COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 130 (2016) 54–64

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Congestive heart failure detection using random forest classifier

Zerina Masetic[a], *Abdulhamit Subasi*[b,*]

[a] *International Burch University, Faculty of Engineering and Information Technologies, 71000 Sarajevo, Bosnia and Herzegovina*

[b] *Effat University, College of Engineering, Computer Science Department, Jeddah 21478, Saudi Arabia*

ARTICLE INFO

Article history:
Received 17 November 2015
Received in revised form
13 March 2016
Accepted 16 March 2016

Keywords:
Electrocardiogram (ECG)
Congestive heart failure (CHF)
Autoregressive (AR) modeling
Machine learning
Random forest

ABSTRACT

Background and objectives: Automatic electrocardiogram (ECG) heartbeat classification is substantial for diagnosing heart failure. The aim of this paper is to evaluate the effect of machine learning methods in creating the model which classifies normal and congestive heart failure (CHF) on the long-term ECG time series.

Methods: The study was performed in two phases: feature extraction and classification phase. In feature extraction phase, autoregressive (AR) Burg method is applied for extracting features. In classification phase, five different classifiers are examined namely, C4.5 decision tree, *k*-nearest neighbor, support vector machine, artificial neural networks and random forest classifier. The ECG signals were acquired from BIDMC Congestive Heart Failure and PTB Diagnostic ECG databases and classified by applying various experiments.

Results: The experimental results are evaluated in several statistical measures (sensitivity, specificity, accuracy, *F*-measure and ROC curve) and showed that the random forest method gives 100% classification accuracy.

Conclusions: Impressive performance of random forest method proves that it plays significant role in detecting congestive heart failure (CHF) and can be valuable in expressing knowledge useful in medicine.

1. Introduction

Human heart is the most important and hardest working muscle in human body, that together with blood vessels compose cardiovascular system. It pumps the blood into the every cell of the human body. Furthermore, heart muscle is the engine of the human body [1]. Heart failure is the common syndrome that progresses slowly but causes cardiac dysfunction resulted from the inability of the heart to pump the blood to all the cells of the human body efficiently. The heart weakens by heart attacks, long-term high blood pressure or an anomaly of one of the heart valves. Yet, heart failure is generally not recognized until it comes to the more progressive phase, entitled congestive heart failure, which causes fluid to flow to lungs, feet and abdominal cavity. Congestive heart failure is a condition that can be caused by heart diseases, such as coronary artery disease, damage of the heart after the heart attack, high blood pressure, valvular heart disease, diabetes and even the alcoholism [2]. According to European Heart Network and European Society of Cardiology [3], over 4 million people die from cardiovascular diseases in

* *Corresponding author*. Tel.: +966 590129599.
E-mail addresses: zmasetic@ibu.edu.ba (Z. Masetic), absubasi@effatuniversity.edu.sa (A. Subasi).
http://dx.doi.org/10.1016/j.cmpb.2016.03.020

Europe and 1.9 million in European Union (EU) which is 47% deaths in Europe and 40% in EU.

Considering that there is no definite diagnosis of heart failure, medical diagnosis such as history or physical examinations, electrocardiography (ECG), chest radiography or echocardiography is crucial for detecting the congestive heart failure. The electrocardiogram (ECG) is noninvasive tool that records electrical activity of the heart and shows irregularities of the heartbeats. It safely examines and records the electrical impulses of the heart and show possible damages of the heart or irregularities of the heartbeats [4]. Thus, ECG is an important tool for determining the function and the health of the cardiovascular system. Moreover, it is significant to define accurate and timely diagnosis of physicians to avoid more damage and to determine proper methods and approaches [5]. Still, the problem occurs when there is insufficient number of physicians to meet the needs of patients. Therefore, it is necessary to develop an effective and automated diagnostic systems based on ECG recordings, combined with application of machine learning techniques for classification of heart diseases. These diagnostic systems will aid medical experts in detecting the irregularities in the cardiovascular system. The diagnostic system will firstly process the ECG signals taken from different subjects and hence decomposed into few features performing feature extraction. Extracted ECG signals are used to detect different types of heart failure by using various machine learning techniques [6].

