Expert Systems With Applications 92 (2018) 334–349

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system

Paweł Pławiak

Institute of Telecomputing, Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology, Warsaw 24 st., F-5, 31-155 Krakow, Poland

ARTICLE INFO

Article history:
Received 14 June 2017
Revised 9 August 2017
Accepted 9 September 2017
Available online 13 September 2017

Keywords:
ECG
Biomedical signal processing and analysis
Classification
Machine learning algorithms
Neural networks
Support vector machine
K-nearest neighbor algorithm
Evolutionary-neural system
Genetic algorithm
Feature extraction and selection
Discrete Fourier transform

ABSTRACT

This article presents an innovative research methodology that enables the efficient classification of cardiac disorders (17 classes) based on ECG signal analysis and an evolutionary-neural system.

From a social point of view, it is extremely important to prevent heart diseases, which are the most common cause of death worldwide. According to statistical data, 50 million people are at risk for cardiac diseases worldwide. The subject of ECG signal analysis is very popular. However, due to the great difficulty of the task undertaken, and high computational complexity of existing methods, there remains substantial work to perform.

This research collected 1000 fragments of ECG signals from the MIH-BIH Arrhythmia database for one lead, MLII, from 45 patients. An original methodology that consisted of the analysis of longer (10-s) fragments of the ECG signal was used (an average of 13 times less classifications). To enhance the characteristic features of the ECG signal, the spectral power density was estimated (using Welch's method and a discrete Fourier transform). Genetic optimization of parameters and genetic selection of features were tested. Pre-processing, normalization, feature extraction and selection, cross-validation and machine learning algorithms (SVM, kNN, PNN, and RBFNN) were used.

The best evolutionary-neural system, based on the SVM classifier, obtained a recognition sensitivity of 17 myocardium dysfunctions at a level of 90.20% (98 errors per 1000 classifications, accuracy = 98.85%, specificity = 99.39%, time for classification of one sample = 0.0023 [s]). Against the background of the current scientific literature, these results are some of the best results to date.

1. Introduction

Diagnosing heart conditions by analyzing ECG signals has been popular for many years and is the basic method used in the prevention of cardiovascular diseases. The wide range of application of ECG signal analysis is due to the fact that it is a simple and non-invasive method that provides substantial valuable information about the function of the circulatory system.

The huge popularity of the ECG signal analysis is also reflected in research. In recent years, the most developed topics related to electrocardiography include: 1) ECG beat detection / classification: (Augustyniak, 2015; Martis, Acharya, & Adeli, 2014; Martis, Acharya, & Min, 2013; Song, Cho, Kim, & Lee, 2015; Yochum, Renaud, & Jacquir, 2016), 2) deep learning: (Acharya et al., 2017; Kiranyaz, Ince, & Gabbouj, 2016; Rahhal et al., 2016), 3) principal component analysis: (Castells, Laguna, Sörnmo, Bollmann, & Roig, 2007; Ceylan & Ozbay, 2007; Chawla, 2009; Elhaj, Salim, Harris, Swee, & Ahmed, 2016; Kallas, Francis, Honeine, Amoud, & Richard, 2012; Kanaan et al., 2011; Kim, Shin, Shin, & Lee, 2009; Martis, Acharya, Lim, & Suri, 2013; Martis, Acharya, Mandana, Ray, & Chakraborty, 2012; Martis, Acharya, & Min, 2013; Polat & Gunes, 2007; Rodriguez, Mexicano, Bila, Cervantes, & Ponce, 2015; Wang, Chiang, Hsu, & Yang, 2013), 4) higher order statistics: (Martis, Acharya, Mandana, Ray, & Chakraborty, 2013; Martis et al., 2013; Martis, Acharya, Ray, & Chakraborty, 2011), 5) feature selection / dimensionality reduction: (Bereta & Burczyński, 2007; Doquire, de Lannoy, Francois, & Verleysen, 2011; 2011; Kishore & Singh, 2015; Lin, Ying, Chen, & Lee, 2008b; Llamedo & Martinez, 2011; Mar, Zaunseder, Martineznez, Llamedo, & Poll, 2011; Martis et al., 2014; Nasiri, Naghibzadeh, Yazdi, & Naghibzadeh, 2009; Oh, Lee, & Moon, 2004; Wang, Yang, Teng, Xia, & Jensen, 2007; Yeh, Wang, & Chiou, 2010; Yu & Lee, 2012; Zhang, Dong, Luo, Choi, & Wu, 2014), 6) noise: (Li, Rajagopalan, & Clifford, 2014; Pasolli & Melgani, 2015; Roonizi & Sassi, 2016), 7) discrete wavelet transform: (Augustyniak, 2003; Daamouche, Hamami, Alajlan, & Melgani, 2012; Elhaj et al., 2016; Guler & Ubeyli, 2005; Islam, Haque, Tangim, Ahammad, & Khondokar, 2012; Kutlu & Kuntalp, 2012; Lin, Du, & Chen, 2008a;

