Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 59 (2020) 15–26

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Application of multi-feature fusion and random forests to the automated detection of myocardial infarction

Action editor: Rajiv Khosla

Zhizhong Wang [a], Longlong Qian [a], Chuang Han [a], Li Shi [a,b,c,*]

[a] Zhengzhou University, Zhengzhou, Henan Province, China
[b] Tsinghua University, Beijing, China
[c] Beijing National Research Center For Information Science And Technology, Beijing City, Beijing, China

Received 3 March 2019; received in revised form 21 August 2019; accepted 2 September 2019
Available online 5 September 2019

Abstract

Myocardial infarction (MI) was one of the most threatening cardiovascular diseases due to its suddenness and high mortality. Electrocardiography (ECG) reflected the electrophysiological activity of the heart which was widely used for the diagnosis of MI. The aim of the paper was to provide a novel method to detect MI leveraging ECG. Firstly, data enhancement technology was employed to extend the database and prevent overfitting. Then, principal component analysis (PCA) features, statistical features, and entropy features were computed as the representation of first layer features for each lead. Furthermore, the second layer features for each lead were extracted by using random forests (RF), and the feature extraction results were quantified as a classification data set. Finally, in order to evaluate the proposed method, two schemes for the intra-patient and inter-patient were employed. The accuracy, sensitivity, specificity and *F*1 values in the intra-patient scheme were 99.71%, 99.7%, 99.73%, and 99.71%, respectively, and 85.82%, 73.91%, 97.73%, and 83.9% in the inter-patient scheme. Meanwhile, compared with different methods including support vector machine (SVM), back propagation neural network (BPNN), and k-nearest neighbor (KNN), RF displayed the best performance.

Keywords: MI; ECG; Principal component analysis; Statistical feature calculation; Entropy calculation; Random forests

1. Introduction

Disruption of the blood flow to the heart muscle may lead to myocardial infarction (MI) (Rajendra Acharya, Kannathal, Mei Hua, & Mei Yi, 2005). Each year, 210,000 Americans have a recurrent heart attack but around 750,000 suffer heart attacks in total (Mozaffarian et al., 2015), and thus approximately 72% of the heart attacks were silent. Electrocardiogram (ECG) is an inexpensive technique and widely used non-invasive diagnostic tool for cardiopulmonary diseases, it can be employed to monitor the patients' heart-beat and accurately diagnose MI. Fig. 1 shown the ECG waveform (including the P wave, QRS complex wave and T wave) of a healthy control (HC) on the left and a MI on the right. The processes of ECG intelligent diagnosis included noise removal, feature point extraction, feature value selection, dimensionality reduction, and classification (Chang, Lin, Hsieh, & Weng, 2012).

Aiming at the detection of MI, many scholars have carried out this topic. Rajendra Acharya et al. (Acharya et al., 2017) employed deep convolutional neural network (CNN)

* Corresponding authors at: School of Electrical Engineering, Zhengzhou University, 100 Science Avenue, Zhengzhou City, Henan Province, China.

E-mail addresses: wzz1982@zzu.edu.cn (Z. Wang), zdhhan@gs.zzu.edu.cn (C. Han), shilits@mail.tsinghua.edu.cn (L. Shi).

