Classifying Heart Sounds using Images of MFCC and Temporal Features

Diogo Marcelo Nogueira[1], Carlos Abreu Ferreira[2], and Alípio M. Jorge[3]

[1] INESC TEC, Portugal
diogo.m.nogueira@inesctec.pt
[2] Instituto Politécnico do Porto and INESC TEC, Portugal
cgf@isep.ipp.pt
[3] FCUP-Universidade do Porto and INESC TEC, Portugal
amjorge@fc.up.pt

Abstract. Phonocardiogram signals contain very useful information about the condition of the heart. It is a method of registration of heart sounds, which can be visually represented on a chart. By analyzing these signals, early detections and diagnosis of heart diseases can be done. Intelligent and automated analysis of the phonocardiogram is therefore very important, to determine whether the patient's heart works properly or should be referred to an expert for further evaluation. In this work, we use electrocardiograms and phonocardiograms collected simultaneously, from the Physionet challenge database, and we aim to determine whether a phonocardiogram corresponds to a "normal" or "abnormal" physiological state. The main idea is to translate a 1D phonocardiogram signal into a 2D image that represents temporal and Mel-frequency cepstral coefficients features. To do that, we develop a novel approach that uses both features. First we segment the phonocardiogram signals with an algorithm based on a logistic regression hidden semi-Markov model, which uses the electrocardiogram signals as reference. After that, we extract a group of features from the time and frequency domain (Mel-frequency cepstral coefficients) of the phonocardiogram. Then, we combine these features into a two-dimensional time-frequency heat map representation. Lastly, we run a binary classifier to learn a model that discriminates between normal and abnormal phonocardiogram signals.
In the experiments, we study the contribution of temporal and Mel-frequency cepstral coefficients features and evaluate three classification algorithms: Support Vector Machines, Convolutional Neural Network, and Random Forest. The best results are achieved when we map both temporal and Mel-frequency cepstral coefficients features into a 2D image and use the Support Vector Machines with a radial basis function kernel. Indeed, by including both temporal and Mel-frequency cepstral coefficients features, we obtain sligthly better results than the ones reported by the challenge participants, which use large amounts of data and high computational power.

Keywords: Phonocardiogram, Electrocardiogram, Mel-frequency cepstral coefficients, Time Features, Classification

1 Introduction

Cardiovascular diseases (CVD) are the single leading cause of death worldwide. According to the estimates of the World Health Organization (WHO) in 2012, CVD account for approximately 17.5 million deaths worldwide, which corresponds to over 31% of all deaths globally. These facts alone show that CVD are a major global threat and any development to aid the prevention of such diseases is of great importance [24]. According to the latest statistics, 20% of people aged over 40 develop heart failure during their life. This condition is the number one reason for hospitalization among those over 65. Half of all patients die within 5 years of diagnosis, and each year heart failure costs the global economy $108 billion, with hospitalizations accounting for 60-70% of direct treatment costs. 14.9 million people in the EU and 5.7 million in the United States have heart failure; the impact on the rest of the world is not sufficiently documented [24].

A more pro-active approach involving low cost cardiac health screening of the general population can help the physician detect possible complications at an early stage. Currently, two effective cardiac screening methodologies are the Electrocardiogram (ECG) and echocardiogram exams but these can be expensive for mass screening and require technical expertise that is not always available. Despite remarkable advances in imaging technologies for the heart, the clinical evaluation

of cardiac defects by auscultation has remained a main diagnostic method for congenital heart diseases. In experienced hands the method is effective, reliable, and cheap.

The auscultation of the heart and lungs with a stethoscope is often conducted on patients thought to have cardiac or pulmonary disease, before recommending additional diagnostic procedures, treatment, or no further action [15]. Because this process is simple, cheap, and quick to detect diseases, the stethoscope still maintains a key position in medicine in the modern era. However, auscultation is a subjective process that depends on the experience and hearing capability of the individual, a feature that may lead to a large variability in findings. The poor sensitivity of human hearing in the low frequency range of the heart sounds makes this task even more difficult. Also, the human hearing system is better at detecting frequency changes than intensity changes. The physical limitations of the human ear make it unable to analyze all the information contained in the acoustic signals of the heart [26].

Physically, a stethoscope covers a broad sound spectrum and the average frequency depends on the point of auscultation. It requires significant practice for a human ear to distinguish between them. The existence of methods that can automatically and successfully analyze heart signals, can be used as a diagnostic tool to help determine if an individual should be referred for expert diagnosis, specially in cases where access to clinicians and medical care is limited.

In this work, we develop a novel approach to classify heart sounds, which uses both temporal and Mel-frequency cepstral coefficients (MFCC) features. The main idea is to translate a 1D Phonocardiogram (PCG) signal into a 2D image that represents temporal and MFCC features. First we segment a PCG signal, using the ECG that was simultaneously recorded, to identify the S1 heart sound. Second, we extract temporal and MFCC features from PCG signals. Third, we combine these features into a 2D image. Last, we run a binary classifier to classify each image as either being normal or abnormal. Our method is evaluated using ECG and PCG signals that were made available in the 2016 PhysioNet Computing in Cardiology Challenge [12], [21].

The remainder of this paper is organized as follows. In Section 2, we present a brief description of ECG and PCG heart signals, stressing their main characteristics and how they relate to each other. In Section 3, we discuss the current challenges in the study of heart signals and the related work, including different methods and approaches that can be used to extract features and classify heart sounds. Section 4 introduces our methodology, and details each step of our procedure. In Section 5 we present and discuss the obtained results. At the end of the paper we present the main conclusions of this work.

