

Automated Diagnosis of Coronary Artery Disease: A Review and Work flow

Abstract

Coronary artery disease (CAD) is the most dangerous heart disease which may lead to sudden cardiac death. However, CAD diagnoses are quite expensive and time-consuming procedures which a patient need to go through. The aim of our paper is to present a unique review of state-of-the-art methods up to 2017 for automatic CAD classification. The protocol of review methods is identifying best methods and classifier for CAD identification. The study proposes two workflows based on two parameter sets for instances A and B. It is necessary to follow the proper procedure, for future evaluation process of automatic diagnosis of CAD. The initial two stages of the parameter set a workflow are preprocessing and feature extraction. Subsequently, stages (feature selection and classification) are same for both workflows. In literature, the SVM classifier represents a promising approach for CAD classification. Moreover, the limitation leads to extract proper features from noninvasive signals.

Introduction

Cardiovascular diseases (CVDs) are the major reasons for mortality around the world. According to World Health Organization (WHO), approximately 17.7 million people died in 2015, representing to 31% of all global death. However, European Heart Network and European Society of Cardiology estimate that over 4 million people died from CVDs in Europe and 1.9 million people died in European Union (EU) which are 47% and 40% deaths, respectively. We all know that the human heart is the most crucial and hardest working organ of the body that combines with blood vessels to form the whole cardiovascular system. CVD is caused by disorders of the heart and blood vessels which result in coronary artery disease (CAD), heart failure, cardiac arrest, and sudden cardiac death. In order to diagnose positive symptoms of CAD, medical specialists prescribed various tests such as angiography, nuclear scan, and C-reactive protein test which are quite expensive and require technical experts; therefore, researchers are seeking interest to develop a less expensive and an effective alternative to the costly prescribed test. In literature, automatic CAD-diagnosing techniques using machine learning algorithms and data mining method have been developed for reducing the medical specialist's efforts and time and save patients' lives and cost. Furthermore, this paper describes a unique review of existing studies found in the literature regarding the identification of CAD symptoms from signal recording and CAD classification using other clinical parameters. The procedure of identifying and classifying CAD diagnosis automatically from noninvasive data is required to follow a proper work direction [1]. In literature, there are two types of studies found, some studies used signal recording to identify CAD symptoms, for instance, electrocardiograph (ECG), photoplethysmography (PPG), and phonocardiography (PCG), and other studies used clinical parameters like age, blood pressure, and smoking habit to classify CAD patients. Therefore, our study proposed two workflows which help to guide the evaluation process of future works. The initial two steps of are preprocessing and feature extraction. These stages are widely used in the literature which used signal recording (noninvasive

John Parker*

Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

*Author for correspondence:
parkerjohnedu345@ma

Received: 04-Jul-2022, Manuscript No. JESTM-22-40111; **Editor assigned:** 07-Jul-2022, PreQC No. JESTM-22-40111 (PQ); **Reviewed:** 21-Jul-2022, QC No. JESTM-22-40111; **Revised:** 27-Jul-2022, Manuscript No. JESTM-22-40111 (R); **Published:** 31-Jul-2022, DOI: 10.37532/jestm.2022.14(4).58-63

data) to identify CAD symptoms. The techniques utilized in first two stages directly influence the classification results; therefore, it is necessary to choose the technique carefully. Subsequently, the remaining two stages of workflows (feature reduction and classification) are the same for both studies. Note that workflows for evaluation process of future works are a significant contribution to this review work. In the recent studies, we find a review of signal recording pattern recognition and classification techniques based on nonlinear transformation. Rajkumar et al. performed an extensive review and comparative analysis of methods used for CAD classification [2]. In specific, they did not focus on the workflows for the evaluation process of future direction and their study only reviews the state-of-the-art classifiers. However, our study focuses on more up-to-date literature review. Furthermore, our study performs a special review of existing classification methods.

Data Acquisition

Various databases are developed for the heart disease and arrhythmia classification which allow the researcher to evaluate their methods on the standardized database. There are few datasets which are more commonly used in studies for CAD and its risk identification.

The benchmark database of PhysioNet contains 86 lengthy ECG recordings of 80 human subjects. In literature, studies consider 23 subjects of this database which are only affected by CAD.

The database was collected from the Cleveland Clinic Foundation. It has 76 parameters, out of which only 14 parameters were selected for use. The selected attributes represent the clinical and noninvasive test results of 303 patients who are undergoing angiography. Removing the cases containing missing values, 270 cases were considered in studies, out of which 120 cases were identified as patients with CHD while 150 cases were diagnosed as patients without CHD [3].

