleaning and Coding

The collected data were used to create a structured database system. The risk factors collected with their corresponding codings are given in Table I. The criteria for data coding were provided by the participating cardiologist and are as coded by the American and European Heart Disease Associations [10].

After data cleaning the number of cases was reduced to 369, mainly due to the unavailability of biochemical results. The number for MI were 265, 160 PCI, and 152 CABG cases respectively.

C. Pattern Evaluation and Knowledge Representation

The following three different sets of runs for association analysis were investigated:

- (i) MI versus PCI or CABG,
- (ii) (ii) PCI versus MI or CABG, and
- (iii) (iii) CABG versus MI or PCI.

For each of these runs, the steps for data mining association and pattern evaluation were carried out. Rules were extracted from different combinations of risk factors.

More specifically, selected rules were evaluated according to the importance of each rule. Each extracted rule was further evaluated by inspection of the number of cases from within the database that support the rule. Rules with a small number of records were ignored. Furthermore, other measures were used when we had rules with similar support and/or confidence.

The hold-out validation was used for evaluating the performance of the proposed runs. The data were split into training and testing partitions representing 70% and 30% of the records respectively. This procedure was repeated five times. It is noted that the extracted rules derived from the different sets were very similar.

The same procedure of runs was carried out for both algorithms, AKAMAS and Apriori. In order to have the same data to compare and to evaluate the performance of our results, only the support and confidence measures were used for the comparison both algorithms. The Apriori algorithm was run using the WEKA tool [11].

D. Performance Measures

The following measures were investigated [11], [12]:

Support: is the number of cases for which the rule applies (or predicts correctly); i.e. if we have the rule $X \& Y \rightarrow Z$, Support is the probability that a record contains $\{X, Y, Z\}$ [13].

Confidence: is the number of cases for which the rule applies (or predicts correctly), expressed as a percentage of all instances to which it applies (i.e. if we have the rule $X \& Y \rightarrow Z$, Confidence is the conditional probability that a record having $\{X, Y\}$ also contains Z [13].

Coverage: shows the percentage of the observations that satisfy the left part of the rule (A).

$$\text{Coverage} = P(A)$$

Prevalence: shows the percentage of the observations that satisfy the right part of the rule, which is the class output value.

$$\text{Prevalence} = P(B)$$

Recall: is the probability that the left part of the rule is true, if the right part of the rule is satisfied.

$$\text{Recall} = P(A|B) = \frac{P(AB)}{P(B)}$$

Specificity: is the probability that the right part of the rule is false, if the left part of the rule is not satisfied.

$$\text{Specificity} = P(\neg B | \neg A) = \frac{P(\neg A \neg B)}{P(\neg A)}$$

Accuracy: is the addition of the probability that the rule is true and the probability that the rule is false. If the value equals to one (1), then the left part of the rule and the right part of the rule are dependent. Else they are independent.

$$\text{Accuracy} = P(AB) + P(\neg A \neg B)$$

Lift: equals to the confidence divided by the percentage of all the observations of the rule. The measure is independent from coverage. The values vary from 0 to infinity (∞). When the value is close to one (1), then the two parts of the rule are independent so the rule is not interesting. When the value equals to zero (0), then the rule is not important. Else if the $P(B|A)$ is one (1), then the rule is interesting.

$$\text{Lift} = \frac{P(B|A)}{P(B)} = \frac{P(AB)}{P(A)P(B)}$$

Leverage: This measure is an important measure of the rule, because it involves the confidence and the coverage. It is the percentage of the observations that cover the left part of the rule and the right part, over the result that is expected if the two parts are independent. The values vary from -1 to 1. Values close to zero (0) show a powerful independence between the two parts of the rule. Values close to one (1), point out that the rule is important.

$$\text{Leverage} = P(B|A) - P(A)P(B) = \frac{P(AB)}{P(A)} - P(A)P(B)$$

Added Value:

$$\begin{aligned} \text{Added_Value} &= P(B|A) - P(B) = \frac{P(AB)}{P(A)} - P(B) \\ &= \frac{P(AB) - P(A)P(B)}{P(A)} \end{aligned}$$

Conviction:

$$\text{Conviction} = \frac{P(A)P(\neg B)}{P(A \neg B)}$$

Odds Ratio:

$$\text{Odds_Ratio} = \frac{P(AB)P(\neg A \neg B)}{P(A \neg B)P(\neg AB)}$$

Chi Square χ^2 [14] is a statistical test of discrete variables – looking for a difference:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O and E represent the Observed and Expected observations respectively.

Expected = (row total observations * column total observations) / total number of observations.

Use the χ^2 test is used to compare observed and expected values. By comparing the observed results with the expected ones, one can decide whether the original hypothesis is valid or not.

ChiDist: Returns the one-tailed probability of the chi-squared distribution. The χ^2 distribution is associated with a χ^2 test. The formula is:

$$\text{ChiDist}(x, \text{degrees_freedom})$$

where x is the value at which the distribution is evaluated.

Degrees_freedom is the number of the degrees of freedom.

A table's degrees of freedom (df) can be expressed by the following formula:

$$df = (r - 1)(c - 1)$$

That is, a table's degrees of freedom equals the number of rows (r) in the table minus one multiplied by the number of columns (c) in the table minus one.

IV. RESULTS

Three different sets of runs were investigated, for extracting rules: (i) MI versus PCI or CABG, (ii) PCI versus MI or CABG, and (iii) CABG versus MI or PCI. The corresponding rules for these runs are given in Table II. For each of the above models we used different or measures to select the most appropriate rules.

A. MI Events

Most of the extracted rules had the same confidence. Thus, we used the measures accuracy and relative risk to select the rules displayed in Table II in the MI section. Risk factors sex, family history and smoking are the most important risk factors as shown by the selected rules.

Rule 1.1 show that men with history of diabetes have no event with support 11% and confidence 52%. Rules 1.2 and 1.3 show that patients without family history, but smokers, have equal chance to have or not have an event. The same situation applies for rules 1.8 and 1.9 for men. Considering the number of cases in these rules we conclude that the factor smoking is not important for women. In rules 1.5, 1.6, 1.10, and 1.11 for patients with a family history, but no history of diabetes, it seems that 15% are women.

All rules gave no significant difference for the χ^2 test.

B. CABG Events

We used the measure relative risk to select the rules displayed in Table II in the CABG section.

Rules 2.4 – 2.11 are the most important ones due to the fact that these rules were statistically significant based on the χ^2 test.

Smokers that do not have history of diabetes have the same possibility to have an event as smokers with history of diabetes (rules 2.4 and 2.5). Patients in the age range of 61-70 with history of hypertension have a 60% risk to have an event (rules 2.6 and 2.7). Patients aged 61-70 with family history have the same possibility to have or not have an

event (rules 2.9 and 2.10). Sex does not influence risk (rule 2.11).

Rules 2.1 and 2.2 show that approximately 30% of the non-smoker patients with CABG were women. Non-smoker patients have a second event more often than smoker patients without a prior event (rules 1.1 and 1.3). Rules 2.6 and 2.7 show that 60% of patients in the age range of 60-70 years old, with a history of hypertension, have an event. Rules 2.12 and 2.13 show that 27 of 28 smokers with a history of hypertension are men. Rules 2.14 and 2.15 show that all smoker patients with a family history are men.

C. PCI Events

We used the measures coverage to select the rules displayed in table II in the PCI section.

Rules 3.3, 3.4, and 3.9 – 3.12 are the most important ones due to the fact that these rules are statistically significant based on the χ^2 test.

Two thirds of patients with family history have an event (rules 3.3 and 3.4). History of diabetes does not seem to affect the risk to have an event (rules 3.9 and 3.10). This also applies to male patients (rules 3.11 and 3.12).

Men have approximately the same chance to have an event (rules 3.1 and 3.2). Family history is more important (rules 3.3 and 3.4) than smoking (3.5 and 3.6) and a history of hypertension (rules 3.7 and 3.8).

Taking into consideration only one risk factor, we observe that there appears no difference between patients having an event and patients not having any event.

D. Comparison of the AKAMAS and Apriori algorithms

Both algorithms were used for association rule extraction and gave similar rules (see [15] for a comparison with the Apriori rules).

However, the advantage of the AKAMAS algorithm is that it offers a wide selection of measures for filtering the best rules.

The Apriori algorithm uses the frequent itemsets with $k-1$ items that satisfied the minimum threshold of support, to create the frequent itemsets with k items. Therefore, the results of the Apriori algorithm are dependent on support and rules with k items will be extracted only if during the previous repetition, the $k-1$ items satisfy the support. This is not the case for the AKAMAS algorithm.

The AKAMAS algorithm is dependent on support only for the first iteration, where it identifies the frequent itemsets with 1 feature which satisfy the minimum support. Then it builds all possible combinations of itemsets in order to find the association rules. Association rules may be identified and filtered with different measures that are selected by the user.

A performance study was carried out to compare the runtime of the AKAMAS and the Apriori algorithms. We have chosen to implement both algorithms in the same platform, Java, running under Windows XP. We ran the two algorithms with the same dataset. First we used a small set of data. The runtime required for both algorithms was the same. We used the same measures to extract the rules and

the extracted rules were the same. Then we used the complete set of data and did the same as above. The performance results were also the same.

V. CONCLUDING REMARKS

In this study, a new algorithm named AKAMAS was developed for extracting association rules for the assessment of heart event related risk factors. The events investigated were: MI, PCI, and CABG.

The advantages of the AKAMAS algorithm are the following:

- i. one pass is needed for extracting frequent itemsets,
- ii. it offers the possibility of selecting from a large number of measures to be used for filtering the best rules,
- iii. its measures include the computing of the chi-square test, and its associated probability for estimating if a rule is statistically significant or not, and
- iv. extracted rules are user friendly represented to the expert.

The use of measures enables the filtering of the extracted rules, but it is not possible to select the ideal measure for the best filtering of rules. For each described model, we used a different measure, due to the fact that there is no single measure ideal for all cases (datasets).

The usefulness of association rules in the analysis of CHD risk factors was also investigated by our group on the same database with this study using the Apriori association rules algorithm [15], and decision tree algorithm with four different splitting criteria [16]. The results regarding the most important risk factors were similar to the results of this study using the AKAMAS algorithm.

The most important risk factors, as extracted from the association rule analysis were: sex, family history, and smoking. Therefore the CHD risk of a patient may be reduced through a proper control of the smoking risk factor.

Ordóñez using the C4.5 decision tree algorithm and association rules for the prediction of cardiac disease based on 25 risk factors documented that association rules generally include simpler predictive rules than decision tree rules [19]. Wang *et al* [20] used the risk factors age, sex, cholesterol, HDL, blood pressure, diabetes, and smoking to predict CHD. They used the Framingham function and concluded that the traditional risk factors have different degrees of impact and/or than other factors are contributing to risk.

The first and second EUROASPIRE surveys showed high rates of modifiable cardiovascular risk factors in patients with coronary heart disease, and indicated that preventive measures might decrease cardiovascular risk [17], [18].

It is anticipated that data mining could help in the identification of high and low risk subgroups of patients, a decisive factor for the selection of therapy, i.e. medical or surgical. Moreover, the extracted rules could help to reduce CHD morbidity and possibly, mortality.

REFERENCES

- [1] D. W. Cheung, J. Han, V. T. Ng, A. W. Fu, and Y. Fu, "A Fast Distributed Algorithm for Mining Association Rules," in *proc. of the fourth int'l conference on Parallel and distributed information systems*, Miami Beach, Florida, United States, pp.: 31 – 43, 1996.
- [2] M. J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li, "New algorithms for fast discovery of association rules," in *proc. of the 8rd Int'l Conf. on KDD and Data Minin 9 (KDD '97)*, Newport Beach, California, 1997.
- [3] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *proc. of the 20th Int'l Conf. on Very Large Databases*, Santiago, Chile, SIGKDD Explorations, vol. 2, issue 1 – pp.: 63, 1994.
- [4] J. S. Park, M-S yan Chen and Ph. S. Yu, "An effective hash-based algorithm for mining association rules", *ACM SIGMOD Record* vol. 24, issue 2, pp.: 175 – 186, 1995.
- [5] A. Savasere, E. Omicinski, and S. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases," *Technical Report No. GIT-CC-95-04*, in *proc. of the 21th Int'l Conference on Very Large Data Bases*, pp.: 432 – 444, 1995.
- [6] H. Toivonen, "Sampling Large Databases for Association Rules", in *proc. of the 22th Int'l Conference on Very Large Data Bases*, pp.: 134 – 145, 1996.
- [7] H. Mannila, H. Toivonen, and I. Verkamo, "Discovery of frequent episodes in event sequences, *Data Mining and Knowledge Discovery*, 1(3), 1997.
- [8] J. Han, J. Pei, J. and Y. Yiwen, "Mining Frequent Patterns Without Candidate Generation", in *proc. ACM-SIGMOD Int'l Conference on Management of Data*, ACM Press, pp.:1-12, 2000.
- [9] J. Han, Y. Cai, and N. Cercone, "Knowledge Discovery in Databases: An Attribute-Oriented Approach, in *proc. of the 18th Int'l Conference on Very Large Data Bases*, pp.: 547- 559, 1992.
- [10] T.A. Pearson, S. N. Blair, S. R. Daniels, R. H. Eckel, J. M. Fair, S. P. Fortmann, B. A. Franklin, L. B. Goldstein, Ph. Greenland, S. M. Grundy, Y. Hong, N. H. Miller, R. M. Lauer, I. S. Ockene, R. L. Sacco, J. F. Sallis, S. C. Smith, N. J. Stone, and K. A. Taubert, "AHA guidelines for primary prevention of cardiovascular disease and stroke," *Circulation*, vol. 106, no. 3, pp.: 388–391, 2002.
- [11] I.H. Witten, E. Frank, "Data Mining – Practical Machine Learning Tools and Techniques," Morgan Kaufmann Publishers, 2005.
- [12] L. Geng and H. J. Hamilton, "Interestingness Measures for Data Mining: A Survey," *ACM Computing Surveys*, Vol. 38, No. 3, Article 9, 2006.
- [13] J. Han and M. Kamber, "Data Mining, Concepts and Techniques," Morgan Kaufmann Publishers, 2001.
- [14] J. Connor-Linton, "Chi Square tutorial", Georgetown University, <http://www.garnetthenley.com/ChiSquareLec.pdf>, Last access: 10/2009.
- [15] M. Karaolis, J.A. Moutiris, C.S. Pattichis, "Assessment of the risk of coronary heart event based on data mining," in 8th IEEE Int'l Conference on BioInformatics and BioEngineering, BIBE 2008, issue , pp.:1 – 5, 2008.
- [16] M. Karaolis, J. A. Moutiris, L. Papaconstantinou, C. S. Pattichis, "Association Rule Analysis for the Assessment of the Risk of Coronary Heart Events," *31st Annual International IEEE EMBS Conference*, September, 2-6, Minneapolis, Minnesota, USA, 4 pages, 2009.
- [17] Euroaspire study group, "A European Society of Cardiology survey of secondary prevention of coronary heart disease: Principal results," *European Heart Journal*, vol. 18, pp. 1569-1582, 1997.
- [18] Euroaspire II Study Group, "Lifestyle and risk factor management and use of drug therapies in coronary patients from 15 countries," *European Heart Journal*, vol. 22, pp. 554-572, 2002.
- [19] D. Ordóñez, E. Omiecinski, L. de Braal, C.A. Santana, N. Ezquerria, J. A. Taboada, D. Cooke, E. Krawczvnska, and E. V. Garcia, "Mining Constrained Association Rules to Predict Heart Disease," in *Proc. International Conference on Data Mining, ICDM, IEEE*, pp. 431-440, 2001.
- [20] Z. Wang and W. E. Hoy, "Is the Framingham coronary heart disease absolute risk function applicable to Aboriginal people?," *The Medical Journal of Australia*, vol. 182, no. 2, pp. 66-69, 2005.