2 Characteristics of ECG and PCG heart signals

The heart is one of the most important organs in our body. The heart beats continuously to pump oxygen and nutrient-rich blood throughout the human body to sustain life. A human heart is made of a strong muscle called myocardium, and is divided into chambers. The two upper chambers are known as atria while the two lower chambers are known as ventricles. The heart beats in regular intervals, controlled by the electrical pulses generated from the sinus node near the heart. This organ is susceptible to a variety of pathologies. One of the techniques used to detect these pathologies is the ECG, which consists in the recording of the variation of bioelectric potentials versus time of human heartbeats [5]. Another technique that can be used to verify the existence of pathologies is the PCG. During the squeezing of the blood from chamber to chamber, the valves keep the blood flowing smoothly in and out of the heart. This is done by automatically opening the valves, to let the blood flow from chamber to chamber, and closing the valves to prevent the backflow of blood [28]. PCG is a graphical representation of the waveform of heart sounds, which are generated by: (1) opening or closing of the heart valves, (2) flow of blood through the valve orifice, (3) turbulence created when the heart valves snap shut, and (4) rubbing of cardiac surfaces. The PCG creates a visual recording of these events and allows the detection of sub-audible heart-sounds and murmurs. This technique is very useful because it contains a great amount of physiological and pathological information regarding the human heart and vascular system.

2.1 ECG Signal

The ECG is a powerful diagnostic tool for heart disease. It can provide accurate information on the functional aspects of the heart and the cardiovascular system. The ECG signal is formed by a set of waves, such as the P-wave, representing the atrial depolarization, the QRS wave, which represents

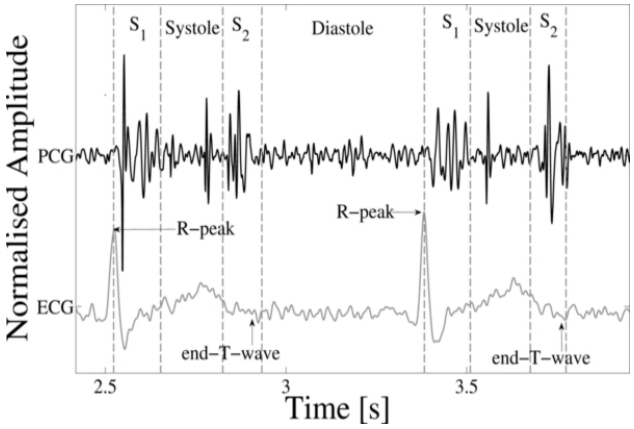

Fig. 1. Example of an ECG-labeled PCG, with the ECG and four states of the heart cycle (S1, Systole, S2, Diastole) shown. The R-peak and end-T-wave are labeled as reference for the approximate positions of S1 and S2, respectively [30].

the depolarization of the ventricles [12], and the T-wave, which corresponds to the repolarization of the ventricles. The QT interval is the most important region for the detection of abnormality, each change that affects these characteristics represents a cardiac abnormality [12]. The ECG waveform is illustrated in the bottom of Figure 1.

2.2 PCG Signal

The top of the Figure 1 shows the heart sounds, composed by four different sounds: S1, S2, S3 and S4. The pumping action of a normal heart is audible by the 1st heart sound (S1) and the 2nd heart sound (S2). During systole, the atrioventricular valves are closed and the blood tries to flow back to the atrium, causing back bulging of the AV valves. This leads to vibration of the valves, the blood and the walls of the ventricles, and corresponds to the 1st heart sound. During diastole, the blood in the blood vessels tries to flow back to the ventricles, causing the semi lunar valves to bulge, but the elastic recoil of the arteries makes the blood bounce forward, thus leading to vibration of the blood, the walls and the ventricular valves, which produces the 2nd heart sound. S1 is a low-pitch sound with longer duration, whereas S2 is a high-pitch sound with a shorter duration. In normal situations, the S1–S2 interval (systole) is shorter than the S2–S1 interval (diastole). The 3rd heart sound (S3) is heard in the mid diastole due to the blood that fills the ventricles. The 4th heart sound (S4), also known as atrial heart sound, occurs when the atrium contracts and pumps blood to the ventricles. S4 appears with a low energy and is almost never heard by the stethoscope [11]. In addition to these components of the normal heart sounds, a variety of other sounds, such as heart murmurs, may be present in the cardiac signal. Murmurs can be benign (physiological) or abnormal (pathological), and are usually caused by turbulent blood-flow, which can happen inside or outside the heart. Abnormal murmurs can occur due to stenosis, which restricts the opening of a heart valve, or to regurgitation related with valves insufficiency, which allows backflow of blood following the partial closure of an inept valve.

