Computer Methods and Programs in Biomedicine 141 (2017) 19–26

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm

Zeinab Arabasadi[a], Roohallah Alizadehsani[b,*], Mohamad Roshanzamir[c], Hossein Moosaei[d], Ali Asghar Yarifard[a]

[a] Department of Computer Engineering, University of Bojnord, Bojnord, Iran
[b] Department of Computer Engineering, Sharif University of Technology, Azadi Ave, Tehran, Iran
[c] Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran
[d] Department of Mathematics, Faculty of Science, University of Bojnord, Iran

ARTICLE INFO

Article history:
Received 11 September 2016
Revised 18 December 2016
Accepted 12 January 2017

Keywords:
Cardiovascular disease
Coronary artery disease
Neural network
Genetic algorithm

ABSTRACT

Cardiovascular disease is one of the most rampant causes of death around the world and was deemed as a major illness in Middle and Old ages. Coronary artery disease, in particular, is a widespread cardiovascular malady entailing high mortality rates. Angiography is, more often than not, regarded as the best method for the diagnosis of coronary artery disease; on the other hand, it is associated with high costs and major side effects. Much research has, therefore, been conducted using machine learning and data mining so as to seek alternative modalities. Accordingly, we herein propose a highly accurate hybrid method for the diagnosis of coronary artery disease. As a matter of fact, the proposed method is able to increase the performance of neural network by approximately 10% through enhancing its initial weights using genetic algorithm which suggests better weights for neural network. Making use of such methodology, we achieved accuracy, sensitivity and specificity rates of 93.85%, 97% and 92% respectively, on Z-Alizadeh Sani dataset.

1. Introduction

Data mining refers to the extraction of valid patterns, hidden information and relationships from large datasets [1]. This interdisciplinary subfield is employed in banking, insurance, and marketing for cost reduction and quality improvement [2]. Recent years have witnessed a surge of interest in utilizing machine learning and data mining in the field of medicine for early diagnosis [3–10].

Cardiovascular disease is the most widespread cause of death the world over. Particularly, coronary artery disease (CAD) is the most common cardiovascular condition.

CAD occurs when at least one of the left anterior descending (LAD), left circumflex (LCX), and right coronary (RCA) arteries is stenotic [11].

In order to diagnose CAD, physicians currently employ different methods, among which angiography is widely regarded as the most precise method. It is, however, associated with high costs and major side effects, hence researchers have long sought to devise precise diagnostic modalities. In fact, a large number of investigations have been conducted in this field [12–34], with the UCI datasets [35] being the most frequently utilized sets. These datasets, however, are not up-to-date.

In the present study, given the risks of invasive diagnostic procedures such as angiography and auspicious experiences in the field of data mining, attempts were made to propose a model for identifying coronary arteries disease.

The suggested detection model, based on artificial neural networks and genetic algorithms, can detect coronary artery disease based on clinical data without the need for invasive diagnostic methods.

The present research primarily introduces the required background information; in the subsequent four sections, the proposed method, experimental results, related works and conclusions will be discussed.

2. Background

2.1. Dataset

The present research used Z- Alizadeh Sani dataset, containing information on 303 patients, 216 of whom suffered from CAD. Fifty-four features were collected for each patient. These features

* Corresponding author.
E-mail addresses: alizadeh_roohallah@yahoo.com, ro_alizadeh@alum.sharif.edu (R. Alizadehsani).

http://dx.doi.org/10.1016/j.cmpb.2017.01.004
0169-2607/

Table 1
Features of Z-Alizadeh Sani dataset.