https://doi.org/10.1016/j.cogsys.2019.09.001
1389-0417/

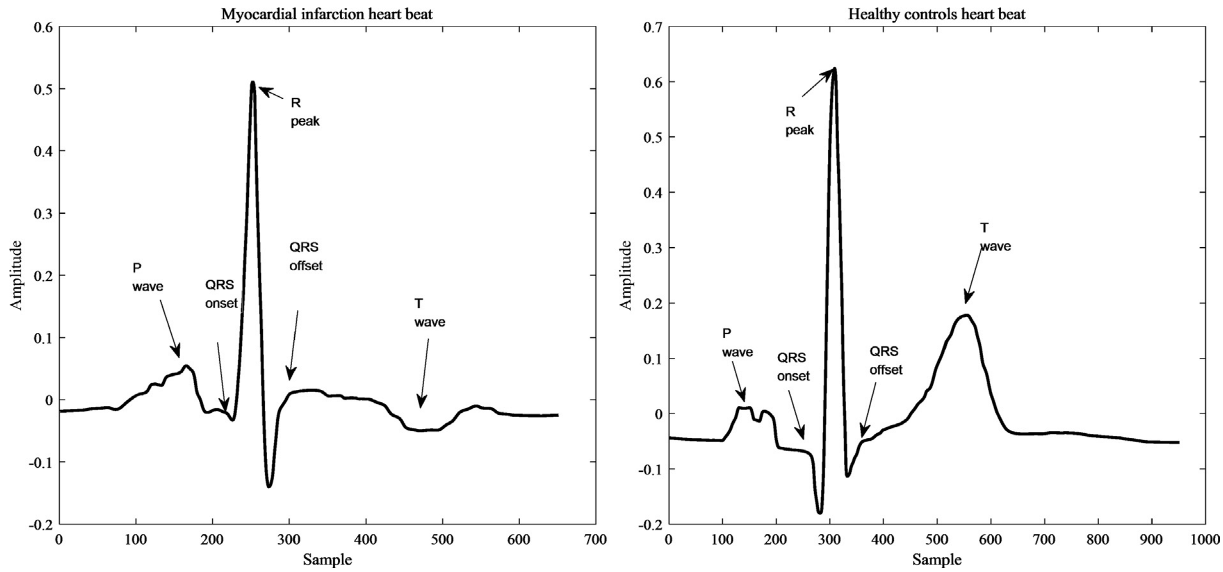

Fig. 1. ECG waveform for HC and MI heartbeats.

for automated detection of HC and MI ECG heartbeats, this method had achieved an detection accuracy of 93.53% have used a noisy dataset and 95.22% with a clean dataset. The Pan-Tompkins algorithm have been employed for R-peak detection in both datasets (with and without noise) (Member, S. IEEE & Pan, 1985). Chang et al. (Chang et al., 2012) have extracted the features from the V1, V2, V3, and V4 leads by hidden Markov models, and Gaussian mixture models have been used to detect the HC and MI ECG beats. Sun et al. (Sun, Lu, Yang, & Li, 2012) have applied latent topic multiple instance learning (MIL) and the support vector machine (SVM) have facilitated the automated diagnosis of MI. The derivative-based algorithm have been employed to detect the characteristic points (such as the R, S and T points) in (Member, S. IEEE & Pan, 1985). In another study (Remya, Indiradevi, & Babu, 2016), discrete wavelet transform (DWT) decomposition have been used to extract specific frequency components as the ECG features. A simple adaptive threshold (SAT) method have been employed to classify multi-lead ECG records. The cross wavelet transform have been used to analysis ECG signals (Banerjee & Mitra, 2014), ECG beat classification have used the threshold-based classifier. In addition, the DWT method have been utilized for denoising and classification (Jayachandran, Joseph, & Acharya, 2010). Noise could be removed by incorporating wavelet shrinkage via soft thresholding (Singh & Tiwari, 2006). The entropy of the wavelet coefficient have been computed, and ECG heart beats have been classified by entropy. Malagore et al. (Arif, Malagore, & Afsar, 2012) have extracted feature point with the wavelet transform; calculated the time domain features, and applyed k-nearest neighbor (KNN) classifier achieved 12-lead ECG beat detection sensitivity of 99.97%, and specificity of 99.9%. Moreover, multiscale wavelet energies and eigenvalues of multiscale have been used to diagnostic features in a previous study (Sharma, Tripathy, & Dandapat, 2015), where SVM with both linear and radial basis function (RBF) kernel and KNN have been used as classifiers. Study (Weng, Lin, Chen, & Chang, 2014) have used polynomial approximation and PCA (Jolliffe, 2002) and the SVM classifier have detected MI. Paper (Chawla, Verma, & Kumar, 2008) achieved heartbeat location by PCA and independent component analysis (ICA). Another study (Nidhyananthan, 2016); R-peak detection with the Pan-Tompkins algorithm (Member, S. IEEE & Pan, 1985); applying the wavelet coefficient have calculated the feature values, and detection MI have used SVM classifier. The wavelet decomposition and eigenspace analysis of 12 lead ECG signals have been used to detect MI (Pereira & Daimiwal, 2016). In another study (Bhaskar, 2015), the features of time domain and wavelet coefficients have been extracted from ECG sianals, PCA and SVM have been used to reduce the dimensionality and classification of the ECG data. In study (Devika, Gopakumar, Aneesh, & Nayar, 2017), ICA (Kužílek, 2013) based on blind source separation (BSS) algorithm (Chibole, 1991; Sugumar, Vanathi, & Mohan, 2014) have been used to denoise, the features of polynomial coefficients have been fed in naive Bayes classifier (Islam, Wu, Ahmadi, & Sid-Ahmed, 2007) that have achieved 12-lead ECG beat classification of MI. In another study (Padhy & Dandapat, 2017), the third-order tensor has been employed to extract the features from multi-lead ECG data, the Pan-Tompkins algorithm (Member, S. IEEE & Pan, 1985) have been used to effectively detecte the R peak. In recent years' studies, in papers (Pławiak & Acharya, 2019; Pławiak, 2018; Pławiak, 2018) the power spectral density as feature have been fed in classifier for heart disease diagnosis. CNN (Hammad, 2019; Yıldırım, Pławiak, Tan, & Acharya, 2018), SVM (Han and Shi, 2019; Hammad, Maher, Wang, Jiang, & Amrani, 2018;