2.3 Relationship between ECG and PCG signals

PCG can provide quantitative and qualitative information of heart sounds and murmurs. Studies on heart sound detection can be divided into two categories: ECG signal-dependent and ECG signal-independent. Our study is ECG signal-dependent. The opening and closing of the cardiac valves, and the sounds they produce, are the mechanical events of the cardiac cycle. They are preceded by the electrical events of the cardiac cycle. In Figure 1 we plot part of both ECG and PCG signals to illustrate the relationship between them in the time domain. In this figure we can see that S1 occurs 0.04s to 0.06s after the onset of the QRS complex, and that S2 occurs towards the end of the T wave. Heart sound segmentation refers to the detection of the exact positions of the first (S1) and second (S2) heart sounds in a PCG. This is an essential step in the automatic analysis of heart sound recordings, as it allows the analysis of the periods between these sounds for the presence of clicks and murmurs. The segmentation becomes a difficult task if the PCG

recordings are corrupted by in-band noise. In our work, we have a set of PCG signals with the corresponding ECG signals collected simultaneously. Therefore, we identify the start of the S1 using the ECG signal and then use this knowledge to segment the PCG. In particular, we use the Springer's segmentation algorithm [30] to identify the fundamental heart sounds (S1, Systole, S2 and Diastole) in the PCG waveform.

3 Current challenges and related work

3.1 Current challenges

As the quality and availability of PCG signals is no longer an issue, the development of appropriate algorithms that are able to detect heart diseases from heart sounds is an important challenge that has become the focus of work for many researchers. The ability to mathematically analyze and quantify the heart sounds represented on the PCG provides valuable information regarding the condition of the heart [25]. Thus, automated analysis and characterization of the PCG signal plays a vital part in the diagnosis and monitoring of valvular heart diseases. The main problems concerning the development of relevant techniques are the wide variety of distinguishable pathological heart sounds and the non-stationary characteristics of the PCG signals. Considering these issues, a question that can be addressed is how to increase the variety of distinguishable heart sounds while improving the performance of such systems in terms of reducing their computational complexity, without compromising their precision. PCG signal processing can be crudely divided into two main research areas. One is focused in the detection of events such as S1 and S2 to perform the segmentation of the PCG. The other deals with the detection of murmurs and, consequently, of cardiac pathologies [8].

3.2 Related work

The segmentation process of PCG signals is a very important task to perform murmur detection and diagnosis of cardiac pathologies with computer analysis. Thus, it is essential that different components of the heart cycle can be timed and separated [10]. A large variety of algorithms that perform PCG segmentation have been presented in the literature. A solution for segmentation based in the time-domain characteristics of the PCG, was presented in [13], and another, based in the frequency-domain characteristics, in [17]. A threshold based on Shannon energy is set to detect peaks that correspond to S1 and S2 [16]. Correlation techniques have been used in [32], but this method may not perform well when the duration and the spectra of sound signal components show huge variations, making impossible to run this technique without user intervention.

In [23], as the heart sound and ECG signals are time varying, the Instantaneous Energy is computed to characterize the temporal behaviors of these signals. The purpose of the study is to perform heart sound segmentation based on the Instantaneous Energy of the ECG. Another important step in signal processing is feature extraction. If the features are not chosen properly, the performance of any classifier will be poor. The objective of the feature extraction procedure is to find the features from the available data and use them later for classification. These features were extracted from different analysis domains to ensure that the segments were described as thoroughly as possible. The analysis of heart sound is difficult to perform in the time domain because of noise interference and the overlapping of heart sound components. Thus, in many cases the processing of heart sound signals is done in the frequency domain. There are a large number of feature extraction algorithms available. These include the Fourier transform [19], the short time Fourier transform (STFT) [6], the time-frequency representation (TFR) [2], the MFCC [9] and the Discrete Wavelet Transform (DWT) coefficients [3]. However, the most widely used algorithms are the MFCC and the DWT. In this work we extract MFCC and time-frequency features from the signal.

After extracting the features from each signal, in a classification problem we need to learn a model that discriminates between normal and abnormal heart sounds. Most of the previous studies that learn models to classify heart sounds use artificial neural networks (ANN) or Support Vector Machines (SVM) [4]. In [14], Gupta et al. addressed the problem of distinguishing between two abnormal and one normal heart states. The methodology used by these authors uses Wavelet analysis of the PCG signal in combination with homomorphic filtering and K-means clustering method. The generalization accuracy of the proposed methodology was 97%. One of the first reported studies using neural networks for classification was presented by Barschdorff et al. [4].

These authors discussed the advantages of using neural networks over traditional classifiers, such as nearest neighbors. Spectral features obtained from short-term Fourier transform (STFT) analysis of the signal and mean values of corresponding sections of the signal envelope were used to train the neural network. Another algorithm that was widely used, and is known to generate highly accurate models, is the SVM classifier. An approach for heart sounds identification presented by Wu et al. reached a generalization accuracy of 95%. This approach uses wavelet transform to extract the envelope of the PCG signals [33]. Almost the same results were obtained by Jiang and Choi [34], who developed a system based in clustering algorithms for in-home use. However, this system was proven only by a case study. Another approach is the use of the tools and techniques of deep learning for the automated analysis of heart sounds [27]. In this paper, an algorithm was presented that accepts PCG waveforms as input and uses a deep convolutional neural network architecture to discriminate between normal and abnormal heart sounds.

4 Methodology

The methodology that we developed in this work is a novel approach to classify heart sounds. We use ECG and PCG collected simultaneously, to identify the fundamental heart sounds and thus segment the PCG signals more accurately. The main idea is to translate a 1D PCG signal into a 2D image that represents temporal and MFCC features. The methodology has four main steps:

- Segmentation of the PCG signal, identifying their four heart sounds states;
- Feature extraction of a group of 8 features in time domain, and MFCC in frequency domain;
- Transformation of the extracted features into two dimensional heat maps, which capture the time-frequency distribution of signal energy;
- Classification of the images generated, using different classifiers, distinguishing between normal and abnormal heat maps.