Some of the studies use ECG signals of 10 CAD patients from the IQRAA Hospital, Calicut, Kerala, India. BIOPAC™ equipment was used to record the ECG signals at a sampling rate of 500 Hz. All the CAD subjects participated in

the studies were on similar medication. The age of all the subjects was under the ranges between 40 and 70 years.

The MIMIC II database contains two types of ICU patient records: waveform dataset and clinical dataset. The waveform dataset contains physiological signal recording (such as ECG, PPG, and arterial blood pressure (ABP), and clinical dataset contains clinical data which are collected by hospital staff.

Methodology Procedure

CAD is caused by atherosclerosis of the coronary arteries that leads to formation of barrier on the blood flow to the heart which may be diagnosed using clinical data of the patient such as blood pressure, age, gender, smoking habit, and random blood sugar and identifies symptoms of CAD using ECG. However, existing studies follow different procedures for diagnosing CAD using machine learning methods and data mining techniques. However, we proposed a common workflow for the new researcher for their work evaluation. Moreover, are based on those studies which utilize raw signals as a dataset and clinical dataset for diagnosing CAD, respectively. CAD classification using raw signal dataset has more steps to diagnose as compared to furthermore; our study reviewed each stage separately [4].

Contaminated recordings were the major problem of detecting coronary artery disease; however, studies used a different method to preprocess data prior to feature extraction step. This section reviews those preprocessing techniques which were used in the context of coronary artery disease detection.

Davari Dolatabadi et al. and Patidar et al. used a low-pass filter and a high-pass filter with a cutoff frequency for removing 20 Hz noise and 0.3 Hz noise, respectively, whereas a 50 Hz notch filter is used to remove power source interference and this filter is also called band-rejection filter [5]. However, the Pan-Tompkins algorithm is used to analyze the R peak for the measurement of two consecutive beats RR interval and QRS detection. The Pan-Tompkins method is widely used in literature because it is simple and easy to implement.

Kumar et al. used baseline wander for low

and high cutoff frequency of 0.3 Hz and 15 Hz, respectively. Similarly, the study used notch filter and Pan–Tompkins methods to eliminate 50 Hz cutoff frequency and identify R-peaks separately.

Ukil et al. proposed the methodology for cleaning PPG signals for CAD detection. The multistage method is used to analyze the presence of noise in the signal. In the first part, the study used dynamic time wrapping technique for segmentation. Secondly, the Hampel filter is used to remove the noise from the signal.

Contrasting with previous mentioned techniques, discretization techniques were presented in for the parameter intervals. Discretization is a process of dividing the continuous parameter in a discretized variable for classification of coronary artery disease parameters [6]. However, this technique had a direct impact on the performance of classification method which is widely used for data analysis.

This stage is one of the major keys to the success route of CAD detection. The feature extraction is the process of revealing clinical features from the signal's morphology in the time and frequency domains. This phase of classification is only used by those studies which refer raw ECG signal as a dataset. According to time, frequency, and nonlinear dynamic features are used to demonstrate CAD patient and non-CAD patients. For the measurement of frequency domain features, the study employed autoregressive (AR) modeling-based method to calculate power spectrum density; AR spectrum is the most popular method for HRV analysis, and this algorithm has the capability to be factorized into separate spectral components. AR model is more complex, and it has the contingency of negative components in spectral factorization. Subsequently, for time domain calculation, the authors used statistical features and geometrical features like SD, RMSSD, and HRV triangular index, respectively.

SVM, NN, KNN, and RF classifiers were used for heart failure. They performed comparison analysis between classifiers and disclosed that the RF classifier stands out with 100% accuracy. Furthermore, the RF classifier successfully achieved significant advantages among other implemented classifiers in

the study. The RF classifier was applied by to differentiate normal and abnormal heartbeats and successfully achieved 92.2% and 93% success, respectively. The decision tree classifiers are non-parameteric supervised learning technique used for classification and regression. The aim of this technique is to create a model that predicts the value of a target variable by learning simple decision rules. Baihaqi et al. performed an experimental research to diagnose CAD using [7], and they successfully obtained accuracy of 78.95%. However, studies reveal that the classifier is not a promising approach for continuous features. Thus, a technique that uses the classifier considers only small dataset. For instance, in the large dataset was considered for the detection of heart disease; in that case, the RF classifier performed better than. Meanwhile, the combination technique was used in, and the authors noticed that the bagged decision tree classifier obtained remarkable progress to discriminate the classes of the feature set. According to the methodology was developed using the Gaussian mixture model (GMM) unsupervised learning of classification where it returns remarkable performance with 99.42% accuracy over ten folds [8]. The study also revealed that using the GMM low probability error of 3.0700×10^{-5} was obtained as an upper bound on the classification error. The GMM is used in to re-estimate attributes from the dataset in order to calculate the class mean and covariance matrix for a priori probability estimation [9]. The GMM model is used for testing the data and noting the performance of unsupervised learning, while classifying CAD resulted in the highest accuracy of 96.8%. Clustering methods are unsupervised learning that are widely used with different supervised learning algorithms in recent studies. According to the combined the NN classifier with unsupervised learning methods taking into account that the accuracy obtained is much better than single NN classifier. Other studies in the same direction used clustering technique with a linear discriminate classifier for heart patients, and it reveals remarkable performance results which are reliable and efficient for real-world application.

