Expert Systems with Applications 40 (2013) 1086–1093

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Association rule mining to detect factors which contribute to heart disease in males and females

Jesmin Nahar [a,*], Tasadduq Imam [a], Kevin S. Tickle [a], Yi-Ping Phoebe Chen [b]

[a] Faculty of Arts, Business, Informatics and Education, Central Queensland University, Queensland, Australia
[b] Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, Victoria 3086, Australia

ARTICLE INFO

*Keywords:*
Cleveland data
Heart disease
Computational intelligence
Association rule mining
Sick
Healthy
Male
Female

ABSTRACT

This paper investigates the sick and healthy factors which contribute to heart disease for males and females. Association rule mining, a computational intelligence approach, is used to identify these factors and the UCI Cleveland dataset, a biological database, is considered along with the three rule generation algorithms – Apriori, Predictive Apriori and Tertius. Analyzing the information available on sick and healthy individuals and taking confidence as an indicator, females are seen to have less chance of coronary heart disease then males. Also, the attributes indicating healthy and sick conditions were identified. It is seen that factors such as chest pain being asymptomatic and the presence of exercise-induced angina indicate the likely existence of heart disease for both men and women. However, resting ECG being either normal or hyper and slope being flat are potential high risk factors for women only. For men, on the other hand, only a single rule expressing resting ECG being hyper was shown to be a significant factor. This means, for women, resting ECG status is a key distinct factor for heart disease prediction. Comparing the healthy status of men and women, slope being up, number of coloured vessels being zero, and oldpeak being less than or equal to 0.56 indicate a healthy status for both genders.



## 1. Introduction

Throughout history, humans have been affected by life-threatening diseases. Of the various life-threatening diseases, heart disease has received a great deal of attention from medical researchers. As shown in Fig. 1, the Australian Bureau of Statistics and Cancer Biology (ABS, 2009; King & Robins, 2006) noted heart disease as one of the two highest causes of mortality in Australia and UK, with the other being cancer. For Australia, cancer is the cause of the greatest number of deaths, followed by heart disease, respiratory disease, mental disorder, accidents and others. On the other hand, in the UK, heart disease is the cause of the greatest number of deaths, followed by cancer, respiratory disease, mental disorder, accidents and others. With such a high mortality rate, it is necessary to gain a clearer understanding of the risk and prevention factors for this disease, as well as improving the accuracy of diagnosis. So, this research has considered factor determinations of coronary disease as the subject for computational diagnostics.

Computational intelligence concepts have recently been used in discovering the relationships between different diseases and patient attributes (Huang, Li, Su, Watts, & Chen, 2007; Ishibuchi, Kuwajima, Nojima, 2007; Karabatak & Ince, 2009; Shin et al., 2010; Wang & Hoy, 2005). So, this research also uses the computational intelligence approach. Particularly, this research presents rule extraction experiments on heart disease data using three different rule mining algorithms – Apriori, Predictive Apriori and Tertius. It also highlights the efficiency of these algorithms for this diagnostic task. A considerable issue in a research on heart disease diagnosis is the privacy issue related to medical data. So, Cleveland dataset (UCI, 2009), a publicly available dataset and widely popular with data mining researchers, has been used.

For heart disease, diagnostic systems are time consuming, costly and prone to errors. Patients suffering from heart disease need to be under constant observation as improper treatment can be fatal. Proper identification of the disease and early treatment are essential. The World Health Organization (WHO) identified the potential of data mining for improving the problems in this medical domain as early as 1997 (Gulbinat, 1997). In the WHO research, emphasis was placed on the usefulness of knowledge detection from medical data repositories that could benefit medical diagnosis and prediction, patient health planning and progress, healthcare system monitoring and assessment, hospital and health services management, and disease prevention. This paper is motivated by these views and the aforementioned issues, and proposes a set of computational intelligence based approaches for diagnosing heart disease.

* Corresponding author.
*E-mail addresses:* j.nahar@cqu.edu.au (J. Nahar), t.imam@cqu.edu.au (T. Imam), k.tickle@cqu.edu.au (K.S. Tickle), phoebe.chen@latrobe.edu.au (Yi-Ping Phoebe Chen).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.08.028

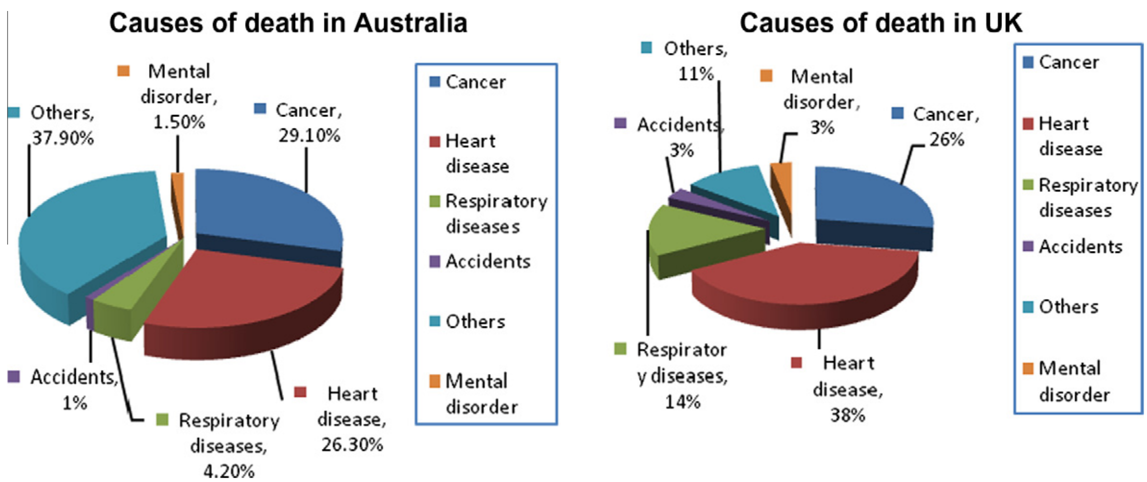


**Fig. 1.** A comparison of death from different causes in Australia (2009) and UK (2006). ABS (2009) and King and Robins (2006).

The plan of this paper is as follows: Section 2 presents different related concepts including heart anatomy, heart disease, association rule mining and Computational Intelligence in heart disease; Section 3 illustrates datasets details; Section 4 presents an association rule mining technique to derive human interpretable rules regarding heart disease for both sick and healthy males and females; and finally, Section 5 concludes the paper with a summary of findings and future research directions.

## 2. Different related concepts

This section will discuss the anatomy of the heart, heart disease, association rule mining algorithms and the use of computational intelligence in heart disease diagnostics.