E-mail addresses: plawiak.pawel@gmail.com, plawiak@pk.edu.pl
URL: https://www.researchgate.net/profile/Pawel_Plawiak

http://dx.doi.org/10.1016/j.eswa.2017.09.022
0957-4174/

Martis, Acharya, & Min, 2013; Mishra, Thakkar, Modi, & Kher, 2012; Thomas, Das, & Ari, 2015; Yan & Lu, 2014), 8) independent component analysis: (Chawla, 2009; Elhaj et al., 2016; Martis, Acharya, & Min, 2013; Sarfraz, Khan, & Li, 2014; Yu & Chou, 2008; 2009), 9) ensemble learning: (Guler & Ubeyli, 2005; Huang, Liu, Zhu, Wang, & Hu, 2014; Javadi, Arani, Sajedin, & Ebrahimpour, 2013; Mert, Kılıç, & Akan, 2012; Osowski, Hoai, & Markiewicz, 2004; Osowski, Markiewicz, & Hoai, 2008; Sambhu & Umesh, 2013), 10) hybrid systems: (Engin, 2004; Meau, Ibrahim, Narainasamy, & Omar, 2006; Osowski & Linh, 2001; Osowski et al., 2008; Ozbay, Ceylan, & Karlik, 2006).

Currently, we observe a very high incidence of cardiovascular disease and the very high mortality caused by them. Despite the preventive measures taken, cardiovascular diseases are the leading cause of death worldwide (17.3 million people per year, accounting for 37% of all deaths (AHA, 2004; 2016; WHO, 2014)) and the most serious and costly health problems facing the world today (Heron & Smith, 2003; National Center for Health Statistics, 2005). Circulatory system diseases are usually chronic diseases that require long-term and expensive treatment. The tendency for the incidence of cardiovascular diseases will increasingly intensify due to the progressive aging of the population (the number of deaths will increase from 17.3 million in 2016 to 23.6 million in 2030 (AHA, 2004; 2016; Healthsquare, 2007; WHO, 2014)).

The classification of cardiac disorders based on existing methods based on the calculation of morphological and dynamic features of individual QRS complexes (heart evolution) is difficult and error prone due to the variability of these features in different patients (Padmavathi & Ramakrishna, 2015). For this reason, solutions currently described in the scientific literature do not achieve a satisfactory efficiency (da S. Luz, Schwartz, Cmara-Chvez, & Menotti, 2016).

The existing approaches are also ineffective for certain cardiac disorders, characterized by complex dependencies between subsequent evolutions of heart, for which the most important are “prolapsed” evolutions of heart (time intervals between subsequent heartbeats) and not the QRS complexes that may be correct. The group of these dysfunctions can include pre-excitation syndromes (e.g., Wolff-Parkinson-White syndrome - WPW), atrio-ventricular and atrial-sinus conduction blocks, and elongate PQ intervals.

This is why it is very important to develop specialized software supporting medical diagnostics to more effectively identify heart pathologies earlier and monitor the conditions of patients in real time. The reduction in computational complexity is also an important aspect in the context of deploying the solution in mobile devices.