Discussion

Data mining techniques play a major role in

medical systems, which will provide the major contribution to enhance the medical field. This study presents coronary artery disease classification review on different methods of data mining and artificial intelligence. Furthermore, we observed that from literature, there are two types of parameters used in CAD classification. However, we represent features in this study as parameter set A and parameter set B which is signal features (Table 1) and patient clinical data (Table 2) respectively [10]. However, (Table 3) describes review of state-of-the-art classifiers and their effectiveness. Furthermore, we proposed two workflows (Figures 1 and 2) for the evaluation process of future works for beginners in this field (Figure 1). That preprocessing and feature extraction are the most important phases for parameter set A, and depicts that it is not necessary to go through preprocessing and feature extraction stages for parameter set B (Figure 2). However, remaining flows are the same for both diagrams [11]. In literature, we found that there is still a room for improvement in CAD classification. ECG is a noninvasive

Table 1. Parameters of ECG.

Features	Description
SDNN	Standard deviation of normal RR intervals
SDSD	The standard deviation of successive RR interval difference
RMSSD	Square root of the mean of the sum of the squares differences between adjacent normal intervals
QRS duration	Area under peak
Mean	Average values

Table 2. Patient clinical data.

Features	Description	Ranges
Age	Age (in years)	30–86
Gender	1: male; 0: female	0–1
HTN	Hypertension, 0: no; 1: yes	0–1
RBS	Random blood sugar	57–180
Chest pain type	0: nonspecific chest pain	0–2
	1: atypical chest pain	
	2: typical angina	
HT	Height (cm)	133–188
WT	Weight (kg)	33–110
DBP	Diastolic blood pressure (mmHg)	46–110
SBP	Systolic blood pressure (mmHg)	100–170
CAD	Coronary artery disease	0: no; 1: yes

Table 3. Review of state-of-the-art classifiers and their effectiveness.

Work	Feature set	Classifiers	Effectiveness
[8]	A	Optimized SVM	Accuracy=99.2% Sensitivity=98.43% Specificity=100%
[66]	B	NN	Accuracy=88.4%
[10]	A	KNN	Accuracy=96.8% Sensitivity=100% Specificity=93.7%
[9]	A	LS-SVM	Accuracy=99.7% Sensitivity=99.6% Specificity=99.8%
[27]	A	SVM	Accuracy=79.71%
[7]	A	LS-SVM	Accuracy=100%
[19]	B	Fuzzy rule	Accuracy=84% Sensitivity=79% Specificity=89%
[32]	B	Fuzzy rule	Accuracy=92.8%
[58]	B	Fuzzy rule	Accuracy=81.2%
[67]	B	Fuzzy rule and ensemble classifier	Accuracy=84.44%
[55]	A	Random forest	Sensitivity=80% Specificity=90%
[44]	A	SVM with RBF	Sensitivity=73% Specificity=87%
[45]	A	SVM	Sensitivity=85% Specificity=78%

technique used to diagnose CAD patients, and ECG signal does not provide the proper information that is required though it is necessary to obtain accurate feature from the ECG. This limitation may also lead to a serious heart disease. Therefore, a suitable method for hidden factor extraction from ECG signal is very intricate due to the irregular shape of bio signals. Some studies like reported that feature extraction method is unable to calculate accurate values of unmasked attributes of the ECG signal. Furthermore, the usage of the small dataset for classification may diagnose misclassification and it is also necessary to avoid small dataset for classification in order to overcome the error rate [12].

