J Med Syst (2016) 40:178
DOI 10.1007/s10916-016-0536-z

TRANSACTIONAL PROCESSING SYSTEMS

A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data

Luxmi Verma[1] · **Sangeet Srivastava**[2] · **P. C. Negi**[3]

Received: 1 March 2016 / Accepted: 1 June 2016

Abstract Coronary artery disease (CAD) is caused by atherosclerosis in coronary arteries and results in cardiac arrest and heart attack. For diagnosis of CAD, angiography is used which is a costly time consuming and highly technical invasive method. Researchers are, therefore, prompted for alternative methods such as machine learning algorithms that could use noninvasive clinical data for the disease diagnosis and assessing its severity. In this study, we present a novel hybrid method for CAD diagnosis, including risk factor identification using correlation based feature subset (CFS) selection with particle swam optimization (PSO) search method and K-means clustering algorithms. Supervised learning algorithms such as multi-layer perceptron (MLP), multinomial logistic regression (MLR), fuzzy unordered rule induction algorithm (FURIA) and C4.5 are then used to model CAD cases. We tested this approach on clinical data consisting of 26 features and 335 instances collected at the Department of Cardiology, Indira Gandhi Medical College, Shimla, India. MLR achieves highest prediction accuracy of 88.4 %.We tested this approach on benchmarked Cleaveland heart disease data as well. In this case also, MLR, outperforms other techniques. Proposed hybridized model improves the accuracy of classification algorithms from 8.3 % to 11.4 % for the Cleaveland data. The proposed method is, therefore, a promising tool for identification of CAD patients with improved prediction accuracy.

Keywords Classification · Particle swarm optimization · Coronary artery disease · Clustering

Introduction

Cardiovascular diseases (CVD) are caused by disorders of the heart and blood vessels and result in coronary heart disease, heart failure, cardiac arrest, ventricular arrhythmias and sudden cardiac death, ischemic stroke, transient ischemic attack, subarachnoid and intracerebral hemorrhage, rheumatic heart disease, abdominal aortic aneurysm, peripheral artery disease and congenital heart disease [1]. According to World Health Organization (WHO), 17.5 million people died from CVD in 2012 amounting to 31 % of all global deaths [2]. CAD is a type of CVD in which presence of atherosclerotic plaques in coronary arteries, leads to myocardial infarction or sudden cardiac death [3]. In order to diagnose positive sign of heart disease and to assess the level of damage of heart muscles, certain tests may be prescribed by a medical practitioner including nuclear scan, angiography, echocardiogram, electrocardiogram (ECG), exercise stress testing [4]. ECG is a noninvasive technique used to identify CAD cases [5, 6], though it could lead to undiagnosed symptoms of CAD [7]. This limitation leads to angiography which is an invasive diagnosis to confirm CAD cases and is considered as the gold standard for disease detection and severity analysis. However, it is costly and requires high level of technical expertise [8]. Researchers are, therefore, seeking less expensive and effective alternatives, say, using

This article is part of the Topical Collection on *Transactional Processing Systems*

✉ Sangeet Srivastava
sangeetsrivastava@ncuindia.edu

[1] Department of Computer Science and Engineering, The NorthCap University, Gurgaon, India

[2] Department of Applied Sciences, The NorthCap University, Gurgaon, India

[3] Department of Cardiology, Indira Gandhi Medical College, Shimla, India

Published online: 11 June 2016