J Med Syst (2019) 43: 168
https://doi.org/10.1007/s10916-019-1286-5

IMAGE & SIGNAL PROCESSING

Classifying Heart Sounds Using Images of Motifs, MFCC and Temporal Features

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

Received: 21 December 2017 / Accepted: 9 April 2019 / Published online: 6 May 2019

Abstract

Cardiovascular disease is the leading cause of death in the world, and its early detection is a key to improving long-term health outcomes. The auscultation of the heart is still an important method in the medical process because it is very simple and cheap. To detect possible heart anomalies at an early stage, an automatic method enabling cardiac health low-cost screening for the general population would be highly valuable. By analyzing the phonocardiogram signals, it is possible to perform cardiac diagnosis and find possible anomalies at an early-term. Therefore, the development of intelligent and automated analysis tools of the phonocardiogram is very relevant. In this work, we use simultaneously collected electrocardiograms and phonocardiograms from the Physionet Challenge database with the main objective of determining whether a phonocardiogram corresponds to a "normal" or "abnormal" physiological state. Our main contribution is the methodological combination of time domain features and frequency domain features of phonocardiogram signals to improve cardiac disease automatic classification. This novel approach is developed using both features. First, the phonocardiogram signals are segmented with an algorithm based on a logistic regression hidden semi-Markov model, which uses electrocardiogram signals as a reference. Then, two groups of features from the time and frequency domain are extracted from the phonocardiogram segments. One group is based on motifs and the other on Mel-frequency cepstral coefficients. After that, we combine these features into a two-dimensional time-frequency heat map representation. Lastly, a binary classifier is applied to both groups of features to learn a model that discriminates between normal and abnormal phonocardiogram signals. In the experiments, three classification algorithms are used: Support Vector Machines, Convolutional Neural Network, and Random Forest. The best results are achieved when both time and Mel-frequency cepstral coefficients features are considered using a Support Vector Machines with a radial kernel.

Keywords Phonocardiogram · Electrocardiogram · Mel-frequency cepstral coefficients · Motifs · Time features

Introduction

Based on data from the World Health Organization relative to 2012, cardiovascular diseases are the main cause of death for approximately 17.5 million people around the world. In total, this leads to 31% of all deaths. Therefore, any contribution to help prevention of these diseases is highly valuable [22] As stated in [22], 20% of people aged over 40 fall ill with heart failure during their lifetime. For people over 65-year-old, this is the main reason for their hospitalization. Fifty percent of all patients die within 5 years of diagnosis. Heart failure costs $108 billion, with hospitalizations accounting for 60-70% of direct treatment costs. Another fact is that 14.9 million people in the EU and 5.7 million in the US have heart failure. The statistics for the rest of the world are not adequately documented.

To support the physicians in the detection of any possible complication at an early stage, a dynamic method that enabled low-cost cardiac health screening for the general population would be highly valuable. Presently, there are two effective cardiac screening procedures: the Electrocardiogram (ECG) and echo-cardiogram exams. The disadvantage of these methods are being expensive for mass screening and requiring technical expertise which is not continually available. Even with impressive progression in imaging technologies for the heart, the principal diagnostic

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Diogo Marcelo Nogueira
diogo.m.nogueira@inesctec.pt

Extended author information available on the last page of the article.

method for congenital heart disease is still the clinical evaluation of cardiac defects by auscultation. Although this is a reliable, and relatively inexpensive method, the auscultation is a subjective process which highly depends on the experience and expertise of the doctor and his/her hearing capability. Considering the weakness of human hearing to detect intensity changes and its sensitivity in the low-frequency range makes the diagnosis task even more troublesome. These limitations of the human ear make it infeasible to extract all the information from the heart signal [25]. The presence of automated methods as a diagnostic tool can be beneficial where access to the clinicians and medical care are limited.

In this paper, we propose a novel method for heart sounds classification using Mel-frequency cepstral coefficients (MFCC), motifs and time features. The main contribution is the methodology for transforming a one-dimensional Phonocardiogram (PCG) signal into a two-dimensional image that represents temporal and frequency features. In this direction, firstly, the PCG signal is divided into several segments using an ECG signal, which is recorded simultaneously, to identify the four heart sound states. Consequently, the time features are calculated, and the MFCC and motifs features are extracted. Finally, several binary classifiers are used to classify the images into two classes: Normal and Abnormal. Our method is evaluated by using ECG and PCG signals that were made available in the 2016 PhysioNet Challenge [19].

