Chapter 6
Heart Disease Diagnosis Using Fuzzy Supervised Learning Based on Dynamic Reduced Features

Walid Moudani
Lebanese University, Lebanon

Mariam Abdelrazzak
Lebanese University, Lebanon

Mohamad Hussein
Lebanese University, Lebanon

Félix Mora-Camino
French Civil Aviation Institute (ENAC), France

ABSTRACT

The health industry collects huge amounts of health data which, unfortunately, are not mined to discover hidden information. However, there is a lack of effective analytical tools to discover hidden relationships and trends in data. Information technologies can provide alternative approaches to the diagnosis of the heart attach disease. In this study, a proficient methodology for the extraction of significant patterns from the Coronary Heart Disease warehouses for heart attack prediction, which unfortunately continues to be a leading cause of mortality in the whole world, has been presented. For this purpose, we propose to develop an innovative fuzzy classification solution approach based on dynamic reduced sets of potential risk factors using the promising Rough Set theory which is a new mathematical approach to data analysis based on classification of objects. Therefore, we propose to validate the classification using Multi-classifier decision tree to identify the risky heart disease cases. This work is based on a dataset collected from several clinical institutions based on the medical profile of patient. Moreover, the experts' knowledge in this field has been taken into consideration in order to define the disease, its risk factors, to follow up the issue results, and to establish significant knowledge relationships between medical factors related to Coronary Heart Disease. To identify cases of heart attack, experiments of several classification techniques have been performed leading to rank the suitable techniques. The reduction of potential risk factors contributes to enumerate dynamically one or more optimal subsets of the potential risk factors of high interest which implicitly leads to reduce the complexity of the classification problems while maintaining the prediction classification quality. The performance of the proposed model is analyzed and evaluated based on set of benchmark techniques applied in this classification problem.

DOI: 10.4018/978-1-5225-8185-7.ch006

1. INTRODUCTION

Medical diagnosis is an important but complicated task that should be performed accurately and efficiently in which its automation would be very useful and advantageous. Unfortunately, all doctors do not possess expertise in every sub specialty. Moreover, they are in many places a scarce resource. However, appropriate computer-based information and/or decision support systems can aid in enhancing medical care and in achieving clinical tests at a reduced cost. Or, efficient and accurate implementation of automated system needs a comparative study of various available techniques. Indeed, most hospitals today employ some kinds of hospital information systems to manage their healthcare or patient data (Obenshain, 2004; Usher, Laakso, James & Rowlands, 2013). These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. The main motivation of our research is to process data in order to get useful information that enables healthcare practitioners to make intelligent clinical decisions.

In healthcare domain, data mining leads to improve decision-making by discovering patterns and trends in large amounts of complex data. Anticipating patient's future behavior on the given history is one of the important applications of data mining techniques that can be used in health care management, by employing appropriate computer-based information and/or decision support systems. Healthcare organizations should have ability to analyze data. So by applying data mining techniques, this may help in answering several important and critical questions related to healthcare. The researchers in the medical field identify and predict the diseases besides proffering effective care for patients (Palaniappan & Awang, 2008; Lemke & Mueller, 2003) with the aid of data mining techniques. The data mining techniques have been utilized by a wide variety of works in the literature to diagnose various diseases including: Diabetes, Hepatitis, Cancer, Heart diseases and the like (Gorunescu, 2008; Bellaachia & Guven, 2006; Nakić & Loškovska, 2012). Information associated with the disease, prevailing in the form of electronic clinical records, treatment information, gene expressions, images and more; were employed in all these works. Recently, the data mining techniques were utilized by several authors to present diagnosis approaches for diverse types of heart diseases (Guru, Dahiya & Rajpal, 2007; Szymanski, Han, Embrechts, Ross, Sternickel & Zhu, 2006; Le Duff, Munteanb, Cuggiaa & Mabob, 2004; Ordonez, 2004). However, such analysis has become increasingly essential as financial pressures have heightened the need for healthcare organizations to make decisions based on the analysis of clinical and financial data. Insights gained from data mining can influence cost, revenue, and operating efficiency while maintaining a high level of care. In this paper, we deal with the diagnosis of one of the real health problem, called Coronary Heart Disease (CHD), because of its increasing frequency over the countries. CHD is assessed as the primary cause of mortality among adults in the world, and it is the top concern of healthcare organizations and medical doctors. The social economic burden of CHD is so large that its etiology, prevention and treatment have become an urgent issue that needs to be coped with worldwide. The diagnosis of this disease is a vital and intricate job in medicine. The recognition of CHD from diverse features or signs is a multi-layered problem that is not free from false assumptions and is frequently accompanied by impulsive effects. Thus, the attempt to exploit knowledge and experience of several specialists and clinical screening data of patients composed in databases to assist the diagnosis procedure is regarded as a valuable option. The methodology applied in this study to build the mining predictive model consists of several phases that start with medical-technical environment understanding, data understanding, data preparation, modeling, implementation and evaluation. The environment understanding phase focuses on