### *2.1. Brief overview of the anatomy of the heart*

This section details the anatomy of heart, the general functions of this organ and the abnormalities that may lead to heart disease. The heart, a muscular organ situated near the middle of chest, is responsible for pumping blood to the other parts of the body and together with network of blood vessels and blood form the human body's cardiovascular system (Caster, 2010; Midgley, 2003; Sherwood, 2009). The organ has four chambers – two atria (the two upper chambers) and two ventricles (the two bottom chambers). Both the *atria* receive deoxygenated blood coming back from the body except that from the lungs, and the left section of the heart also receives oxygenated blood from the lungs. The right and left ventricles pump the blood back out into the body. There are four valves: the aortic valve, the pulmonary valve, the mitral valve and the tricuspid valve that control the forward and backward flows of the blood through the heart (Caster, 2010; Midgley, 2003; Sherwood, 2009; TEXASH, 2010). Disruptions to this circulation of blood can result in serious health problems including death. Scientists are, however, still unclear about the specific causes of heart disease. Details on the symptoms of heart disease and known diagnostics are provided in the following sub-section.

### *2.2. Heart disease*

This section will briefly discuss the characteristics of heart disease, its symptoms, causes, and known diagnostic techniques and the treatment options for this disease.

#### *2.2.1. Characteristics of heart disease*

Current research on heart disease research has established that it is not a single condition, but refers to any condition in which the heart and blood vessels are injured and do not function properly, resulting in serious and fatal health problems (Chilnick, 2008; HEALTHS, 2010; King, 2004; Silverstein et al., 2006). There are different types of heart diseases, among which the major types are: atherosclerosis, coronary, rheumatic, congenital, myocarditis, angina and arrhythmia (Health, 2010).

#### *2.2.2. Symptoms of heart disease*

Symptoms of this disease, however, differ from person to person. In majority of the cases, there is no early symptom and the disease is identifiable only in the advanced stage. Some common symptoms of heart disease are (Chilnick, 2008; Crawford, 2002; HEALTHS, 2010):

- chest pain (Angina pectoris);
- strong compressing or flaming sensation in the chest, neck or shoulders;
- discomforts in chest area;
- sweating, light-headedness, dizziness, shortness of breath;
- pain spanning from the chest to arm and neck, and that amplifying with exertion;
- cough;
- palpitations;
- fluid retention.

#### *2.2.3. Causes of heart disease*

The causes of heart disease are unclear, but age, gender, family history, and ethnic background are all considered to be the major causes in different investigatons (Chilnick, 2008; HEALTHS, 2010; King, 2004; Silverstein et al., 2006). Other factors like eating habits, fatty foods, lack of exercise, high cholesterol, hypertension, pollution, life style factors, obesity, high blood pressure, stress, diabetes and lack of awareness have also been claimed to increase the chance of developing heart disease (Chilnick, 2008; HEALTHS, 2010); King, 2004; Silverstein et al., 2006. Heart research, further, has found that the majority of the disease occurrence is noticed in people between the ages of 50–60 (Chilnick, 2008; HEALTHS, 2010); Silverstein et al., 2006.

#### *2.2.4. Diagnostic techniques of heart disease*

The diagnosis of heart disease patient depends on clinical history and physical examination, even though different diagnostic procedures exist (Chilnick, 2008; Crawford, 2002; HEALTHS, 2010; MERCK., 2010) including:

- **Electrocardiography (ECG):** a fast process monitoring the heart's electrical impulses.
- **Stress testing:** a heart-functionality test used by doctors to differentiate the problems limiting exercise due to a heart disorder or other issues.
- **Magnetic resonance imaging (MRI):** use of a powerful magnetic field and radio waves to form detailed images of the heart and chest.
- **Electrophysiologic testing:** test assessing serious abnormalities in heart rhythm or electrical conduction using a catheter with tiny electrodes.
- **Tilt table testing:** diagnostic method for people experiencing fainting (syncope) due to undetected reasons and who do not have a structural heart disorder (aortic valve stenosis).
- **Radiologic procedures (X-rays):** chest X-rays from the front and the side of the body to realize the shape and size of the heart, as well as the related blood vessels.
- **Angiography:** test involving the injection a dye through a catheter for viewing on X-rays and determining heart disease.
- **Other notable procedures are:** continuous ambulatory electrocardiography (Holter monitor), radionuclide imaging, cardiac catheterization, central venous catheterization. Fluoroscopy is also used to detect heart disease, though the use is generally infrequent. Blood tests involving measurement of sugar levels, cholesterol and other substances are also used.

Majority of these procedures have very small risk, however the risk intensifies with complex and severe heart disorder (Chilnick, 2008; Crawford, 2002; HEALTHS, 2010; Merck, 2010).

*2.2.5. Treatment options for heart disease*

Treatment of heart disease depends on the type, patient's age, health condition and the patient's choice. The major types of treatment are:

- **Medications:** Including beta-blockers, like atenol (Tenormin), nadolol (Corgard), and propranolol (Inderal); calcium channel blockers like, amlodipine (Norvasc), and nisoldipine (Sular); diuretics like chlorothiazide (Diuril) and hydrochlorothiazide (Esidrix); angiotensin-converting enzyme (ACE) Inhibitors like captopril (Capoten), enalapril (Vasotec), and lisinopril (Prinivil, Zestril); statins like atorvastatin (Lipitor) and rosuvastatin calcium (Crestor) (Crawford, 2002; HEALTHS, 2010).
- **Balloon angioplasty:** A nonsurgical process designed to enlarge narrowed coronary arteries (Crawford, 2002; HEALTHS, 2010), involving the insertion of a catheter into an artery in the arm or leg and a second smaller catheter, having a deflated balloon on its tilt, within the first catheter. The balloon is then inflated, to increase the diameter of the blood vessel's opening and resulting in increased circulation. A special X-ray camera (called fluoroscope) is used to monitor the process.
- **Bypass graft surgery:** In this procedure, cardiac surgeons remove a blood vessel from a part of the body and graft it to the narrowed or blocked coronary artery (Crawford, 2002; HEALTHS, 2010) with the aim to nourishing the vessel with blood flow again.
- **Electrophysiologic devices (Pacemakers):** Another surgical procedure involving the insertion of a device, called a pacemaker, into the body to maintain a minimum safe heart rate and regulated blood flow (Crawford, 2002; HEALTHS, 2010).

