Available online at www.sciencedirect.com
ScienceDirect

Journal of Biomedical Informatics 41 (2008) 15–23

Journal of Biomedical Informatics

www.elsevier.com/locate/yjbin

Medical diagnosis of atherosclerosis from Carotid Artery Doppler Signals using principal component analysis (PCA), *k*-NN based weighting pre-processing and Artificial Immune Recognition System (AIRS)

Fatma Latifoğlu [a], Kemal Polat [b], Sadık Kara [a,*], Salih Güneş [b]

[a] *Erciyes University, Department of Electronics Engineering, 38039 Kayseri, Turkey*
[b] *Selcuk University, Department of Electrical and Electronics Engineering, 42075 Konya, Turkey*

Received 18 July 2006
Available online 11 April 2007

Abstract

In this study, we proposed a new medical diagnosis system based on principal component analysis (PCA), *k*-NN based weighting pre-processing, and Artificial Immune Recognition System (AIRS) for diagnosis of atherosclerosis from Carotid Artery Doppler Signals. The suggested system consists of four stages. First, in the feature extraction stage, we have obtained the features related with atherosclerosis disease using Fast Fourier Transformation (FFT) modeling and by calculating of maximum frequency envelope of sonograms. Second, in the dimensionality reduction stage, the 61 features of atherosclerosis disease have been reduced to 4 features using PCA. Third, in the pre-processing stage, we have weighted these 4 features using different values of *k* in a new weighting scheme based on *k*-NN based weighting pre-processing. Finally, in the classification stage, AIRS classifier has been used to classify subjects as healthy or having atherosclerosis. Hundred percent of classification accuracy has been obtained by the proposed system using 10-fold cross validation. This success shows that the proposed system is a robust and effective system in diagnosis of atherosclerosis disease.

Keywords: Atherosclerosis; Carotid artery; Fast Fourier Transformation; Welch; PCA; *k*-NN based weighting pre-processing; AIRS

1. Introduction

Atherosclerosis is the build-up of fatty deposits called plaque on the inside walls of arteries. Plaques can grow large enough to significantly reduce the blood's flow through an artery. As an artery becomes more and more contracted, less blood could flow through. The artery may also become less elastic (resulting in the condition called "hardening of the arteries"). Atherosclerosis is the main cause of a group of cardiovascular diseases [1].

Atherosclerosis is usually diagnosed after symptoms or complications have arisen. There are a number of tests for diagnosing vascular diseases, including blood tests, electrocardiogram, stress testing, angiography, ultrasound, and X-ray CT. Angiography is used to look inside arteries to see the presence and significance of any blockage [2,3]. This is the most accurate way to assess the presence and severity of vascular disease. On the other hand this technique involves injecting dye directly into the arteries. Therefore this is much more invasive.

Since angiography is invasive and has a relatively high cost, noninvasive ultrasonic Doppler sonography is generally recommended. Recent advances in Doppler imaging technique have made the evaluation of the temporal and spatial flow characteristics possible in the different portions of the arterial system, such as aorta, coronary, carotid and

* Corresponding author. Fax: +90 352 437 5784.
E-mail addresses: fdirgenali@tse.org.tr (F. Latifoğlu), kpolat@selcuk.edu.tr (K. Polat), kara@erciyes.edu.tr (S. Kara), sgunes@selcuk.edu.tr (S. Güneş).

1532-0464/$ - see front matter
doi:10.1016/j.jbi.2007.04.001

peripheral arteries. The analysis of Doppler signals provides information about physiology and pathology, through applications of flow measurement and the detection of significant changes in "Doppler waveform" [4].

It is often useful to analyze the spectrum of the Doppler-shifted signal to assess the degree of a disease. The use of spectrum analysis to display Doppler frequency shifts provides not only the best means of measuring blood flow velocity but also information about the presence of disturbed flow [5,6]. The Doppler sonograms describe how the power of a time series is distributed with frequency and in this way spectral analysis of the Doppler signal produce information concerning the velocity distribution in the artery [7,8]. In a sonogram, the horizontal axis represents time, the vertical axis (f) represents frequency and the gray level intensity at coordinates (t,f) denotes the signal power at frequency f and time instant t. As the color tone of the sonogram turns darker, the power level is increased and as it becomes lighter, the power level diminishes.