The remainder of this paper is organized as follows. In "Characteristics of ECG and PCG heart signals", a brief description of ECG and PCG heart signals is presented, stressing their main characteristics and how they relate to each other. In "State of the art", the current challenges in the study of heart signals and the related work are discussed, including different methods and approaches that can be used to extract features and classify heart sounds. In "Methodology", our methodology is described. In "Experimental results" and "Conclusions", results and discussion are presented, respectively. Finally, the paper is concluded in the last section.

Characteristics of ECG and PCG heart signals

One of the most important organs of a human body is the heart. It beats constantly to pump blood through the circulatory system. The heart is divided into four chambers: upper left and right atria; and lower left and right ventricles. The heart pumps blood with a rhythm determined by a group of pace-making cells in the sinoatrial node. This organ is vulnerable to a variety of diseases. The ECG signal consists of the recording of the variation of bioelectric potentials versus time of human heartbeats [4] and can be used to detect any damage to the heart's muscle cells or conduction system.

Another useful signal to detect problems in the heart is the PCG signal which is the recording of all the sounds made by the heart during a cardiac cycle. Graphically, it represents the waveforms of heart sounds which are generated by (1) opening/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.

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 the depolarization of the ventricles [19], and the T-wave, which corresponds to the re-polarization 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 [19]. The ECG waveform is illustrated in the bottom of Fig. 1.

PCG signal

The top of Fig. 1 shows the heart sounds, composed of four different heart sound states: S1, Systole, S2, and Diastole.

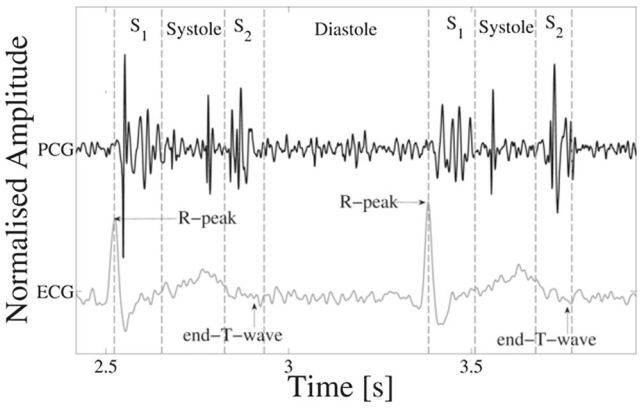

Fig. 1 Example of an ECG-labeled PCG, with the ECG and four states of the heart cycle (S1, Systole, S2, Diastole) shown. The R-peak and end-T-wave are labeled as reference for the approximate positions of S1 and S2, respectively [29]

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 semilunar 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) [9]. 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. Abnormal murmurs can occur due to stenosis, which restricts the opening of a heart valve, or to regurgitation related with valves insufficiency, which allows back-flow of blood following the partial closure of an inept valve.

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 Fig. 1 we plot part of both ECG (the bottom of figure) and PCG signals (the top of the figure) to illustrate the relationship between them in the time domain. The S1 occurs 0.04s to 0.06s after the onset of the QRS complex and the S2 occurs towards the end of the T wave [27]. 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 [29] to identify the four heart sound states (S1, Systole, S2 and Diastole) in the PCG waveform.

State of the art

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 [24]. Some cardiovascular conditions are well-reflected in the heart sound before their signatures appear in other signals such as ECG [13]. Thus, automated analysis and characterization of the PCG signal play a vital part in the diagnosis and monitoring of 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. The challenge currently faced by the scientific community is to find approaches with reasonable computational cost that can improve the performance in automatic diagnostic.

Related work

The approaches for heart sound classification usually comprises three main stages: the first stage is focused on the detection of events such as S1 and S2 to identify individual cardiac cycles, and perform the segmentation of the PCG; the second stage corresponds to the feature extraction that provides the input for the last stage: feature classification.