*2.3. Association rule mining*

This research will make extensive use of association learning. Association rule mining is a well recognised data mining procedure. In its basic structure, every association rule fulfilling the minimum support and confidence are extracted (Ishibuchi et al., 2007). A rule generated by association learning has the form 'LHS (left hand side) => RHS (right hand side)', where LHS and RHS are the disjoint sets of items. This rule expresses that the RHS set is likely to occur whenever the LHS set occurs. *Support* and *confidence* are two measures of rule interestingness that reflect the usefulness and certainty of a rule respectively (Hastie, Tibshirani, & Friedman, 2001). For a rule, $X \geqslant Y$,

$$\text{Support } (X \leqslant Y) = P(X \cap Y)$$

$$\text{Confidence } (X \leqslant Y) = P(Y|X) = \frac{\mathrm{P}(\mathrm{X} \cap Y)}{\mathrm{P}(\mathrm{X})}$$

Association learning is a promising knowledge discovery tool and has been widely explored in different areas, including the medical domain (Chen, Mabu, Shimada, & Hirasawa, 2010; Fu, Zhou, & Guo, 2010; Lei, Cui, & Mi, 2010; Ordonez, 2006; Ordonez et al., 2001; Shin et al., 2010). Different algorithms have been proposed to generate association rules, and depending on the method, the rules vary (Agrawal, Imielinski, & Swami, 1993; Flach & Lachiche, 2001; Scheffer, 2001). In this research, three techniques have been explored: Apriori, Predictive Apriori and Tertius.

*2.3.1. Apriori*

The Apriori algorithm has become a standard approach in association rule mining. It was first introduced by Agrawal and Srikant (1994). The algorithm starts with a dataset containing transactions and aims to construct frequent item sets, having at least a user specified threshold. In the algorithmic process of Apriori, an item set $X$ of length $k$ is frequent if and only if every subset of $X$, having length $k-1$, are also frequent. This consideration results in substantial reduction of search space and allows rule discovery in a computationally feasible time. Confidence is basically the accuracy of the rule and is used in Apriori to rank the rules (Agrawal et al., 1993; Mutter, Hall, & Frank, 2005; Taihua & Fan, 2010).

*2.3.2. Predictive Apriori*

Predictive Apriori (Scheffer, 2001), an algorithm motivated by Apriori, maximises the expected accuracy of an association rule on unseen data in contrast to the confidence related focus of Apriori. While Apriori ranks the rules based on confidence only, Predictive Apriori considers both the confidence and support in ranking the rules. Thus the algorithm has higer generalisation capacity (Scheffer, 2001). Considering a database $D$ with a set of records $r$ and a static process $P$, the predictive accuracy of the rule: $X \geqslant Y$ is the conditional probability of $Y$ being a subset of $r$, given that $X$ is a subset of $r$. A Bayesian framework is used to calculate the predictive accuracy out of the support and confidence of a rule. The technique is well suited for classification tasks and, with the datasets considered in this research mainly originating from the task of disease identification (in other words, classification), this algorithm has been chosen as a potential computational intelligence technique in this research.

*2.3.3. Tertius*

Tertius (Flach & Lachiche, 2001; Flach, Maraldi, & Fabrizio, 2006) is an inductive logic programming algorithm that looks for clauses with the highest value of a confirmation evaluation function. Different confirmation measures have been investigated, with the simplest one being weighted relative accuracy. A confirmation measure indicates the unexpectedness of a rule and the fraction of expected counter-instances. Two values, that are the expected probability and the observed probability, are calculated in the algorithm. Tertius can extract first order rules and has been applied in association rule mining tasks along with other schemes (Gunasekara et al., 2009; Sadri, 2002). The algorithm has shown promising results in these applications, so it is an interesting issue to see how well this rule mining algorithm performs compared to other schemes for the datasets in this research. Hence, the technique has also been considered in our study.

### 2.4. Computational Intelligence in heart disease detection

Computational intelligence techniques have been used in the diagnosis and analysis of heart diseases. Avci (2009), for example, proposed an intelligent system, based on using a Genetic-Support Vector machine (GSVM) approach to classify the Doppler signals of heart valve. Palaniappan and Awang (2008) developed a CAD (Computer Aided Diagnostics) prototype, called the Intelligent Heart Disease Prediction System (IHDPS), using a number of data mining techniques. Eberhart, Dobbins, and Webber (2002) used an adaptive neural network to classify multichannel Electrocardiograph (ECG) patterns. Cios, Chen, and Langenderfer (2002) used Neural Networks to detect cardiac diseases from echocardiographic images. Das, Turkoglu, and Sengur (2009) introduced a method that employs an SAS based software for heart disease diagnosis. Shin et al. (2010) analysed records of patients diagnosed with critical hypertension and showed, using association rule mining (ARM), the strong association of critical hypertension with non-insulin dependent diabetes mellitus (NIDDM) and cerebral infarction.While some researchers have utilized the different proprietary datsets (Kaya, 2010; Kou, Peng, Shi, & Chen, 2007; Luukka & Lampinen, 2010; Michelakos, Papageorgiou, & Vasilakopoulos, 2010; Patil, Joshi, & Toshniwal, 2010; Vijaya, Khanna Nehemiah, Kannan, & Bhuvaneswari, 2010), several other researchers have used the publicly available Cleveland dataset (UCI, 2009) in their research. These include: application of the database in determining the effectiveness of instance based learning algorithms (Aha & Kibler, 1988), probabilistic algorithm for diagnosing the risk of coronary artery disease (Detrano et al., 1989), comparison of global evolutionary computation approaches (Edmonds, 2005), and ensemble of classifiers for improved diagnostics (Fida, Nazir, Naveed, & Akram, 2011)

It is mentionable that while medical communities have noted different risk factors, it is still unclear what causes heart disease and more specifically, how these factors vary for men and women. This paper contributes a computational intelligence based approach to determine the risk factors, including consideration of the factor of gender.

## 3. Dataset details

As mentioned earlier, we use the publicly available UCI heart disease dataset in our research. The heart disease dataset consists of a total of 76 attributes, however majority of the studies use a maximum of 14 attributes (, 2010; UCI, 2009) as these are considerably linked to the heart disease. These 14 attributes are as follows (, 2010; UCI, 2009).

1. Age: *numeric*;
2. Sex: *nominal* – 2 values: male, female;
3. Chest pain type: *nominal* – 4 values: typical angina (angina), atypical angina (abnang), non-anginal pain (notang), asymptomatic (asympt).[1]
4. Trestbps: *numeric*, indicates resting blood pressure on admission;
5. Chol:: *numeric*, indicates Serum cholesterol in mg/dl;
6. Fbs: *nominal* – 2 values: True, False, indicates whether fasting blood sugar is greater than 120 mg/dl;
7. Restecg: *nominal* – 4 values: normal (norm), abnormal (abn): ST–T wave abnormality, ventricular hypertrophy (hyp) – indicates resting electrocardiographic outcomes;
8. Thalach: *numeric*, indicates maximum heart rate achieved;
9. Exang: *nominal* – 2 values: yes, no – highlights existence of exercise induced angina;
10. Oldpeak: *numeric*: ST depression induced by exercise relative to rest;
11. Slope: *nominal* – 3 values: upsloping, flat, downsloping – the slope characteristics of the peak exercise ST segment;
12. Ca: *numeric* – number of fluoroscopy colored major vessels (0–3);
13. Thal: *nominal* – 3 values: normal, fixed defect, reversible defect- the heart status;
14. The class attribute: value is either healthy or existence of heart disease (sick type: 1, 2, 3, and 4).