Each component is described below in detail. To better present each step of the methodology, we explore the dataset that was made available at the 2016 Physionet/ Computing in Cardiology Challenge [21]. In this challenge the goal was to discriminate between normal and abnormal hearts using PCG and ECG signals.

4.1 Heart sound database

The challenge database provides a large collection of heart sound recordings, obtained from different real-world clinical and nonclinical environments. They include clean heart sounds but also very noisy recordings. The data were recorded from both normal and pathological subjects, and from both children and adults. We only use the training set A, which contains a total of 400 heart sound recordings (PCG signals lasting from 5 seconds to just over 120 seconds) and 400 ECG signals collected at the same time. The heart sound recordings were divided into two types: normal and abnormal heart sound recordings. The former were recorded from healthy subjects and the latter from patients with a confirmed cardiac diagnosis. These patients suffered from a variety of illnesses, but a more specific classification of the abnormal recordings was not provided. It is noteworthy that the number of normal recordings does not equal that of abnormal recordings, i.e., the dataset used is unbalanced. The distribution of the two classes in the dataset is approximately 70% of normal recordings and 30 % of abnormal recordings. More detailed information about the dataset can be found in [21].

4.2 Segmentation

The first step of our method is to segment the PCG. In this work, the segmentation of the PCG was performed with the Springer's segmentation algorithm [30]. This algorithm is based on a logistic regression hidden semi-Markov model to predict the most likely sequence of states by incorporating information about the expected duration of each heart sound state. By applying this segmentation algorithm (which uses the ECG signals as reference, as explained above) to the PCG signals, we were able to identify the beginning and end of the four fundamental heart sound states (S1, Systole, S2 and Diastole). In our approach, we divide the original PCG signals into shorter segments. Using the information obtained with the segmentation algorithm, we selected the beginning of each heartbeat

(S1) as a starting point for each segment that would be created. This was performed to ensure that sequences were aligned during classification. After the S1 heart sound was identified, we decided to create segments with a period of three seconds. We then extracted overlapping segments, and produced a total of 13404 segments of three seconds from the original 400 PCG signals. In Figure 3 we can see the result of applying the segmentation algorithm.

4.3 Feature Extraction

After the segmentation step, we have a signal that is partitioned into several segments, with the heart sound states identified. Now we need to extract a set of features that describe each portion of the PCG signal.

In this step, two types of features are extracted from the heart sound signal. We extract a set of time domain features and MFCC features from the frequency domain. In the next lines we present the extracted features.

Time Features

After the segmentation of the PCG, and the identification of the four fundamental states of the heart cycle, some features were extracted. Currently, we are using eight time-domain features:

- Average duration of states S1;
- Average duration of states Systole;
- Average duration of states S2;
- Average duration of states Diastole;
- Average duration of the intervals RR;
- Ratio between the duration of the Systole and the RR period, of each heart beat;
- Ratio between the duration of the Diastole and the RR period, of each heart beat;
- Ratio between the duration of the Systole and the Diastole, of each heart beat.

MFCC features

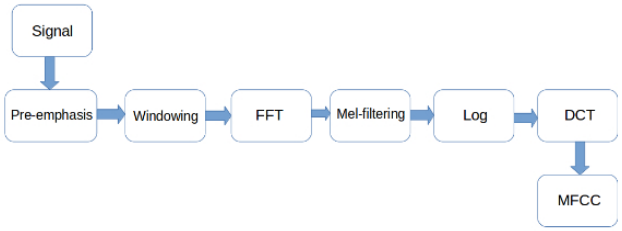

Fig. 2. MFCC feature extraction process.

We use the MFCC to extract features from the audio signal. The MFCC is a linear representation of the cosine transforms of a short duration of logarithmic power spectrum of the sound signal on a non-linear scale Mel frequency [20]. It perceives frequency in a logarithmic way, inspired in the behavior of the human ear. It is a powerful signal processing algorithm, widely used in the field of sound recognition. The advantage of extracting MFCC parameters is that all features of the sound signal are concentrated in the first coefficients, thus facilitating the extraction task for operations in clustering algorithms or sound recognition [18]. Obtaining the MFCCs involves analyzing and processing the sound, according to the following steps: pre-emphasis, windowing, fast Fourier transform (FFT), Mel-filtering, nonlinear transformation, and discrete cosine transform (DCT). The stages of MFCC coefficient extraction are shown in Figure 2. The pre-emphasis operation enhances the received signals to compensate for signal distortions. The windowing operation divides a given signal into a sequence of frames. The FFT operation is applied to the windowed signals for spectral analysis. The Mel-filtering operation is designed based on human perception, and it integrates the frequency compositions from one Mel-filter band into one energy intensity.

The non-linear transformation operation takes the logarithm of all Mel-filter band intensities. The transformed intensities are then converted into MFCC using DCT. The computation of the MFCC includes Mel-Scale filter-banks, an they are computed as follows [22]:

$$m = 1127 log_e \left(\frac{f}{700} + 1 \right) \tag{1}$$

where f is the frequency in the linear scale and m is the resulting frequency in Mel-Scale. The power spectral density (PSD) of the spectrum is mapped onto the Mel-Scale by multiplying it with the filter-banks constructed earlier, and the log of the energy output of each filter is calculated as follows [22]:

$$s[m] = log_e \left(\sum_{k=10}^{N-1} |X[k]|^2 H_m[k] \right) \tag{2}$$

where $H_m[k]$ is the filter-banks and m is the number of the filter-bank. To obtain the MFCC, the discrete cosine transform (DCT) of the spectrum is computed [22]:

$$c[n] = \sum_{m=0}^{N-1} S[m] cos \left(\frac{\pi n}{M} \left(m - \frac{1}{2} \right) \right), n = 0, 1, 2....M \tag{3}$$

where M is the total number of filter banks.

