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 old-peak 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

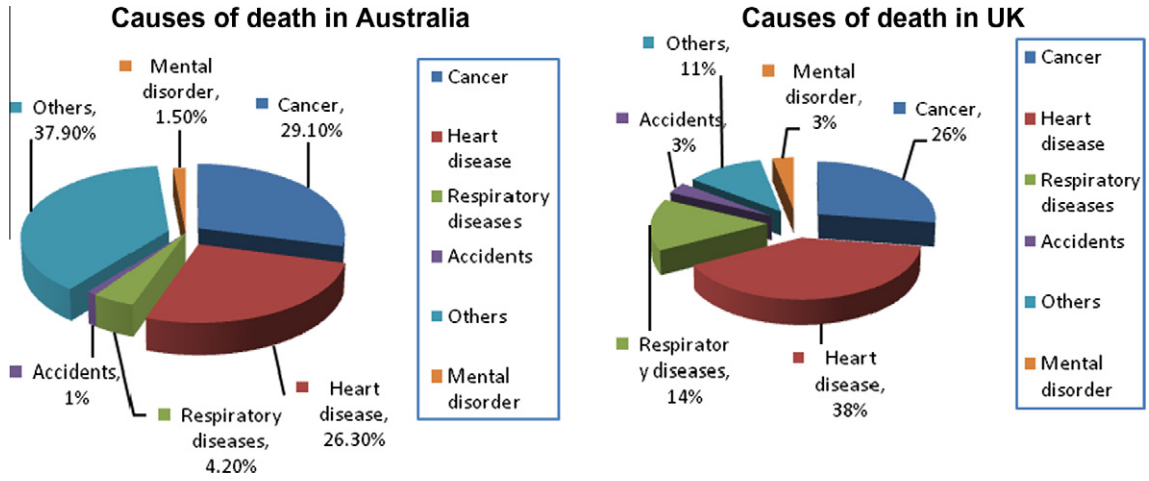

Fig. 1. A comparison of death from different causes in Australia (2009) and UK (2006). ABS (2009) and King and Robins (2006).

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