## 4. Association rule mining on heart disease data

While most existing works have considered the Cleveland database as a classification problem, we view, in this research, the dataset as a knowledge extraction problem and explore the use of association rule mining. Two experiments have been performed. The first experiment sets out extracting rules to indicate healthy and sick conditions. In the medical domain, the gender of a person has been found to be an important factor influencing heart disease (Andersen & Haraldsdottir, 2009; Barrett-Connor, Cohn, Wingard, & Edelstein, 1991; Dalaker, Smith, Arnesen, & Prydz, 2009; Ferrara et al., 2008; Flint et al., 2010; Haley, Roth, Howard, & Safford, 2010; Jeppesen, Hein, Suadicani, & Gyntelberg, 1998; Pencina, D'Agostino, Larson, Massaro, & Vasan, 2009; Schenck-Gustafsson, 2009; Tucker et al., 2009). A second experiment is so performed to discover rules based on gender. Details of these two experiments are provided in the following sub-sections.

### 4.1. Association rule mining to detect sick and healthy conditions

In this first experiment, all sick individuals were regarded to be in one class and healthy individuals to be in another class. Three popular association rule mining algorithms, Apriori, Predictive Apriori and Tertius, were used in this experiment. Results of the experiment are shown in Tables 1–3. Rules with confidence levels above 90%, with accuracy levels above 99% and confirmation levels above 79% were selected respectively for Apriori, Predictive Apriori and Tertius. As there can be many such rules, only the rules containing the 'sick' or 'healthy' class in the right-hand side (RHS) were considered. If no such rules were available, rules containing the 'sick' or 'healthy' class in the left-hand side (LHS) were reported.

For Apriori, four of the five rules for the 'healthy' class were

[1] Typical angina is the condition with patient showing the general symptoms and indicating high possibility of coronary artery blockages (Baliga and Eagle, 2008; Kaul, 2010; Diagnosis, 2010). Atypical angina refers to the condition in which the symptoms are not detailed and the probability of blockages is low (Baliga and Eagle, 2008; Kaul, 2010; Diagnosis, 2010). Non-angina pain is related to a stabbing, prolonged, dull, or painful condition (Mengel and Schwiebert, 2005; Society, 1945; Diagnosis, 2010). Asymptomatic pain is the condition with no symptoms of illness or disease (Pickett, 2000; FREEDC, 2010).

**Table 1**
Rule extraction for healthy and sick through the Apriori algorithm.

| Algorithms | Rules | Time (s) |
|---|---|---|
| Apriori | Healthy rules:<br>Healthy rule: If {Sex = female ∩ exercise_induced_angina = fal ∩ number_of_vessels_colored=0 ∩ thal = nom} => class healthy (conf., 0.98).<br>Healthy rule: If {Sex = female ∩ fasting_blood_sugar = fal ∩ exercise_induced_angina = fal ∩ number_of_vessels_colored = 0} => class healthy (conf., 0.98).<br>Healthy rule: If {Sex=female ∩ exercise_induced_angina = fal ∩ number_of_vessels_colored = 0} => class healthy (conf., 0.98).<br>Healthy rule: If {Sex = female ∩ fasting_blood_sugar = fal ∩ exercise_induced_angina = fal ∩ thal = norm} => class healthy (conf., 0.95).<br>Healthy rule: If {Resting_blood_pres less or = '(115.2,136.4]' ∩ exercise_induced_angina = fal ∩ number_of_vessels_colored = 0 ∩ thal = norm} => class healthy (conf., 0.94).<br>Sick Rules:<br>Sick rule: If {Chest_pain_type = asympt ∩ slope = flat ∩ thal = rev} => class sick (conf., 0.96).<br>Sick rule: If {Chest_pain_type=asympt ∩ exercise_induced_angina=TRUE ∩ thal=rev} => class sick (conf., 0.94). | 0 |

**Table 2**
Rule extraction for healthy and sick through the Predictive Apriori algorithm.

| Algorithms | Rules | Time (s) |
|---|---|---|
| Predictive apriori | Healthy rules:<br>Healthy rule: If {Sex=female ∩ fasting_blood_sugar = fal ∩ resting_ecg = norm ∩ exercise_induced_angina = fal ∩ thal=norm} => class healthy (acc., 0.9938).<br>Healthy rule: If {Sex = female ∩ chest_pain_type = notang ∩ thal = norm} => class healthy (acc., 0.9935).<br>Healthy rule: If {Age = '(48.2–57.8]' ∩ max_heart_rate = '(149.6,175.8]' ∩ exercise_induced_angina = fal ∩ number_of_vessels_colored=0} => class healthy (acc., 0.99314).<br>Healthy rule: If {Sex = female ∩ chest_pain_type = notang ∩ max_heart_rate = '(149.6,175.8]'} => class healthy (acc., 0.9921).<br>Healthy rule: If {Age = '(38.6–48.2]' ∩ resting_blood_pres = '(115.2,136.4]' ∩ thal = norm} => class healthy (acc., 0.9918).<br>Healthy rule: If {Sex = female ∩ exercise_induced_angina = fal ∩ number_of_vessels_colored=0} => class healthy (acc., 0.9901).<br>Sick Rules:<br>Sick rule: If {Age = '(48.2, 57.8]' ∩ slope = flat ∩ number_of_vessels_colored = 1} => class sick (acc., 0.9902).<br>Sick rule: If {Max_heart_rate = '(123.4,149.6]' ∩ exercise_induced_angina = TRUE ∩ thal = rev} => class sick (acc., 0.9931).<br>Sick rule: If {Sex = male ∩ chest_pain_type = asympt ∩ number_of_vessels_colored = 2} => class sick (acc., 0.9915).<br>Sick rule: If {Age = '(57.8,67.4]' ∩ sex = male ∩ number_of_vessels_colored = 2} => class sick (acc., 0.9902). | 2 min 43 see. |