According to the number of researches done, the field of heartbeats classification using different techniques is very popular. Beth Israel Deaconess Medical Center (BIDMC) congestive heart failure database were used in different studies. Baim et al. [7] were used this database to show the effect of treatment with oral milrinone. The larger group of 100 patients, with severe congestive heart failure belonging to NYHA3 and NYHA4 groups was treated with oral milrinone with an average initial dose of 27 ± 8 mg/day. Causes of congestive heart failure were different (e.g. ischemic heart disease, dilated cardiomyopathy, valve replacement, etc.). Thuraisingham used [8] BIDMC congestive heart failure database in detecting CHF from normal heartbeats, using k-nearest neighbor algorithm and features from the second-order difference plot (SODP) obtained from Holter monitor cardiac RR intervals. Six features are obtained from second-order difference plot and k-nearest neighbor algorithm is applied with the value of $k = 1$. Authors obtained a success rate of 100% in separating CHF and normal heartbeats. The same database was used by Kuntamalla and Reddy [9] and they applied sequential trend analysis to differentiate CHF patients from patients with normal heart beats. The accuracy of the proposed method is 96.68%. Furthermore, Kuntamalla and Reddy [10] applied multiscale entropy (MSE) to HRV signals, to differentiate healthy young and elderly subjects from CHF patients. They applied Reduced Data Dualscale Entropy Analysis method to reduce the data size for clear differentiation of subjects. They achieved 100% accuracy. Hossen and Al-Ghunaimi [11] used technique based on recognition of power spectral densities pattern of decomposed sub-bands of R–R interval to identify patients with congestive heart failure. They used 12 subjects from BIDMC congestive heart failure database and reached accuracy of 90%. The same authors, Hossen and Al-Ghunaimi [12] used different wavelet decomposition filters with soft decision algorithm to estimate power spectral density of RR interval data for screening patients with congestive heart failure. The accuracy value of test data with almost all wavelet filters used was 88.6%. Yu and Lee [13] focused on selecting the best feature selector to get high accuracy in recognizing the CHF. The authors proposed conditional mutual information feature selector (CMIFS) with support vector machine (SVM) classifier. The combination achieved 97.59% accuracy using only 15 features. Işler and Kuntalp [14] combined classical HRV and wavelet entropy measures to distinguish healthy patients from patients with congestive heart failure. They applied genetic algorithm to select the best ones among all possible combinations of measures. Furthermore, they used k-nearest neighbor classifier with different values of k. Finally, they got accuracy value of 96.39% with value of $k = 5$ and 7, using 8–10 features. Similar study was done by Asyali [15] and author used linear discriminant analysis to examine the discrimination power of 9 long-term HRV measures and validated the result by applying Bayesian classifier. Sensitivity and specificity value of the stated classifier are 81.8% and 98.1%, respectively. Pecchia et al. [16] investigated discrimination power of heart rate variability (HRV) in distinguishing normal subject and subjects with congestive heart failure (CHF). They performed time and frequency analysis in order to measure HRV features. The result is evaluated by applying classification and regression tree (CART) classifier. Furthermore, authors introduced two non-standard features: average of normal intervals and low/high frequencies for recording over the 24 h and reached the success rate of 89.7% for sensitivity and 100% for specificity.

Feature extraction methods are important in ECG signal classification. In this paper, features are extracted by use of autoregressive Burg method. Consequently, specified signal is categorized as normal or with congestive heart failure by using machine learning techniques. In previous works [8,9,11–15,17], various machine learning techniques were applied. However, k-NN method is not capable of dealing with high dimensional data without dimension reduction, where SVM is not strong enough to handle large number of trivial data without data selection [18]. Decision tree method might not perform well with more complex interactions and overfitting might lead to the instability of the model [19]. Method efficient to achieve excellent performance, but not applied in previously mentioned studies [8–15] is random forest (RF). In this paper, RF method is used to investigate the performance of classifier for ECG signals classification in diagnosing the congestive heart failure. Moreover, performance of different machine learning techniques such as C4.5 decision tree, ANN, k-NN and SVM in ECG signal classification is compared (see Fig. 1.).

The rest of the paper is organized as follows: Section 2 gives explanation about the data used and present methods applied in each step of the ECG signal classification process. Experimental results are presented and discussed in Section 3. Finally, conclusions derived from the results are summarized in Section 4.