TABLE II
ASSOCIATION RULES EXTRACTED WITH THE AKAMAS ALGORITHM

Rule	Sex	Age	Fh	Smokingbe	HxHTN	HxDM	class	NumOf Attributes	Support Count	Support	Confidence	Coverage	Recall	Specificity	Accuracy	Lift/Interest	Leverage	Relative Risk	x ²	p-value
MI																				
1.1	M					Y	N	2	32	0,11	0,52	0,21	0,21	0,50	0,51	1,03	0,41	0,50	0,83	NS
1.2			N	Y			N	2	41	0,14	0,50	0,27	0,27	0,50	0,50	1,00	0,36	0,50	2,29	NS
1.3			N	Y			Y	2	41	0,14	0,50	0,27	0,27	0,50	0,50	1,00	0,36	0,50	2,29	NS
1.4			Y	N			Y	2	42	0,14	0,51	0,27	0,28	0,50	0,51	1,02	0,38	0,50	2,29	NS
1.5			Y			N	N	2	53	0,18	0,50	0,36	0,35	0,50	0,50	0,99	0,32	0,50	2,05	NS
1.6			Y			N	Y	2	54	0,18	0,50	0,36	0,36	0,50	0,50	1,01	0,33	0,50	2,05	NS
1.7	M	3				N	Y	3	42	0,14	0,51	0,27	0,28	0,50	0,51	1,02	0,38	0,50	18,3	NS
1.8	M		N	Y			N	3	39	0,13	0,50	0,26	0,26	0,50	0,50	1,00	0,37	0,50	3,71	NS
1.9	M		N	Y			Y	3	39	0,13	0,50	0,26	0,26	0,50	0,50	1,00	0,37	0,50	3,71	NS
1.10	M		Y			N	N	3	46	0,15	0,50	0,31	0,31	0,50	0,50	1,00	0,35	0,50	4,19	NS
1.11	M		Y			N	Y	3	46	0,15	0,50	0,31	0,31	0,50	0,50	1,00	0,35	0,50	4,19	NS
CABG																				
2.1				N			Y	1	42	0,25	0,40	0,61	0,49	0,35	0,38	0,81	0,10	0,65	0,65	NS
2.2	M			N			Y	2	29	0,17	0,38	0,45	0,34	0,40	0,39	0,76	0,16	0,60	0,61	NS
2.3			Y				N	1	23	0,14	0,35	0,39	0,27	0,40	0,38	0,70	0,15	0,60	0,65	NS
2.4			Y			N	N	2	22	0,13	0,47	0,28	0,26	0,49	0,48	0,94	0,33	0,51	9,75	S
2.5			Y			N	Y	2	25	0,15	0,53	0,28	0,29	0,51	0,52	1,06	0,39	0,49	9,75	S
2.6		3			Y		N	2	18	0,11	0,41	0,26	0,21	0,47	0,45	0,82	0,28	0,53	62,5	S
2.7		3			Y		Y	2	26	0,15	0,59	0,26	0,31	0,53	0,55	1,18	0,46	0,47	62,5	S
2.8	M	3			Y		Y	3	20	0,12	0,63	0,19	0,24	0,53	0,55	1,25	0,53	0,47	34,4	S
2.9		3	Y				N	2	18	0,11	0,46	0,23	0,21	0,49	0,48	0,92	0,35	0,51	25,3	S
2.10		3	Y				Y	2	21	0,12	0,54	0,23	0,25	0,51	0,52	1,08	0,42	0,49	25,3	S
2.11	M	3	Y				Y	3	19	0,11	0,59	0,19	0,22	0,52	0,54	1,19	0,50	0,48	31	S
2.12				Y	Y		Y	2	28	0,16	0,76	0,22	0,33	0,57	0,61	1,51	0,65	0,43	2,59	NS
2.13	M			Y	Y		Y	3	27	0,16	0,77	0,21	0,32	0,57	0,61	1,54	0,67	0,43	2,59	NS
2.14			Y	Y			Y	2	23	0,14	0,64	0,21	0,27	0,54	0,56	1,28	0,53	0,46	2,21	NS
2.15	M		Y	Y			Y	3	23	0,14	0,66	0,21	0,27	0,54	0,56	1,31	0,55	0,46	3,17	NS
PCI																				
3.1	M						N	1	58	0,39	0,48	0,80	0,74	0,33	0,45	0,93	0,07	0,67	0,01	NS
3.2	M						Y	1	62	0,41	0,52	0,80	0,86	0,67	0,55	1,08	0,13	0,33	0,01	NS
3.3			Y				N	1	33	0,22	0,43	0,51	0,42	0,38	0,41	0,82	0,16	0,62	10,5	S
3.4			Y				Y	1	44	0,29	0,57	0,51	0,61	0,62	0,59	1,19	0,33	0,38	10,5	S
3.5				Y			N	1	41	0,27	0,54	0,51	0,53	0,50	0,52	1,04	0,28	0,50	1,89	NS
3.6				Y			Y	1	35	0,23	0,46	0,51	0,49	0,50	0,48	0,96	0,22	0,50	1,89	NS
3.7					Y		N	1	47	0,31	0,53	0,59	0,60	0,49	0,51	1,02	0,22	0,51	0,4	NS
3.8					Y		Y	1	42	0,28	0,47	0,59	0,58	0,51	0,49	0,98	0,19	0,49	0,4	NS
3.9						N	N	1	50	0,33	0,46	0,72	0,64	0,33	0,43	0,89	0,09	0,67	5,61	S
3.10						N	Y	1	58	0,39	0,54	0,72	0,81	0,67	0,57	1,12	0,19	0,33	5,61	S
3.11	M					N	N	2	39	0,26	0,42	0,62	0,50	0,32	0,38	0,81	0,10	0,68	9,43	S
3.12	M					N	Y	2	54	0,36	0,58	0,62	0,75	0,68	0,62	1,21	0,28	0,32	9,43	S

S: significantly different at $p \leq 0.05$, NS: non-significantly different at $p > 0.05$