**Table 3**
Rule extraction for healthy and sick through the Tertius algorithm.

| Algorithms | Rules | Time (Sec) |
|---|---|---|
| Tertius | Healthy rules:<br>Healthy rule: If {Chest_pain_type = angina or cholesteral = '(476.4, inf)' or thal = norm} => class healthy (conf., 0.30).<br>Healthy rule: If {Chest_pain_type = angina or max_heart_rate = '(175.8, inf)' or thal = norm} => class healthy (conf., 0.35).<br>Healthy rule: If {Chest_pain_type = abnang or max_heart_rate = '(175.8, inf)' or thal = norm} => class healthy (conf., 0.40).<br>Healthy rule: If {Cholesteral = '(476.4, inf)' or max_heart_rate = '(175.8, inf)' or thal = norm} => class healthy (conf., 0.32).<br>Healthy rule: If {Age = '(67.4, inf)' or chest_pain_type = notang or number_of_vessels_colored = 0} => class healthy (conf., 0.85).<br>Sick rules:<br>Sick rule: If {Chest_pain_type = asympt or resting_blood_pres = '(178.8, inf)' or oldpeak = '(2.48-3.72]'} => class sick (conf., 0.97).<br>Sick rule: If {Exercise_induced_angina = TRUE or oldpeak = '(2.48,3.72]' or thal = rev} => class sick (conf., 1).<br>Sick rule: If {Chest_pain_type = asympt or max_heart_rate = '(−inf, 97.2]' or oldpeak = '(2.48,3.72]'} => class sick (conf., 0.54).<br>Sick rule: If {Old peak = '(2.48,3.72]' or number_of_vessels_colored = 3 or thal = rev} => class sick (conf., 1).<br>Sick rule: If {Chest_pain_type = asympt or oldpeak = '(2.48,3.72]' or thal = fix} => class sick (conf., 0.80).<br>Sick rule: If {Exercise_induced_angina = TRUE or oldpeak = '(3.72, 4.96]' or thal = rev} => class sick (conf., 0.86). | 02 min 49 s 672 ms |

attributed to the female gender indicating that, based on this particular dataset, females have more chance of being free from coronary heart disease. Also if the results showed that when exercise induced angina (chest pain) was false, it was a good indicator of a person being healthy, irrespective of gender (exercise induced angina = false has appeared in the LHS of all the high confidence rules). The number of coloured vessels being zero and thal (heart status) being normal were also shown to be good indicators of health. Rules mined for the 'sick' class, on the other hand, showed that chest pain type being asymptomatic and thal being reversed were probable indicators of a person being sick (both the high confidence rules have these two factors in LHS).

The Predictive Apriori gave different results (Table 2). As mentioned earlier, in contrast to Apriori, which selects rules based on confidence, Predictive Apriori selects rules based on accuracy. Similar to Apriori, most of the rules for 'healthy' were attributed to females. However, the factors in the LHS varied. Exercise-induced angina being false and thal being normal, however, were again shown to be good indicators of health. Maximum heart rate in the interval (149.6, 175.8), the number of coloured vessels being zero and chest pain being notang (non-anginal pain) were also shown to be factors indicating healthy conditions. Considering the 'sick' class, the rules again varied considerably. Two of the rules were attributed to males. Slope (the peak exercise ST segment) being flat, age being over 48 and the existence of coloured vessels were also shown to be risk indicators for heart disease (by the presence of these factors in at least two of the rules).

When running Tertius (Table 3), it was noted that thal being normal is again a good indicator of health. Other attributes, such as maximum heart rate and cholesterol level were also shown to

**Table 4**
Rule extraction for females using Apriori algorithm for healthy and sick conditions. (Table terms are details in Section 3).

| Algorithms | Sex | Sick and healthy rules |
|---|---|---|
| Apriori | Female | Healthy rules: |
| | | Healthy rule: If {Cholesterol = '(225.6, 310.2]' ∩ resting ecg = hyp ∩ max heart rate = '(153.6, 172.8]' ∩ oldpeak = '(−inf, 1.24]' slope = up ∩ number of vessels colored = 0} => class healthy (conf., 1). |
| | | Healthy rule: If {Resting blood pres = '(115.2, 136.4]' ∩ resting ecg = norm ∩ max heart rate='(153.6, 172.8]' ∩ oldpeak = '(−inf, 1.24]' ∩ slope = up ∩ number of vessels colored = 0} => class healthy (conf., 1). |
| | | Healthy rule: If {Age = '(59.2, 67.6]' ∩ resting_ecg = norm ∩ oldpeak='(−inf, 1.24] ' ∩ slope = up ∩ number_of_vessels_colored = 0} => class healthy (conf., 1). |
| | | Healthy rule: If {Age = '(50.8, 59.2]' ∩ resting_blood_pres = '(115.2, 136.4]' ∩ max_heart_rate = '(153.6, 172.8]' ∩ oldpeak = '(−inf, 1.24]' ∩ number_of_vessels_colored = 0} => class healthy (conf., 1). |
| | | Healthy rule: If {Resting blood pres = '(115.2, 136.4]' ∩ cholesterol = '(225.6, 310.2]' ∩ max heart rate = '(153.6, 172.8]'∩ oldpeak = '(−inf, 1.24]' ∩ slope = up} => class healthy (conf., 1). |
| | | Healthy rule: If {Chest_pain_type = notang ∩ cholesteral = '(225.6, 310.2]' ∩ resting_ecg = hyp ∩ oldpeak = '(−inf, 1.24]'∩ slope = up} => class healthy (conf., 1). |
| | | Healthy rule: If {Resting blood pres = '(115.2, 136.4]' ∩ cholesterol = '(225.6, 310.2]' ∩ max heart rate = '(153.6, 172.8]' ∩ oldpeak = '(−inf, 1.24] ' ∩ number of vessels colored = 0} => class healthy (conf., 1). |
| | | Healthy rule: If {Chest pain type = notang ∩ resting blood pres = '(115.2, 136.4]' ∩ resting ecg = norm ∩ oldpeak = '(−inf, 1.24]' ∩ number of vessels colored = 0} => class healthy (conf., 1). |
| | | Sick rules: |
| | | Sick rule: If {Chest pain type = asympt ∩ cholesterol = '(225.6, 310.2]' ∩ resting ecg = norm ∩ exercise induced angina = TRUE} => class sick (conf., 1). |
| | | Sick rule: If {Cholesterol = '(225.6, 310.2]' ∩ resting ecg = norm ∩exercise induced angina = TRUE ∩ slope = flat} => class sick (conf., 1). |
| | | Sick rule: If {Chest pain type = asympt ∩ resting ecg = norm ∩exercise induced angina = TRUE ∩ slope = flat} => class sick (conf., 1). |
| | | Sick rule: If {Age = '(50.8, 59.2]' ∩ chest pain type = asympt ∩ exercise induced angina = TRUE ∩ slope = flat} => class sick (conf., 1). |
| | | Sick rule: If {Chest pain type = asympt ∩ exercise induced angina = TRUE ∩ slope = flat ∩ thal = rev} => class sick (conf., 1). |
| | | Sick rule: If {Chest pain type = asympt ∩ resting ecg = hyp ∩ slope = flat ∩ thal = rev} => class sick (conf., 1). |

**Table 5**
Rule extraction for males using the Apriori algorithm for healthy and sick. (Table terms are details in section 3).