Feature type	Feature name	Range
Demographic	Age	*30–86*
	Weight	*48–120*
	Sex	*Male, Female*
	BMI (Body Mass Index Kg/m^2)	*18–41*
	DM (Diabetes Mellitus)	*Yes, No*
	HTN (Hypertension)	*Yes, No*
	Current smoker	*Yes, No*
	Ex-smoker	*Yes, No*
	FH (Family History)	*Yes, No*
	Obesity	*Yes if MBI > 25, No otherwise*
	CRF (Chronic Renal Failure)	*Yes, No*
	CVA (Cerebrovascular Accident)	*Yes, No*
	Airway disease	*Yes, No*
	Thyroid disease	*Yes, No*
	CHF (Congestive Heart Failure)	*Yes, No*
	DLP (Dyslipidemia)	*Yes, No*
Symptom and examination	BP (Blood Pressure mm Hg)	*90–190*
	PR (Pulse Rate ppm)	*50–110*
	Edema	*Yes, No*
	Weak peripheral pulse	*Yes, No*
	Lung rales	*Yes, No*
	Systolic murmur	*Yes, No*
	Diastolic murmur	*Yes, No*
	Typical chest pain	*Yes, No*
	Dyspnea	*Yes, No*
	Function class	*1, 2, 3, 4*
	Atypical	*Yes, No*
	Nonanginal chest pain	*Yes, No*
	Exertional chest pain	*Yes, No*
	Low Th Ang (low-Threshold angina)	*Yes, No*
ECG	Rhythm	*Sin, AF*
	Q wave	*Yes, No*
	ST elevation	*Yes, No*
	ST depression	*Yes, No*
	T inversion	*Yes, No*
	LVH (Left Ventricular Hypertrophy)	*Yes, No*
	Poor R-wave progression	*Yes, No*
Laboratory and echo	FBS (Fasting Blood Sugar mg/dL)	*62–400*
	Cr (Creatine mg/dL)	0.5–2.2
	TG (Triglyceride mg/dL)	37–1050
	LDL (Low-Density Lipoprotein mg/dL)	18–232
	HDL (High-Density Lipoprotein mg/dL)	15–111
	BUN (Blood Urea Nitrogen mg/dL)	6–52
	ESR (Erythrocyte Sedimentation Rate mm/h)	1–90
	HB (Hemoglobin g/dL)	8.9–17.6
	K (Potassium mEq/lit)	3.0–6.6
	Na (Sodium mEq/lit)	128–156
	WBC (White Blood Cell cells/mL)	3700–18,000
	Lymph (Lymphocyte %)	7–60
	Neut (Neutrophil %)	32–89
	PLT (Platelet 1000/mL)	25–742
	EF (Ejection Fraction %)	15–60
	Region with RWMA	0,1,2,3,4
	VHD (Valvular Heart Disease)	Normal, Mild, Moderate, Severe

encompass the data on the patients' demographic characteristics, symptoms and the results of physical examinations, electrocardiography, echocardiography, and laboratory tests. These features are shown in Table 1. In this dataset, if at least one of the LAD, LCX, and RCA has a stenosis of higher than 50%, the patient is diagnosed with CAD.

2.2. Feature selection

For feature selection, four famous ranking methods were considered, namely Gini index, weight by SVM, information gain and principal component analysis (PCA). Gini Index shows the probability of incorrectly labeling a randomly chosen element according to the distribution of labels in the subset if it is randomly labeled [36]. Information Gain shows the expected reduction in entropy owing to partitioning records based on a given attribute [36]. In weight by SVM, attribute weights are the coefficients of the normal vector of a linear SVM [37]. PCA converts a set of correlated variables into a smaller number of uncorrelated variables employing an orthogonal transformation [38].

Weight by SVM uses F-score so as to measure feature weights. If we have training instance x_i, and i = 1, 2, ..., L (L: Number of instances), the F-score of feature j is calculated as shown in Eq. (1). The higher the F-score, the more discriminative the feature will be [39].

$$F(j) = \frac{\left(\bar{x}_j^{(+)} - \bar{x}_j\right)^2 + \left(\bar{x}_j^{(-)} - \bar{x}_j\right)^2}{\frac{1}{n_+ - 1}\sum_{i=1}^{n_+}\left(x_{i,j}^{(+)} - \bar{x}_j^{(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{i=1}^{n_-}\left(x_{i,j}^{(-)} - \bar{x}_j^{(-)}\right)^2} \quad (1)$$

where n_+ and n_- are the number of positive and negative samples, respectively. The jth feature average of the whole, and the positive-labeled and negative labeled samples are shown with