In our case, in the windowing stage, we run overlapping sliding windows over the segments of three seconds that were created in the segmentation process. We chose a window length of 25 ms and a step size of 10 ms. By applying the described procedure, we calculated a total of 12 MFCC filterbanks per sliding window, which makes a total of 300 time frames for each signal of three seconds.

4.4 Transformation of features in images

At this stage, for each segment of three seconds of signal, we have a total of eight features extracted in the time domain and a collection of 3600 cepstral coefficients resulting from the 12 MFCC filterbanks and 300 time frames. We joined the two sets of features. To adjust the dimensions, zero padding was performed. The merge of the features produced an array with 12 rows and 301 columns.

Figure 3 illustrates one segment of three seconds from the original one-dimensional PCG waveforms, with the identification of the heart sound states calculated during signal segmentation. In addition, the heat map resulting from the conversion of the extracted features is also shown. The heat map has a total of 12 rows by 301 columns. The features from the time domain are in column number 301; the first 300 columns correspond to MFCC features. In these, the horizontal axis represents the sliding window and the vertical axis presents the 12 filterbank frequencies that were used in the calculation of the MFCC. We have also done some experiments, using the two feature sets separately, to evaluate the impact that their joint use has on the results.

4.5 Classification of heart sound images

The aim of the classification procedure is to develop a rule whereby any new observation, represented by a feature vector, can be classified into one of the existing classes.

In this step, we run an algorithm to learn a classification model that is able to discriminate between normal and abnormal heart sounds. We learn the classification model by using the heart sound images obtained in the previous step.

There is a wide variety of classification methods applied in several areas, including the study of cardiac signals (ECG and PCG). Some of the classifiers used in this area are SVM [31], K-nearest neighbor(kNN) classifiers [22], Gaussian Mixture Model (GMM) [22], and several types of Neuronal Networks (NN) [7]. In our methodology, we study and evaluate several algorithms: the SVM, the Random Forest, the K-means Clustering and the Convolutional Neural Network (CNN), using as input parameters, the images created from the extracted features. While the CNN allows the use of images directly as an input parameter, for the other classifiers it is necessary to convert the image in a vector line, so that it can be used as an input parameter.

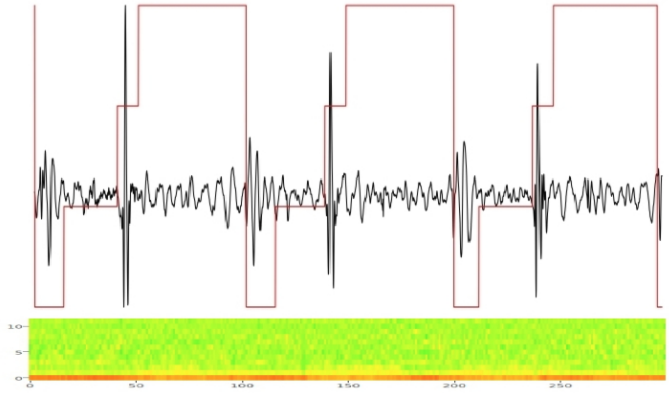

Fig. 3. Example of a PCG with the four states of the heart cycle (S1, systole, S2, diastole) identified (red line). MFCC heat map visualization of 3-second segment of heart sound data.

SVM

In our model, we used an SVM with a radial basis function (RBF) kernel. The RBF kernel has the formula:

$$K\left(x^{(i)}, x^{(j)}\right) = \phi(x^{(i)})^T \phi(x^{(j)}) \tag{4}$$

$$K\left(x^{(i)}, x^{(j)}\right) = exp\left(-\gamma \parallel x^{(i)} - x^{(j)} \parallel^2\right), \gamma > 0 \tag{5}$$

where the $x^{(i)}$ and $x^{(j)}$ represents two features vectors in some input space. The γ factor is a free parameter. In our method we adjust the γ parameter and the cost value, in order to optimize the results, avoiding falling in overfitting.

Convolutional Neural Networks

With the transformation of the extracted features into images, the CNN was chosen to perform the training image classifier, given their ability to automatically learn appropriate convolutional filters. We decided to train a CNN, using the features images as inputs. The architecture, and the parameters selected, were based on the work of Rubin et al. [27], who built a PCG signal classifier using deep convolutional neural networks.

Random Forest

The Random Forest is a classification method that works by creating an ensemble of decision trees at training time.

In our case, we compose a random forest with a number of trees and a number of variables randomly sampled as candidates at each split, in order to optimize the results, avoiding falling in overfitting.

K-means Clustering

Cluster Analysis is a process of aggregating the objects into various groups on the basis of their similarities. K-Means algorithm is one of many methods used to perform clustering that is included in a group of unsupervised methods.

In this study, we tried to form two clusters, in order to divide the signals into the two existing classes, normal and abnormal. We used the Euclidean Distance method to measure the shortest distance between several signals. We also defined how many random sets were chosen, in order to optimize the results.