| Algorithms | Sex | Rule |
|---|---|---|
| | Male | Healthy rules: |
| | | Healthy rule: If {Sex = male ∩ slope = up ∩ number_of_vessels_colored = 0 ∩ thal = norm} => class healthy (conf., 0.93). |
| | | Healthy rule: If {Sex = male ∩ exercise_induced_angina = fal ∩ slope = up ∩ number_of_vessels_colored = 0 ∩ thal = norm} => class healthy (conf., 0.92). |
| | | Healthy rule: If {Sex = male ∩ fasting_blood_sugar = fal ∩ slope = up ∩ number_of_vessels_colored = 0 ∩ thal = norm} => class healthy (conf., 0.92). |
| | | Healthy rule: If {Fasting_blood_sugar = fal ∩ slope = up ∩ number_of_vessels_colored = 0 ∩ thal = norm} => class healthy (conf., 0.92). |
| | | Healthy rule: If {Sex = male ∩ oldpeak = '(−inf, 0.56]' ∩ number_of_vessels_colored = 0 ∩ thal = norm} => class healthy (conf., 0.92). |
| | | Sick rules: |
| | | Sick rule: If {Sex = male ∩ chest_pain_type = asympt ∩ slope = flat ∩ thal = rev} => class sick (conf., 0.95). |
| | | Sick rule: If {Sex = male ∩ chest_pain_type = asympt ∩ exercise_induced_angina = TRUE ∩ thal = rev} => class sick (conf., 0.93). |
| | | Sick rule: If {Sex = male ∩ chest_pain_type = asympt ∩ fasting_blood_sugar = fal ∩ exercise_induced_angina = TRUE ∩ slope = flat} => class sick (conf., 0.92). |
| | | Sick rule: If {Sex = male ∩ chest_pain_type = asympt ∩ fasting_blood_sugar = fal ∩ exercise_induced_angina = TRUE ∩ thal = rev} => class sick (conf., 0.92). |
| | | Sick rule: If {Sex = male ∩ chest_pain_type = asympt ∩ resting_ecg = hyp ∩ exercise_induced_angina = TRUE} => class sick (conf., 0.92). |
| | | Sick rule: If {Chest_pain_type = asympt ∩ exercise_induced_angina = TRUE ∩ slope = flat} => class sick (conf., 0.92). |

be factors influencing healthy conditions. Results showed when oldpeak (when ST depression induced by exercise relative to rest) is greater than 2.48, it is a contributing factor to a person being sick (oldpeak appeared at RHS of all 'sick' rules). Chest pain type and presence of exercise-induced angina also showed an influence on sick conditions.

Overall, the three algorithms generated different rules. But the above results indicated women to be less at risk of developing heart disease. This finding is further investigated in the next section. Of the three algorithms, Apriori is the fastest and the resultant rules show a pattern. As a result of this and also because of its popularity, it was chosen as the rule mining algorithm for the next section.

### 4.2. Rule extraction for males and females

The findings in the previous section, that females are at lower risk of developing heart disease, are investigated here in more detail. The dataset was split, based on 'males' and 'females', and rules were extracted again for sick and healthy data. The Apriori algorithm was used to extract the rules for males and females. The aim was to see which factors are significantly related to heart disease in men and women separately. The results are presented in Tables 4 and 5, listing the rules with above 90% confidence.

Abbreviations used in the subsequent discussion have been detailed in Section 3. As described, the abbreviation 'hype' stands for ventricular hypertrophy, 'assumption' stands for the type of chest pain, 'slope' stands for peak exercise ST segment, 'thal' stands for heart status, and 'oldpeak' is a depression persuaded with exercise relative to rest.

The results show that for women (Table 4), oldpeak values less than or equal to 1.24 are a good indicator of being healthy. Other factors indicating a woman being healthy include resting blood pressure being between 115.2 and 136.4, cholesterol being between 225.6 and 310.2, slope being up and the number of colored vessels being zero. Similarly, resting ECG being hyp and normal, age being less than or equal to (5.0.8, 67.6), maximum heart rate being less than or equal to (153.6, 172.8), were all possible factors decreasing the risk of heart disease in women. For sick women, chest pain type being asympt, resting ECG is normal or hyp, exercise induced angina being true, slope being flat were all found to be significant risk factors for heart disease. Cholesterol being less than or equal to (225.6, 310.2), thal being reverse and age being less than or equal to (50.8, 59.2) were shown to be important risk factors for heart disease.

From Table 5, the significant factors (confidence of 1) for healthy men were the number of coloured vessels being 0, slope being up, and thal being normal. These factors were repeated the maximum number of times in the rules. On the other hand, for sick men, asymptomatic type chest pain and exercise-induced angina being true were shown to be high risk factors (appearing in the majority of the disease related rules). Thal (heart status) being reverse and slope being flat were also indicative of heart disease in men (the presence of these factors are in two high confidence rules out of five). Resting ECG being hyp in combination with the presence of exercise-induced angina and asymptomatic type chest pain also appeared as a high confidence 'sick' rule. This means males with these symptoms are at a high risk of severe heart disease. Similarly, fasting blood sugar in combination with asymptomatic chest pain, occurrence of exercise-induced angina and flat slope indicates the presence of heart disease in men.

Comparing these results with the factors impacting women, it can be seen that factors such as chest pain being asymptomatic and the presence of exercise-induced angina can indicate the existence of heart disease for both men and women. However, resting ECG being either normal or hyp (as detailed in Section 3), and slope being flat are potential high risk factors for women only. For men, on the other hand, only a single rule expressing resting ECG being hyp was shown to be a significant factor. This means resting ECG status is a key distinct factor for heart disease prediction. Comparing the healthy status of men and women, slope being up, the number of coloured vessel being zero, and oldpeak being less than or equal to 0.56 indicate a healthy status for both genders.

## 5. Conclusion

This research has presented a rule extraction experiment on heart disease data using different rule mining algorithms (Apriori, Predictive Apriori and Tertius). Further rule-mining-based analysis was undertaken by categorising data based on gender and significant risk factors for heart disease were found for both men and women. Interestingly, it is found from the set of healthy rules, being 'female' is one of the factors for a healthy heart condition. In other words, the results indicated females to have more chance of being free from coronary heart disease then males. This is supported by existing medical research as well. Research, for example, has identified that before the start of menopause, women have lower rates of coronary heart disease compared to their male counterparts of the same age (Castelli, 2007). The research also mentioned that estrogen, the female sex hormone, aids in storing heart-healthy HDL cholesterol contribute in this respect. Other research has also indicated that women are relatively more iron-deficient compared to men, especially younger woman in their late teens and early 20s due to menstruation (Blue, 2011). This research mentioned that iron reacts with cells producing damaging free radicals and the iron deficiency in younger women results in delayed occurrence of cardiovascular disease. The results of rule mining have, thus, highlighted useful knowledge.

