Expert Systems with Applications 42 (2015) 3643–3652

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Classification of electrocardiogram and auscultatory blood pressure signals using machine learning models

Manjeevan Seera [a,*], Chee Peng Lim [b], Wei Shiung Liew [c], Einly Lim [c], Chu Kiong Loo [a]

[a] Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
[b] Centre for Intelligent Systems Research, Deakin University, Australia
[c] Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia

ARTICLE INFO

Article history:
Available online 20 December 2014

Keywords:
Machine learning
Data classification
Medical signals
Electrocardiogram
Auscultatory blood pressure

ABSTRACT

In this paper, two real-world medical classification problems using electrocardiogram (ECG) and auscultatory blood pressure (Korotkoff) signals are examined. A total of nine machine learning models are applied to perform classification of the medical data sets. A number of useful performance metrics which include accuracy, sensitivity, specificity, as well as the area under the receiver operating characteristic curve are computed. In addition to the original data sets, noisy data sets are generated to evaluate the robustness of the classifiers against noise. The 10-fold cross validation method is used to compute the performance statistics, in order to ensure statistically reliable results pertaining to classification of the ECG and Korotkoff signals are produced. The outcomes indicate that while logistic regression models perform the best with the original data set, ensemble machine learning models achieve good accuracy rates with noisy data sets.

1. Introduction

Data classification constitutes one of the fundamental requirements in undertaking many decision-making tasks (Örkcü & Bal, 2011). A classification task involves building a model that depicts a mapping from the input feature space to the target output space (Oza & Tumer, 2008). In general, there are a number of classification methods, which include statistical methods, mathematical programming methods, and a variety of machine learning methods (Örkcü & Bal, 2011). Researchers in the medical domain have used many methods to perform data classification. Methods with higher classification accuracy are desirable to correctly identify potential diseases; therefore improving diagnosis accuracy (Fan, Chang, Lin, & Hsieh, 2011).

The main contribution of this study is a comprehensive performance evaluation and analysis pertaining to a number of machine learning models for undertaking real medical data classification problems. Specifically, we use two sets of real data collected from patients, i.e., the electrocardiogram (ECG) and auscultatory blood pressure (Korotkoff) signals. ECGs are signals related to electrical activity of the heart, which can be recorded by placing surface electrodes on a patient's body (Mitra, Mitra, & Chaudhuri, 2006). It is an effective non-invasive clinical tool for the diagnosis of certain cardiovascular diseases, and it provides useful information pertaining to pathological physiology of heart activity (Chen & Yu, 2012). ECG signals carry valuable information about the heart function, and provide a cardiologist with useful insight about the rhythm and functioning of the heart (Chen & Yu, 2012).

As stated in Mele (2008), an estimated 300 million ECGs are performed each year. As such, there is a clear need for reliable and accurate interpretation tools of ECG readings. Although trained cardiologists can discover different cardiac abnormalities in ECG recordings, it is time-consuming and laborious for them to examine a large number of ECG recordings (Kiranyaz, Ince, Pulkkinen, & Gabbouj, 2011). Moreover, visual inspection can take considerable time, and some vital information can be neglected due to fatigue in carrying out the tedious manual procedure (Sun, Lu, Yang, & Li, 2012). As such, automated tools to help accurately analyze a large number of ECG data samples are required. Similarly, blood pressure (BP) is an established link in determining coronary heart disease and cardiovascular incidents (Mendiola, Luna, Guerra, & Ramírez, 2013). The most commonly used methods for measuring BP in clinical activities is the use of a manual sphygmomanometer and a stethoscope to detect the Korotkoff sounds (Kurl et al., 2001). Korotkoff waveforms are one of the most reliable means for monitoring blood pressure (Mendiola et al., 2013). To exploit the advantages of using computerized tools to help medical prognosis

* Corresponding author.
E-mail addresses: mseera@gmail.com (M. Seera), chee.lim@deakin.edu.au (C.P. Lim), liew.wei.shiung@gmail.com (W.S. Liew), einly_lim@um.edu.my (E. Lim), ckloo.um@um.edu.my (C.K. Loo).

http://dx.doi.org/10.1016/j.eswa.2014.12.023
0957-4174/