4.6 Evaluation metrics

In the classification process, the 13404 images generated from the extracted features were classified. Once the dataset used was unbalanced, consisting of approximately 70 % normal segments and 30 % abnormal segments, we performed a 10-fold stratified cross validation. In a typical (k-fold)

cross-validation method, a dataset S is first randomly partitioned into k equally-sized, disjoint subsets (folds) $S_1, S_2, ..., S_n$. Each k fold is then in turn used as the test set, while the remaining $(k-1)$ folds are used as the training set. A classifier is then constructed from the training set, and its accuracy is evaluated on the test set. This process repeats k times, with a different fold used as the test set each time. The estimated true accuracy by this method is the average over the k folds. One distinct feature of cross-validation is that all the k test sets are disjoint, and thus each case in the original training set is tested once and only once. An extension of regular cross-validation is stratified cross-validation. In k-fold stratified cross-validation, a dataset S is partitioned into k folds such that each class is uniformly distributed among the k folds. The result is that the class distribution in each fold is similar to that in the original data set S. In this sense, the partition is "balanced" in terms of class distributions. In contrast, regular cross-validation randomly partitions S into k folds without considering class distributions. A possible scenario with regular cross-validation is that a certain class could be distributed unevenly (some folds contain more cases of the class than other folds). This distortion in class distributions can cause a less reliable accuracy estimation [35].

Given that the classification models are trained using the images generated from the segments of three seconds signal, it was necessary to group the predictions of the various segments to classify the original PCG signals. In the evaluation of the classification of the segments belonging to the same signal, a metric was used, in which only the signals with more than 60% of the segments classified as normal, would be classified as normal.

Once the normal PCG signals came from healthy subjects and the abnormal ones from patients with a confirmed cardiac diagnosis, the labels of the signals were assigned taking into account the patients' medical history, and not through the analysis of signals by a physician. This fact may lead to the existence of signals with the abnormal label that do not present the characteristics to be integrated in this class, in the whole signal studied, or in some of the segments of three seconds studied. By using this metric, we are considering that it is possible for an abnormal signal to contain segments of three seconds that are classified as normal. Different values were applied for this metric, and the best results were obtained for the 60%.

Equations (6), (7) and (8) show the sensitivity, specificity and overall metrics, respectively, which were used to evaluate the results. The measures were defined using True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN):

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

$$Overall = \frac{Sensitivity + Specificity}{2} \tag{8}$$

5 Experimental results

In this work we explore the dataset that was made available at Physionet databases. The goal is to discriminate between normal and abnormal hearts using PCG and ECG signals. We present a new approach, whose main idea is to translate a 1D PCG signal into a 2D image that represents temporal and MFCC features. After that, we use a binary classifier to learn a model that discriminates between normal and abnormal PCG signals. This algorithm was developed using the R statistical package. Next, we present and discuss the obtained results.

5.1 Results

During classification, the number of MFCC features to be used was treated as a hyper-parameter. Several tests were performed, with a different number of MFCC features, in order to optimize the classification results. Table 1 shows the results obtained with the different approaches, using 5 and 6 MFCC features, performed with the SVM, Random Forest (RF), K-means and CNN. The set of features composed by 5 MFCC and the TF, showed better results than the set composed by 6 MFCC and the TF (with the exception of CNN). Among the classification algorithms, the best results were obtained with SVM, followed by RF, K-means and CNN, in this order, as can be seen in Table 1.

Type of features	Classifier	Sensitivity	Specificity	Overall
5 MFCC + TF	SVM	0.9187	0.8205	0.8696
6 MFCC + TF	SVM	0.9081	0.8034	0.8558
5 MFCC + TF	RF	0.9789	0.4017	0.6903
6 MFCC + TF	RF	0.9823	0.3418	0.6621
5 MFCC + TF	K-means	0.7456	0.5556	0.6506
6 MFCC + TF	K-means	0.7420	0.5556	0.6488
5 MFCC + TF	CNN	0.8622	0.1538	0.5080
6 MFCC + TF	CNN	0.1343	0.9487	0.5415

Table 1. Results obtained in the various tests performed.

Table 2 shows some results of the experiments where we studied the contribution of mapping temporal features to improve the results obtained with the MFCC features alone. In this table we present the results obtained with the classifiers SVM, RF, K-means and CNN, for two sets of features, one composed by 4 MFCC only and another composed by 4 MFCC and TF.

Type of features	Classifier	Sensitivity	Specificity	Overall
4 MFCC + TF	SVM	0.9647	0.7265	0.8456
4 MFCC	SVM	0.9435	0.7094	0.8264
4 MFCC + TF	RF	0.9788	0.4188	0.6988
4 MFCC	RF	0.9647	0.4017	0.6832
4 MFCC + TF	K-means	0.7951	0.4615	0.6283
4 MFCC	K-means	0.8021	0.4274	0.6147
4 MFCC + TF	CNN	0.5724	0.5812	0.5768
4 MFCC	CNN	0.5018	0.5641	0.5329

Table 2. Results obtained, with or without inclusion of time features.