Several authors have reported decent results using the diagnostic performance of the carotid artery waveform in last decades [10,11]. However, none of these methods have gained wide-spread use, mainly because of their complexity and need for additional equipment. Still, subjective visual examination of the carotid artery waveform profile has a potential for inter- and intra-observer variability. To overcome this, several methods of numerical analysis have been surveyed. Of these, the most straightforward ones are the waveform profile indices, such as the pulsatility index (PI) [12] and the Pourcelot or resistance index (RI) from sonogram [13]. More sophisticated methods have also been developed, such as the Laplace transform [14,15]. But, neither the simple nor the more complex analytical techniques have yielded an acceptable diagnostic accuracy to make them common place in the vascular clinic.

One of the central problems of the information age is dealing with the enormous amount of raw information that is available. More and more data is being collected and stored in databases or spreadsheets. As the volume increases, the gap is widening between generating and collecting the data and actually being able to understand it. In order to bridge this knowledge gap, a variety of techniques known as data mining or knowledge discovery is being developed. Knowledge discovery can be defined as the extraction of implicit, previously unknown, and potentially useful information from real world data, and communicating the discovered knowledge to people in an understandable way [16–18].

The use of classification systems in medical diagnosis is increasing gradually. There is no doubt that evaluation of data taken from patient and decisions of experts are the most important factors in diagnosis. However, expert systems and different Artificial Intelligence techniques for classification also help experts in a great deal.

Artificial Immune Systems (AISs) is a new Artificial Intelligence (AI) technique which is beginning to mature through the collaborative effort of many interdisciplinary researchers [19]. By modeling some metaphors existing in natural immune system or by inspiring from these metaphors, successful applications have being conducted in AI literature. Classification is among these and there have been some promising studies in this branch of AISs. Considering medical diagnosis as an application domain for AISs, there are several studies such as [20–24] in AIS literature. Almost all of these studies use the datasets from UCI (University of California at Irvine) database [25] when conducting the classification process. While these datasets reflect real-world medical problems, they are recorded and processed datasets which are ready to use directly in a classification system and so they are mainly used for comparison of a proposed system with other studies in the literature. Therefore, this study is the first attempt to apply an Artificial Immune System to a real-world medical classification problem.

In this study, a diagnostic system leading to more effective usage of the Doppler technique is presented. Our primary research motivation was to progress the research on atherosclerosis. The proposed method consists of four stages. First, we have obtained the features about atherosclerosis disease using Fast Fourier Transformation (FFT) modeling and calculation of maximum frequency envelope of sonograms. Second, dimensionality of atherosclerosis disease that has 61 features has been reduced to 4 features using PCA. Third, we have weighted atherosclerosis disease data that has 4 features using different values of k in the new weighting scheme based on k-NN based weighting pre-processing. Finally, AIRS classifier has been used to classify subjects as healthy or atherosclerosis.

2. Data acknowledgments

2.1. Hardware and demographic acknowledgments

Carotid arterial Doppler ultrasound signals were acquired from 60 patients and 54 healthy volunteers. The patient group included 33 males and 27 females with an established diagnosis of atherosclerosis through coronary or aortofemoropopliteal (lower extremity) angiography (mean age: 45 years; range: 25–69 years). Healthy volunteers including 35 males and 19 females (mean age: 26 years; range: 20–39 years) were young non-smokers who appeared not to bear any risk of atherosclerosis. The two study groups represent the upper and lower extremes of the arterial compliance. We have utilized Toshiba PowerVision 6000 Doppler Ultrasound Unit in the Radiology Department for data acquisition.

A linear ultrasound probe of 10 MHz was used to transmit pulsed ultrasound signals to the proximal left common carotid artery. In all tests performed on the patients and healthy subjects, the insonation angle and the presetting of the ultrasound were kept constant.

In order to obtain the Doppler responses at the carotid arteries, we used the audio output port on the ultrasound