For recent years, we can distinguish two main approaches in the literature on the automatic recognition of cardiac disorders based on the analysis of ECG signals:

- classification of QRS complexes (Alvarado, Lakshminarayan, & Principe, 2012; de Chazal, O'Dwyer, & Reilly, 2004; Mateo, Torres, Aparicio, & Santos, 2016; Oster et al., 2015; Ye, Kumar, & Coimbra, 2012a; Zhang & Luo, 2014),
- analysis of longer ECG signal fragments (Abawajy, Kelarev, & Chowdhury, 2013; Padmavathi & Ramakrishna, 2015; Romero & Serrano, 2001; Vafaie, Ataei, & Koofigar, 2014).

It should be noted that the first approach concerning the classification of QRS complexes is substantially more popular. A key element of this approach is the effective detection of QRS complexes. On this basis, it is possible to segment an entire signal into individual QRS complexes and then analyze them using morphological features (determining the shape of the heart evolutions) and dynamic features (determining dependencies between subsequent heart evolutions).

An alternative approach is the analysis of longer, from a single QRS complex (lasting approximately 1 s), signal fragments, usually lasting approximately 10 s; this is the time period corresponding to a standard ECG examination at a cardiologist. Such analysis is based on distinctive feature extraction, for a given disorder, for whole, longer fragments. The identification of heart pathology is based on the extracted features.

Based on a current literature review (Augustyniak & Tadeusiewicz, 2009; da S. Luz, Nunes, de Albuquerque, Papa, & Menotti, 2013; da S. Luz et al., 2016), the typical research methodology in the field of ECG signal analysis consists of

1. obtaining data from public databases (MIT-BIH, EDB, AHA, CU, and NST),
2. pre-processing and signal normalization,
3. QRS detection and ECG signal segmentation,
4. extraction of characteristic signal features and rejection of redundant and erroneous information (extraction and selection of features),
5. classification of QRS complexes (recognition of heart disorders), e.g., data cross-validation, training, testing and optimization of classifier parameters, and
6. evaluation of the obtained results.

In the literature, the most popular method for creating training and test sets is cross-validation, where the two most popular validation schemes are Afkhami, Azarnia, and Tinati (2016) and da S. Luz et al. (2016)

- class-oriented validation schemes (intra-patient paradigm) - the selection of elements for training and test sets based on signals from the same patient, and
- subject-oriented validation schemes (inter-patient paradigm) (de Chazal et al., 2004) - the selection of elements for training and test sets based on signals from other patients.

Designing universal algorithms for the general population, not for an individual person, using a subject-oriented validation scheme is a better solution. This solution demonstrates lower effectiveness on the test set but is more reliable and stable and performs better in practice due to the smaller fit of the models to the training set and better knowledge generalization (Afkhami et al., 2016).

The evolutionary-neural system (Rutkowski, 2008) is a hybrid that combines the advantages of two computational intelligence methods: broadly defined Neural Networks (Prieto et al., 2016) and Evolutionary Computation (Back, Hammel, & Schwefel, 1997). With this synergy, we can achieve greater efficiency through better optimization of the classifier tuning by Genetic Algorithm (Holland, 1992) that are parts of the system. In the field of heart disorders recognition, evolutionary-neural systems are also popular and used with success: (Daamouche et al., 2012; Dilmac & Korurek, 2015; Ince, Kiranyaz, & Gabbouj, 2009; Khazaee & Ebrahimzadeh, 2010; Korurek & Dogan, 2010; Lessmann, Stahlbock, & Crone, 2006; Melgani & Bazi, 2008; Shadmand & Mashoufi, 2016).

The main aims of the research were the following:

Aim 1 Develop new and effective methods for the automatic recognition of myocardium dysfunctions based on ECG signals modeled on the work of cardiologists.

Aim 2 Design algorithms for use in tele-medicine and mobile devices for patient self-control and prevention applications (low computational complexity).

Aim 3 Design universal algorithms not for individuals but for the general population.

Based on a literature review (da S. Luz et al., 2013; da S. Luz et al., 2016), it can be stated that the innovative elements of this research include the following: