computer methods and programs in biomedicine 117 (2014) 257–266

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

Cardiovascular risk analysis by means of pulse morphology and clustering methodologies

Vânia G. Almeida [a,*], J. Borba [a], H. Catarina Pereira [a,b], Tânia Pereira [a], Carlos Correia [a], Mariano Pêgo [c], João Cardoso [a]

[a] *Physics Department, Electronics and Instrumentation Group, University of Coimbra, Portugal*
[b] *Intelligent Sensing Anywhere, Portugal*
[c] *Cardiology Department, Hospital and University Coimbra Center, Portugal*

ARTICLE INFO

Article history:
Received 16 September 2013
Received in revised form
12 June 2014
Accepted 17 June 2014

Keywords:
Arterial stiffness
Pulse wave analysis
Risk scores
Clustering analysis

ABSTRACT

The purpose of this study was the development of a clustering methodology to deal with arterial pressure waveform (APW) parameters to be used in the cardiovascular risk assessment. One hundred sixteen subjects were monitored and divided into two groups. The first one (23 hypertensive subjects) was analyzed using APW and biochemical parameters, while the remaining 93 healthy subjects were only evaluated through APW parameters. The expectation maximization (EM) and k-means algorithms were used in the cluster analysis, and the risk scores (the Framingham Risk Score (FRS), the Systematic COronary Risk Evaluation (SCORE) project, the Assessing cardiovascular risk using Scottish Intercollegiate Guidelines Network (ASSIGN) and the PROspective Cardiovascular Münster (PROCAM)), commonly used in clinical practice were selected to the cluster risk validation. The result from the clustering risk analysis showed a very significant correlation with ASSIGN ($r = 0.582$, $p < 0.01$) and a significant correlation with FRS ($r = 0.458$, $p < 0.05$). The results from the comparison of both groups also allowed to identify the cluster with higher cardiovascular risk in the healthy group. These results give new insights to explore this methodology in future scoring trials.

1. Introduction

The atherosclerotic cardiovascular disease (CVD) is the most common cause of death worldwide, resulting from the combination of several risk factors [1]. The international guidelines [2,3] consider that individuals with established CVD should be the first priority for preventive measures application. The concern in changing the current healthcare paradigm, from reactive towards preventive care, aims at identify individuals for risk in early stages of disease development, and then, direct more efforts and attention to the risk factors modification [4,5]. Fortunately, this is an emergent tendency that can be addressed using the traditional risk scores, but also using innovative predictive algorithms.

During the last years many risk estimation systems have been developed in order to assist clinicians in the risk assessment, and in the individual chances prediction, for CVD development. The major challenges of these tools are the capabilities to: (1) identify high risk individuals, (2) weight the individual effects of all risk factors, (3) stratify or organize who needs lifestyle advice or medical therapy, and finally (4) avoid

* *Corresponding author*. Tel.: +351 239410109.
E-mail address: vaniagalmeida@lei.fis.uc.pt (V.G. Almeida).
http://dx.doi.org/10.1016/j.cmpb.2014.06.010
0169-2607/

overmedicalization of individuals at low risk [6]. Taking these challenges into account several risk factors were identified, by their association with an increased risk for CVD development. CVD risk assessment tools differ from each other on the selected risk factors, the disease for what they were designed (Coronary Heart Disease (CHD), heart failure, *etc.*), the selected event type, the considered period of time (long or short term) and the cohort location. The most popular are Framingham Risk Score (FRS), PROspective CArdiovascular Münster (PROCAM), ASsessing cardiovascular risk using Scottish Intercollegiate Guidelines Network (ASSIGN) and Systematic COronary Risk Evaluation (SCORE) project.

These tools are important to help physicians in their daily practice. However, its application in different populations remains a topic concerning attention. The research needs to be directed at refining the accuracy of prediction models and, most importantly, examining ways of turning them into effective clinical tools. Several risk prediction models for cardiovascular disease are available today and their head to head comparison and application in different populations would benefit from standardized reporting and formal, consistent statistical comparisons. The work presented by [7] reinforces this statement. The limitations of the comparison of different methods are associated to missing information, which makes difficult to reach robust conclusions about the best model or the ranking of models' performance. And, additionally most studies did not statistically compare the models that were examined. The inclusion of standardized reporting of discrimination, calibration, and reclassification metrics with formal statistical comparisons would contribute to the successful application of different risk scores in distinct populations.