Comparing the experimental results with the factors impacting women, it is seen that factors such as chest pain being asymptomatic and the presence of exercise-induced angina can indicate the existence of heart disease for both men and women. However, resting ECG being either normal or hyper and slope being flat are potential high risk factors for women only. For men, on the other hand, only a single rule expressing resting ECG being hyper was shown to be a significant factor. This means, for women, resting ECG status is a key distinct factor for heart disease prediction. Comparing the healthy status of men and women, slope being up, number of coloured vessels being zero, and oldpeak being less than or equal to 0.56 indicate a healthy status for both genders.

Overall, this research has demonstrated the use of rule mining to determine interesting knowledge. In medical literature, doctors are in discrepancies about the factors highlighted. This research has focused on the application of computational intelligence, in particular, association rule mining-based classifiers, to identify the key factors behind the disease, as well as considered gender diversity.

## References


ABS. 2009. *Causes of Death, Australia.* Australian Bureau of Statistics. <http://abs.gov.au/ausstats/abs@.nsf/Products/696C1CF9601E4D8DCA25788400127BF0?opendocument> Accessed 11.06.11.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the 20th international conference on very large data bases, 1994 Santiago, Chile. Citeseer* (pp. 487–499).

Agrawal, R.T., Imielinski, L., & Swami, A.N. (1993). Mining association rules between sets of items in large databases. In *International conference on. management of data* (*SIGMOD-93*) (pp. 207–216).

Aha, D., & Kibler, D. (1988). Instance-based prediction of heart-disease presence with the Cleveland database. Technical Report, University of California, Irvine, Department of Information and Computer Science, Number ICS-TR-88-07.

Andersen, L., & Haraldsdottir, J. (2009). Tracking of cardiovascular disease risk factors including maximal oxygen uptake and physical activity from late teenage to adulthood An 8-year follow-up study. *Journal of Internal Medicine, 234*, 309–315.

Avci, E. (2009). A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier. *Expert Systems with Applications, 36*, 10618–10626.

Baliga, R. R., & Eagle, K. A. (2008). *Practical cardiology: evaluation and treatment of common cardiovascular*. Lippincott Williams & Wilkins.

Barrett-Connor, E., Cohn, B., Wingard, D., & Edelstein, S. (1991). Why is diabetes mellitus a stronger risk factor for fatal ischemic heart disease in women than in men? The Rancho Bernardo Study. *JAMA, 265*, 627–631.

Blue, L., (2011). Why do women live longer than men? <http://www.time.com/time/health/article/0,8599,1827162,00.html> Accessed 20.04.11.

Castelli, W. P. (2007). *Cholesterol cures (revised): featuring the breakthrough menu plan to slash cholesterol by 30 points in 30 days*. Rodale Publisher.

Caster, S. (2010). *Heart*. Rosen Publishing Group.

Chen, C., Mabu, S., Shimada, K., & Hirasawa, K. (2010). Network intrusion detection using class association rule mining based on genetic network programming. *IEEE Transactions on Electrical and Electronic Engineering, 5*, 553–559.

Chilnick, L. D. (2008). *Heart disease: An essential guide for the newly diagnosed*. Da Capo Press.

Cios, K., Chen, K., & Langenderfer, R. (2002). Use of neural networks in detecting cardiac diseases from echocardiographic images. *IEEE Engineering in Medicine and Biology Magazine, 9*, 58–60.

Crawford, M. H. (2002). *Current diagnosis & treatment in cardiology*. McGraw-Hill Professional.

Dalaker, K., Smith, P., Arnesen, H., & Prydz, H. (2009). Factor VII-phospholipid complex in male survivors of acute myocardial infarction. *Acta Medica Scandinavica, 222*, 111–116.

Das, R., Turkoglu, I., & Sengur, A. (2009). Effective diagnosis of heart disease through neural networks ensembles. *Expert Systems with Applications, 36*, 7675–7680.

Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., et al. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology, 64*, 304–310.

Diagnosis, E. (2010). Coronary disease or heart attack from expert system: chest pain. <http://www.easydiagnosis.com/cgi-bin/expert/explain2.cgi?mod=Chest+Pain&ask=ddisease2&title=Coronary+Disease+or+Heart+Attack&showmod=yes> Accessed 12.12.10.

Eberhart, R., Dobbins, R., & Webber, W. (2002). Casenet: A neural network tool for EEG waveform classification. *Proceedings, second annual IEEE symposium on computer-based medical systems (current version), Minneapolis, MN, USA* (pp. 60–68). IEEE.

Edmonds, B. H. (2005). Using localised 'Gossip' to structure distributed learning, centre for policy modelling. In *Proceedings of the joint symposium on socially inspired computing engineering with social metaphors* (*AISB*), University of Hertfordshire, Hatfield, UK, (pp. 127–134).

Ferrara, A., Mangione, C., Kim, C., Marrero, D., Curb, D., Stevens, M., et al. (2008). Sex disparities in control and treatment of modifiable cardiovascular disease risk factors among patients with diabetes. *Diabetes Care, 31*, 69–74.

Fida, B., Nazir, M., Naveed, N., & Akram, S. (2011). *Heart disease classification ensemble optimization using Genetic algorithm*. IEEE.

Flach, P., Maraldi, V., & Fabrizio, R. (2006). *Algorithms for efficiently and effectively using background knowledge in Tertius*. Department of computer science, University of Bristol, Udine, Italy, Citeseer.

Flach, P. A., & Lachiche, N. (2001). *Confirmation-guided discovery of first-order rules with tertius*. Springer.

Flint, A., Rexrode, K., Hu, F., Glynn, R., Caspard, H., Manson, J., et al. (2010). Body mass index, waist circumference, and risk of coronary heart disease: A prospective study among men and women. *Obesity Research &Clinical Practice, 4*, e171–e181.

FREEDC (2010). Asymptomatic. <http://www.thefreedictionary.com/asymptomatic> Accessed 12.12.10.

Fu, D., Zhou, S., & Guo, P. (2010). Research on a distributed network intrusion detection system based on association rule mining (current version, 2010). *1st international conference on information science and engineering (ICISE), 2010 Nanjing, China* (pp. 1816–1818). IEEE.

Gulbinat, W. (1997). What is the role of WHO as an intergovernmental organization in the coordination of telematics in health care? <https://www.hon.ch/Library/papers/gulbinat.html> Accessed 13.12.10..

Gunasekara, R., Geethal, B., Hassan, M., Mathew, C., Perera, A., Subasinghe, H., et al. (2009). Prophetia: Artificial Intelligence for TravelBox® Technology. *Advances in, Computational Intelligence*, 21–34.