5.2 Analysis of the results

As already mentioned, the number of MFCC features used during classification was treated as a hyper-parameter, and several experiments with a different number of MFCC have been performed. The best results obtained are presented on tables 1 and 2. As can be seen in the tables, the best results were obtained with the SVM radial basis. The best result was obtained with a set of five MFCC features, together with the temporal features, with which a sensitivity of 0.9187, a specificity of 0.8205 and an overall of 0.8696 were obtained. With the RF and K-means classifiers, the results are not so good, because these classifiers have very low Specificity values, which consequently reduces the Overall. This was due to the high number of false positives returned by the classifier, as a consequence of this being the minority class of the dataset (approximately 30%), and the classifier having a lower recognition rate of the signals belonging to this class. Regarding the results obtained with the CNN, they fall below those obtained by Rubin et al. [27]. This can be related with the amount of data explored: in this work we were able to use only 10% of the dataset used by Rubin et al. As is well known, CNN performance is heavily related with the amount of data used to learn the network. Typically, if more data is used to train, the better the results will be [29]. In our work, we only used a small portion of the dataset due to the computational power that was available.

In Table 2 we present a set of results performed with the various classifiers, in which two types of features were used. In one case 4 MFCC were used together with the temporal features and in the other case only 4 MFCC were used without the temporal features. Analyzing the results, it is possible to conclude that the use of time features together with the MFCC, presents better results than using the 4 MFCC alone, with cases where the overall gain is approximately 0.04.

The overall scores for the top entries of the PhysioNet Computing in Cardiology challenge were very close [1]. The difference between the top place finisher overall (0.8602) and the 10th place (0.8263) was just approximately 0.04. Although the dataset we use is only part of the one used in the challenge, the class distribution is similar. Our best result is about 0.01 higher than the winner of the challenge. The use of time features, along with the MFCC, had a fundamental role

in the obtained results, as it was demonstrated in the analysis of Table 2. Their presence led to an improvement of the results in all the classifiers used and, in the case of SVM, led to an improvement of approximately 0.02. Furthermore, our best performance was achieved using a single SVM radial basis, whereas other top place finishers of the challenge achieved strong classification accuracies with an ensemble of classifiers. In practical terms, a system that relies on only a single classifier, as opposed to a large ensemble, has the advantage of limiting the amount of computational resources required for classification.

6 Conclusions

We have used a SVM radial basis algorithm in the classification of heart sounds as normal or abnormal, obtaining an accuracy of approximately 86.97%. The approach included the segmentation of the heart signal, identifying the four states of the heart cycle, and creating three second signal segments. From these segments we extracted a group of MFCC features, which capture the time-frequency distribution of signal energy, and a group of eight temporal features. The group of features of each segment of three seconds were converted into an image, in the form of heatmap, which was the input to our classifier - a SVM radial basis. The performance of the model was evaluated and compared to other classifiers. The proposed approach outperforms all the other classifiers, achieving 86.97% accuracy in the binary classification task of identifying normal and abnormal heart sounds. Another classifier used was the CNN, which had worse results than the other classifiers. One possible cause for this is the small size of the dataset used, since this algorithm requires a large volume of data to converge. In the future we will investigate the usage of our methodology in larger datasets, and explore other types of features (wavelets). The analysis of the results showed that the unbalanced dataset might be problematic for identifying the minority class, and the results could be improved by collecting more training data, and by balancing the dataset. Furthermore, we intend to use CNN in larger datasets, in order to take full advantage of its ability.

Acknowledgments

This work is supported by the *NanoSTIMA Project: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016* which is financed by the *North Portugal Regional Operational Programme (NORTE 2020)*, under the *PORTUGAL 2020 Partnership Agreement*, and through the *European Regional Development Fund (ERDF)*.

References

1. Physionet Challenge 2016, "classification of normal/abnormal heart sound recordings". https://www.physionet.org/challenge/2016/papers/
2. Avendaño-Valencia, L.D., Godino-Llorente, J.I., Blanco-Velasco, M., Castellanos-Dominguez, G.: Feature extraction from parametric time–frequency representations for heart murmur detection. Annals of Biomedical Engineering 38(8), 2716–2732 (2010)
3. Balili, C.C., Sobrepena, M.C.C., Naval, P.C.: Classification of heart sounds using discrete and continuous wavelet transform and random forests. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). pp. 655–659 (Nov 2015)
4. Barschdorff, D., Bothe, A., Rengshausen, U.: Heart sound analysis using neural and statistical classifiers: a comparison. In: [1989] Proceedings. Computers in Cardiology. pp. 415–418 (Sep 1989)
5. Boussaa, M., Atouf, I., Atibi, M., Bennis, A.: Ecg signals classification using mfcc coefficients and ann classifier. In: 2016 International Conference on Electrical and Information Technologies (ICEIT). pp. 480–484 (May 2016)
6. Boutana, D., Benidir, M., Barkat, B.: Segmentation and identification of some pathological phonocardiogram signals using time-frequency analysis. IET Signal Processing 5, 527–537 (September 2011)
7. Chen, T.E., Yang, S.I., L. T. Ho, e.a.: S1 and s2 heart sound recognition using deep neural networks. IEEE Transactions on Biomedical Engineering 64(2), 372–380 (Feb 2017)
8. Choi, S., Jiang, Z.: Cardiac sound murmurs classification with autoregressive spectral analysis and multi-support vector machine technique. Computers in Biology and Medicine 40(1), 8 – 20 (2010)
9. Colonna, J., Peet, T., Ferreira, C.A., Jorge, A.M., Gomes, E.F., Gama, J.a.: Automatic classification of anuran sounds using convolutional neural networks. In: Proceedings of the Ninth International C* Conference on Computer Science & Software Engineering. pp. 73–78. C3S2E '16, ACM (2016)

