Fuzzy Decision Support System for Coronary Artery Disease Diagnosis Based on Rough Set Theory

Noor Akhmad Setiawan, Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia

ABSTRACT

The objective of this research is to develop an evidence based fuzzy decision support system for the diagnosis of coronary artery disease. The development of decision support system is implemented based on three processing stages: rule generation, rule selection and rule fuzzification. Rough Set Theory (RST) is used to generate the classification rules from training data set. The training data are obtained from University California Irvine (UCI) data repository. Rule selection is conducted by transforming the rules into a decision table based on unseen data set. Furthermore, RST attributes reduction is proposed and applied to select the most important rules. The selected rules are transformed into fuzzy rules based on discretization cuts of numerical input attributes and simple triangular and trapezoidal membership functions. Fuzzy rules weighing is also proposed and applied based on rules support on the training data. The system is validated using UCI heart disease data sets collected from the U.S., Switzerland and Hungary and data set from Ipoh Specialist Hospital Malaysia. The system is verified by three cardiologists. The results show that the system is able to give the approximate possibility of coronary artery blocking.

Keywords: Coronary Artery Disease, Decision Support System, Diagnosis, Fuzzy, Rough Set Theory

INTRODUCTION

Artificial Intelligence (AI), including Artificial Neural Networks (ANN), fuzzy logic, evolutionary computing, machine learning and expert systems has been widely accepted as a tool for intelligent decision support systems. In recent year, there are broad applications of AI in medical decision support systems for diagnosis, prognosis and treatment of patients.

One of the first killer diseases (Zaret et al., 1992) in developed countries is coronary artery disease as well as in developing countries as reported by the World Health Organization (Alwan, 2011) and British Heart Foundation (Weissberg, 2012). Thus, researches on the AI application in the diagnosis of coronary artery disease are very important to support the medical doctor to give the better and accurate diagnosis.

Coronary Artery Disease (CAD) is the development of plaque inside the wall of coronary arteries. The development of plaque will narrow or block the coronary arteries and will make the wall of arteries less elastic. This
plaque will make the blood difficult to flow freely. The CAD is considered presence when the narrowing of at least one of the coronary arteries is more than 50%. Coronary angiogram or cardiac catheterization is considered as the “gold standard” method to diagnose the presence of CAD. This method has high accuracy, but it is invasive, risky, expensive and not possible as a diagnosis for a large population. Many research works were conducted to diagnose the CAD using less expensive and non-invasive methods such as electrocardiogram (ECG) based analysis, heart sound analysis, medical image analysis and others (Phibbs, 2007; Randall & Romaine, 2005; Selzer, 1992; Zaret et al., 1992).

RELATED WORKS

AI and Knowledge Discovery from Data (KDD) methods have been used as a decision support tool to diagnose CAD (Setiawan et al., 2011). The AI and KDD methods use heterogeneous data from the patients such as physical and historical data, exercise tests, and laboratory tests to develop a computer system that can classify and diagnose the presence of CAD.

There are several research works on coronary artery disease diagnosis. A multilayer perceptron based medical decision support system was developed to diagnose five types of heart diseases which are hypertension, coronary heart disease, rheumatic valvular heart disease, chronic cor pulmonale and congenital heart disease, simultaneously (Yan et al., 2006). However, ANN itself cannot explain its knowledge, even though the system has high accuracy. Bayesian network model of heart disease is proposed (Jayanta & Marco, 2000). The system could predict the probabilities of heart diseases and dependency among attributes related to heart diseases. A set of machine learning methods was evaluated on the atherosclerotic coronary heart disease (Gamberger et al., 2000). However, the objective is not to compare different machine learning results but to explore the possibilities of both machine learning and medical expertise improving the quality of a regular medical practice. Prognosis of cardiac events (cardiac death or non-fatal myocardial infarction) was proposed (Komorowski & Ohm, 1999). Diagnosis of ischemic heart disease using various machine learning techniques was developed (Kukar et al., 1997). The data consist of 4000 patients. Extension of multilayer perceptron for coronary heart disease diagnosis by making it interpretable was introduced (Bologna et al., 1997). Fuzzy discrimination analysis for diagnosis of valvular heart disease was also proposed (Watanabe et al., 1994). Most of these research works use a large number of patients. Fuzzy expert system was built to diagnose heart disease (Adeli & Neshat, 2010). Their work used fuzzy rules and inference. The fuzzy rules and membership function were not discovered from evidences but determined subjectively. Adaptive Neuro Fuzzy Inference System (ANFIS) was proposed to diagnose heart disease (Kumar, 2012). ANFIS can learn from data to optimize the fuzzy rules, but the ANFIS parameters must be determined first before optimization process, in other words it needs human interventions. ANFIS also has a high computational burden. There are research works that concentrate on the risk prediction of CAD using risk factor categories (J. Hippisley-Cox et al., 2008a; J. Hippisley-Cox et al., 2007; Julia Hippisley-Cox et al., 2008b; Wilson et al., 1998). The works used a large amount of data and performed in a long-term study. The objectives are not the diagnosis but risk prediction of CAD during certain years based on the Framingham score with modification such as QRISK1 and QRISK2. A combination of data mining technique, namely C4.5 and fuzzy modeling was used to diagnose CAD (M. G. Tsipouras et al., 2006; M. G. Tsipouras et al., 2008). Their work emphasized on fuzzy modeling of C4.5 generated rules. This system used 19 attributes of 199 objects to predict the presence of 50% or more narrowing in at least one of coronary artery vessels. Similar work was proposed using Reduced Error Pruning (REP) instead of C4.5. Particle Swarm Optimization was used to optimize the fuzzy membership functions (Muthukaruppan & Er, 2013). Computational intelligence methods were used to diagnose