The trends for the risk overestimation in low-risk populations and underestimation in high-risk groups have been successfully demonstrated by Cooney et al. [6]. It is known that an examination of 5% SCORE can equate to a 10–25% FRS risk, depending on which of the several FRS functions is selected [3]. Haq et al. [8] studied several methods for risk estimation (FRS, PROCAM, Dundee, and British regional heart-BRHS) and the results demonstrated a close agreement between all these, regarding average risk and showed moderate agreement for estimation among individuals. Finally, it was also concluded that FRS function is acceptably accurate in northern European populations.

The arterial stiffness measurement currently assumes an increasing role in clinical assessment due to its predictive value in cardiovascular events in patients with various risk levels, such as it was demonstrated by several studies [9–12]. There are several advantages of using non-invasive methods over invasive measurements, *e.g.*, the potential use in follow-up strategies in populations without symptomatic CVD, such as children or young adults. Furthermore, non-invasive tools can be essential to the CVD assessment in addition to the established risk factors in populations at high risk aiming the prevention of coronary vascular diseases. Inferences about CVD progressive development can be assessed by the analysis of the mechanical properties of arteries through a variety of indices based on the Pulse Wave Analysis (PWA) [13,14]. The analysis is based on the identification of the key features in the arterial pressure wave profile, such as systolic wave transit time (SWTT), reflected wave transit time (RWTT) and dicrotic notch (evaluated by left ventricular ejection time (LVET)), and can include time or amplitude considerations, as well as variability based parameters [15]. The wave reflections are often addressed, in terms of the augmentation index (AIx), which expresses the ratio of the "augmented pressure" assigned to the reflected wave towards each overall pulse.

Data mining techniques have attracted a great deal of attention due to their ability to extract implicit and potentially useful information from large volumes of data [16]. Their feasible implementation in Computer-Aided Diagnosis (CAD) methodologies has given new insights in the development of innovative and effective decision support systems for CVD premature risk assessment [15,17–19]. An interesting approach is the exploration of different classifiers, as it was proposed by Jovic and Bogunovic [20]. The electrocardiogram (ECG) classification problem was addressed using a combination of several features in the analysis of the Heart Rate Variability (HRV). Other approach presented by Tsipouras et al. [18] was based on the development of a fuzzy rule-based decision support system for CAD diagnosis. On the other hand, multi-classifiers should perform better in some situations, overcoming errors from single classifier analysis [21]. The incorporating of the prediction outcome of each one of the individual classifier was suggested, as a way to reduce the classification errors [22].

On the other hand, clustering analysis is another important branch of unsupervised learning that allows the arrangement of objects into groups (*i.e.*, the clusters), wherein the objects in the same cluster are more similar (in one or more characteristics), than those in different clusters [23]. There is a wide variety of clustering methodologies available in literature, essentially organized in three general classes [24]. The three types include parametric model-based, hierarchical and partitioning algorithms. Shah et al. [25] have proved the usefulness and feasibility of using clustering risk factors in the detection of CVD in youth, by the comparison with the Patholobiological Determinants of Atherosclerosis in Youth (PDAY) risk score. Other studies have also referred the role of clustering methodologies for CVD assessment, such as the work developed by Haseena et al. [26], where a fuzzy C-mean clustered probabilistic neural network for ECG beats discrimination was described. Clustering methodologies were also successful applied in other medical fields, such as in the identification of patterns in blood glucose measurements and regular insulin doses taken before meal time [27].

Our aim is the development of a clustering methodology to deal with arterial pressure waveform (APW) based parameters to cardiovascular risk assessment. The evaluation was performed through the strength of the relationships with traditional risk scores. In the current paper, Section 2 details the subjects and methods used during data analysis, including a quick and up-to-date literature survey on attempts for risk scores and clustering methods used. The results are presented and discussed in Sections 3 and 4, respectively. Finally, in Section 5 some guidelines for further research are presented along with the main conclusions of the current work.