10. El-Segaier, M., Lilja, O., Lukkarinen, S., Slrnmo, L., Sepponen, R.: Computer-based detection and analysis of heart sound and murmur. Annals of Biomedical Engineering 33(7), 937–942 (2005)
11. Ergen, B., Tatar, Y., Gulcur, H.O.: Time–frequency analysis of phonocardiogram signals using wavelet transform: a comparative study. Computer Methods in Biomechanics and Biomedical Engineering 15(4), 371–381 (2012)
12. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation 101(23), e215–e220 (2000)
13. Groch, M.W., Domnanovich, J.R., Erwin, W.D.: A new heart-sounds gating device for medical imaging. IEEE Transactions on Biomedical Engineering 39(3), 307–310 (March 1992)
14. Gupta, C.N., Palaniappan, R., Swaminathan, S., Krishnan, S.M.: Neural network classification of homomorphic segmented heart sounds. Applied Soft Computing 7(1), 286 – 297 (2007)
15. Hanna, I.R., Silverman, M.E.: A history of cardiac auscultation and some of its contributors. The American Journal of Cardiology 90(3), 259–267 (Aug 2002)
16. Huiying, L., Sakari, L., Iiro, H.: A heart sound segmentation algorithm using wavelet decomposition and reconstruction. In: Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE. vol. 4, pp. 1630–1633 vol.4 (Oct 1997)
17. Iwata, A., Ishii, N., Suzumura, N., Ikegaya, K.: Algorithm for detecting the first and the second heart sounds by spectral tracking. Medical and Biological Engineering and Computing 18(1), 19–26 (1980)
18. Kishore, K.V.K., Satish, P.K.: Emotion recognition in speech using mfcc and wavelet features. In: 2013 3rd IEEE International Advance Computing Conference (IACC). pp. 842–847 (Feb 2013)
19. Kumar, D., Carvalho, P., Antunes, M., Paiva, R.P., Henriques, J.: Heart murmur classification with feature selection. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. pp. 4566–4569 (Aug 2010)
20. Lalitha, S., Geyasruti, D., Narayanan, R., M, S.: Emotion detection using mfcc and cepstrum features. Procedia Computer Science 70, 29 – 35 (2015)
21. Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., Castells, F., Roig, J.M., Silva, I., Johnson, A.E.W., Syed, Z., Schmidt, S.E., Papadaniil, C.D., Hadjileontiadis, L., Naseri, H., Moukadem, A., Dieterlen, A., Brandt, C., Tang, H., Samieinasab, M., Samieinasab, M.R., Sameni, R., Mark, R.G., Clifford, G.D.: An open access database for the evaluation of heart sound algorithms. Physiological Measurement 37(12), 2181 (2016)
22. Lubaib, P., Muneer, K.A.: The heart defect analysis based on pcg signals using pattern recognition techniques. Procedia Technology 24, 1024 – 1031 (2016)
23. Malarvili, M.B., Kamarulafizam, I., Hussain, S., Helmi, D.: Heart sound segmentation algorithm based on instantaneous energy of electrocardiogram. In: Computers in Cardiology. pp. 327–330 (Sept 2003)
24. Mozaffarian, D., Benjamin, E.J., Go, Alan S., e.a.: Heart disease and stroke statistics—2016 update. Circulation (2015)
25. Obaidat, M.S.: Phonocardiogram signal analysis: techniques and performance comparison. Journal of medical engineering & technology 17 6, 221–7 (1993)
26. Rangayyan, R., Lehner, R.: Phonocardiogram signal analysis: a review. Crit Rev Biomed Eng. 15(3), 211–236 (1987)
27. Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., Sricharan, K.: Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients. In: Computing in Cardiology Conference (CinC), 2016. pp. 813–816. IEEE (2016)
28. Santos, M.A.R., Souza, M.N.: Detection of first and second cardiac sounds based on time frequency analysis. proceedings of the 23rd annual EMBS international conference (October 2001)
29. Shi, W., Gong, Y., Wang, J.: Improving cnn performance with min-max objective. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. pp. 2004–2010. IJCAI'16, AAAI Press (2016)
30. Springer, D.B., Tarassenko, L., Clifford, G.D.: Logistic regression-hsmm-based heart sound segmentation. IEEE Transactions on Biomedical Engineering 63(4), 822–832 (April 2016)
31. V.Rathikarani, P.Dhanalakshmi: Automatic classification of ecg signal for identifying arrhythmia. International Journal of Advanced Research in Computer Science and Software Engineering 3(9) (September 2013)
32. White, P.R., Collis, W.B., Salmon, A.P.: Time-frequency analysis of heart murmurs in children. In: IEE Colloquium on Time-Frequency Analysis of Biomedical Signals (Digest No. 1997/006). pp. 3/1–3/4
33. b. Wu, J., Zhou, S., Wu, Z., m. Wu, X.: Research on the method of characteristic extraction and classification of phonocardiogram. In: 2012 International Conference on Systems and Informatics (ICSAI2012). pp. 1732–1735 (May 2012)
34. Z, J., S, C.: A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope. In: Expert Systems with Applications. vol. 31, pp. 286–298 (2006)
35. Zhang, Y.D., Yang, Z.J., Lu, H.M., Zhou, X.X., Phillips, P., Liu, Q.M., Wang, S.H.: Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. IEEE Access 4, 8375–8385 (2016)