Haley, W., Roth, D., Howard, G., & Safford, M. (2010). Caregiving strain and estimated risk for stroke and coronary heart disease among spouse caregivers: Differential effects by race and sex. *Stroke, 41*, 331–336.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). Springer.

Health, M. (2010). Heart disease. <http://www.mamashealth.com/Heart_disease.asp> Accessed 20.02.11.

HEALTHS. (2010). Definition of Heart Disease. <http://www.healthscout.com/ency/68/458/main.html#DefinitionofHeartDisease> Accessed 1.04.10.

Huang, Z., Li, J., Su, H., Watts, G. S., & Chen, H. (2007). Large-scale regulatory network analysis from microarray data: Modified Bayesian network learning and association rule mining. *Decision Support Systems, 43*, 1207–1225.

Ishibuchi, H., Kuwajima, I., & Nojima, Y. (2007). *Prescreening of candidate rules using association rule mining and Pareto-optimality in genetic rule selection. Lecture notes in computer science. Knowledge-based intelligent information and engineering systems*. Springer. 509-516.

Jeppesen, J., Hein, H., Suadicani, P., & Gyntelberg, F. (1998). Triglyceride concentration and ischemic heart disease: an eight-year follow-up in the Copenhagen Male Study. *Circulation, 97*, 1029–1036.

Karabatak, M., & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications, 36*, 3465–3469.

Kaul, U. (2010). What is typical and atypical angina? <http://doctor.ndtv.com/faq/ndtv/fid/2907/What_is_typical_and_atypical_angina.html> Accessed 12. 02.10.

Kaya, M. (2010). Autonomous classifiers with understandable rule using multi-objective genetic algorithms. *Expert Systems with Applications, 37*, 3489–3494.

King, L. (2004). *Taking on heart disease*. Rodale.

King, R., & Robins, M. (2006). Pearson.

Kou, G., Peng, Y., Shi, Y., & Chen, Z. (2007). Privacy-preserving data mining of medical data using data separation-based techniques. *Data Science Journal, 6*, 429–434.

Lei, L., Cui, M., & Mi, Z. (2010). Study on association rule mining in multi-component Chinese medicine research and development for treatment of lung cancer. *China Journal of Chinese Materia Medica, 35*, 2192–2195.

Luukka, P., & Lampinen, J. (2010). A Classification method based on principal component analysis and differential evolution algorithm applied for prediction diagnosis from clinical EMR heart data sets. *Computational Intelligence in Optimization*, 263–283.

Mengel, M. B., & Schwiebert, L. P. (2005). *Family medicine: ambulatory care & prevention*. McGraw-Hill Professional.

Merck. 2010. *Diagnosis*. <http://www.merck.com/mmhe/sec03/ch021/ch021c.html> Accessed 13.09.10.

Michelakos, I., Papageorgiou, E., & Vasilakopoulos, M. (2010). A hybrid classification algorithm evaluated on medical data. *19th IEEE international workshops on enabling technologies: infrastructures for collaborative enterprises, 2010 Larissa, Greece* (pp. 98–103). IEEE.

Midgley, M. (2003). *The myths we live*. Routledge.

Mutter, S., Hall, M., & Frank, E. (2005). Using classification to evaluate the output of confidence-based association rule mining. *AI 2004: Advances in, Artificial Intelligence*, 133–148.

Ordonez, C. (2006). Association rule discovery with the train and test approach for heart disease prediction. *IEEE Transactions on Information Technology in Biomedicine, 10*, 334–343.

Ordonez, C., Omiecinski, E., De Braal, L., Santana, C., Ezquerra, N., Taboada, J., et al. (2001). Mining constrained association rules to predict heart disease. In *IEEE international conference on data mining, ICDM, 2001. Citeseer* (pp. 433–440).

Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. *International conference on computer systems and applications, AICCSA. IEEE/ACS, 2008 Doha* (pp. 108–115). IEEE.

Patil, B., Joshi, R., & Toshniwal, D. (2010). Effective framework for prediction of disease outcome using medical datasets: clustering and classification. *International Journal of Computational Intelligence Studies, 1*, 273–290.

Pencina, M., D'agostino, R., Sr, Larson, M., Massaro, J., & Vasan, R. (2009). Predicting the 30-year risk of cardiovascular disease: The Framingham Heart Study. *Circulation, 119*, 3078–3084.

Pickett, J. P. (2000). *The american heritage dictionary of the english language*. Houghton Mifflin.

Sadri, F. (2002). *Computational logic: Logic programming and beyond*. Springer.

Scheffer, T. (2001). Finding association rules that trade support optimally against confidence. *Proceedings of the 5th European conference on principles and practice of knowledge discovery in databases (PKDD'01)*. Freiburg, Germany: Springer-Verlag.

Schenck-Gustafsson, K. (2009). Risk factors for cardiovascular disease in women. *Maturitas, 63*, 186–190.

Sherwood, L. (2009). *Human physiology: From cells to systems*. Brooks Cole Pub Co.

Shin, A., Lee, I., Lee, G., Park, H., Park, H., Yoon, K., et al. (2010). Diagnostic analysis of patients with essential hypertension using association rule mining. *Healthcare Informatics Research, 16*, 77–81.

Silverstein, A., Silverstein, V. B. & Nunn, L. S. (2006). *Heart disease*. Twenty-First Century Books.

Society, M. M. (1945). The New England journal of medicine. Massachusetts Medical Society, New England Surgical Society, vol. 232. HighWire Press new.

Taihua, W., & Fan, G. (2010). Associating IDS alerts by an improved apriori algorithm. *Third international symposium on intelligent information technology and security informatics, 2010 Jinggangshan, China* (pp. 478–482). IEEE.

TEXASH. 2010. *Anatomy of the Human Heart with Flash Illustration*. <http://www.texasheart.org/hic/anatomy/anatomy.cfm> Accessed 28.09.10.

Tucker, A., Vogel, R., Lincoln, A., Dunn, R., Ahrensfield, D., Allen, T., et al. (2009). Prevalence of cardiovascular disease risk factors among National Football League players. *JAMA, 301*, 2111–2119.

UCI. 2009. Heart disease dataset <http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/cleve.mod> Accessed 5.03.09.

UCI. 2010. Cleveland Heart disease data details. <http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/heart-disease.names> Accessed 8. 02.10.

Vijaya, K., Khanna Nehemiah, H., Kannan, A., & Bhuvaneswari, N. (2010). Fuzzy neuro genetic approach for predicting the risk of cardiovascular diseases. *International Journal of Data Mining, Modelling and Management, 2*, 388–402.

Wang, Z., & Hoy, W. (2005). Is the Framingham coronary heart disease absolute risk function applicable to Aboriginal people. *Medical Journal of Australia, 182*, 66–69.