Conclusion

In this paper, we reviewed automated CAD classification state-of-the-art methods. In literature, we found that SVM classifier performance is better than another classifier for automated detection of CAD. Our study proposed two workflows for parameter sets A

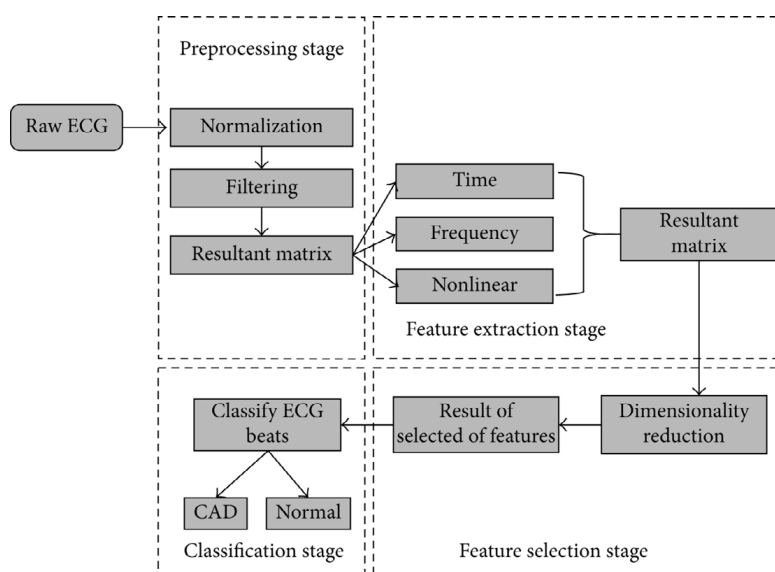


Figure 1. Workflow for parameter set A.

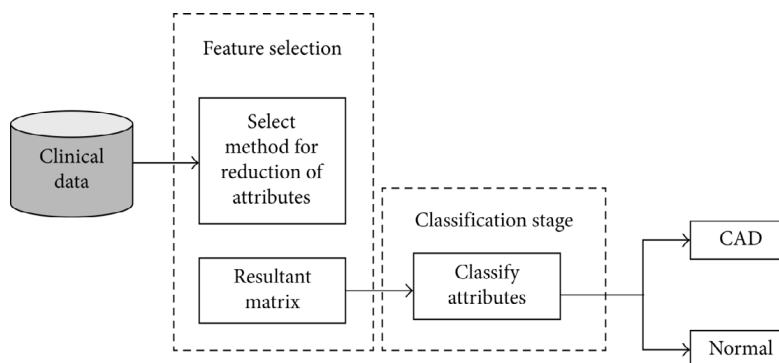


Figure 2. Workflow for parameter set B.

and B in which we analyzed those two stages are most important while using parameter set A. Furthermore, we also suggest that performance of the classifier also relies on dataset's nature and size.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Tsipouras MG, Exarchos TP, Fotiadis DI et al. Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling. *IEEE Tran Inf technol biomed.* 12, 447–458 (2008).
2. Wong ND Epidemiological studies of CHD and the evolution of preventive cardiology. *Nat Rev Cardiol.* 11, 276–289 (2014).
3. Acharya UR, Faust O, Sree V et al. Linear and nonlinear analysis of normal and CAD-affected heart rate signals. *Comput Methods Programs Bio.* 113, 55–68 (2014).
4. Kumar M, Pachori RB, Rajendra Acharya U et al. An efficient automated technique for CAD diagnosis using flexible analytic wavelet transform and entropy features extracted from HRV signals. *Expert Syst Appl.* 63, 165–172 (2016).
5. Davari Dolatabadi A, Khadem SEZ, Asl BM et al. Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. *Comput Methods Programs Bio.* 138, 117–126 (2017).
6. Patidar S, Pachori RB, Rajendra Acharya U et al. Automated diagnosis of coronary artery disease using tunable-Q wavelet transform applied on heart rate signals. *Knowl Based Syst.* 82, 1–10 (2015).
7. Giri D, Acharya UR, Martis RJ et al. Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform. *Knowl Based Syst.* 37, 274–282 (2013).
8. Maglaveras N, Stamkopoulos T, Diamantaras K et al. ECG pattern recognition and classification using nonlinear transformations and neural networks: a review. *Int J Med Inform.* 52,191–208 (1998).
9. Rajkumar R, Anandakumar K, Bharathi A et al. Coronary artery disease (CAD) prediction and classification-a survey. *Breast Cancer.* 90, 945-955 (2006).
10. Goldberger AL, Amaral LA, Glass L et al. PhysioBank, physioToolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation.* 101, 215–220 (2000).

11. Lahasna A, Ainon RN, Zainuddin R *et al.* Design of a fuzzy-based decision support system for coronary heart disease diagnosis. *J Med Syst.* 36, 3293–3306 (2012).
12. Kumar SU, Inbarani HH Neighborhood rough set based ECG signal classification for diagnosis of cardiac diseases. *Soft Comput.* 21, 1–13 (2016).