The segmentation process of PCG signals is a very important task to perform diagnosis of cardiac pathologies with computer analysis. Thus, it is essential that different components of the heart cycle can be timed and separated. A large variety of algorithms that perform PCG segmentation have been presented in the literature. One of the most robust techniques to perform heart sound segmentation is using ECG gating. As was said above, there is a direct relationship between ECG and PCG main components. Correlation techniques have been used in [30], but this method shows limitations when the duration and the spectra of sound signal components show huge variations. It therefore requires user intervention. In [21], as the heart sound and ECG signals are time varying, the Instantaneous Energy is computed to characterize the temporal behavior of these signals and perform segmentation.

Regarding the second stage, the feature extraction, the features can be obtained using different analysis domains to ensure that the segments were described as thoroughly as possible. Since the analysis of HS is difficult to perform in the time domain due of noise interference and

the overlapping of HS components, feature extraction is typically done over the frequency domain. Important feature extraction methods include the Fourier Transform [16], the MFCC [8], motifs [11] the Discrete and Continuous Wavelet Transform coefficients [2]. Another type of features, which have already shown good results [10], are the features of the PCG signal, collected in the time domain.

In this work, we will collect a group of time features from signals, together with the MFCC and the motifs.

After extracting the features from each signal, in a classification problem we need to learn a model that discriminates between the different classes of our signals. Most of the previous studies that learn models to classify heart sounds use Random Forest (RF) or Support Vector Machines (SVM) [3]. The SVM algorithm is known to generate highly accurate models. An approach for heart sound classification presented by Wu et al. reached an accuracy of 95%. This approach uses wavelet transform to extract the envelope of the PCG signals. The envelope is used to achieve the accurate position of S1 and S2. After that, they use as features, the area of PCG envelope and the wavelet energy, and their goal is to determine if the heart sounds are "normal" or "abnormal" [31]. Although they achieved good results, the methodology was implemented in a dataset with only 35 PCG signals.

In [23], Nogueira et. al uses ECG and PCG collected simultaneously, with the aim of distinguishing if the PCG corresponds to a "normal" or "abnormal" physiological state. To do that, they segment the PCG signals into the fundamental HS using Springer's segmentation algorithm. After, they extract time domain and MFCC features before performing classification. In the classification process, they evaluate three algorithms, and the best results (accuracy of 86.96%) is obtained with the SVM with radial basis function kernel. They explore the dataset from the 2016 Physionet/ Computing in Cardiology Challenge.

In [2], an integrated approach to heart sound classification using wavelet analysis and RF classifiers is proposed. The heart sounds were first segmented through detection of the S1 and S2 heart sounds using Shannon energy. Time and frequency based features derived from Discrete and Continuous Wavelet Transforms were used as feature vectors for the RF classifier that categorized heart sounds into four classes (Normal, Murmur, Ex-trasystole, and Artifact). The proposed approach uses data obtained from the 2011 PASCAL Classifying Hearts Sounds Challenge, and performed better than previous related studies. Another approach is presented in [11], where the authors use motifs (frequent subsequences) as features in cardiac audio time series. The general idea is to find frequent motifs in a discretized version of the audio time series using a frequent pattern mining algorithm. Such discovered motifs are regarded as features. The results have demonstrated that these features contain valuable information for discrimination tasks.

As can be seen from the list of works mentioned above, despite achieving good accuracy results so far, there is still space to test new features in combination with well-performing classifiers in order to improve not only accuracy, but also develop approaches with reasonable computational cost.

Methodology

In this work we explore the dataset from the 2016 Physionet/ Computing in Cardiology Challenge [19]. The goal is to discriminate between normal and abnormal heart conditions using PCG and ECG signals. For this, we propose a novel methodology to classify heart sounds. We use ECG and PCG collected simultaneously. The ECG are important to identify the fundamental heart sounds and segment the PCG signals more accurately. Afterwards, we obtain for each PCG signal values of temporal features, MFCC features and motif based features. These numerical values are then arranged into two-dimensional images that represent the one-dimensional signal.

The methodology has five main steps:

1. Segmentationcrt sounds states;
2. Feature extraction of a group of eight features in time domain, and the features in the frequency domain - MFCC and motifs;
3. Combination of the two types of features collected (from the time and frequency domain) in one image, for each of the PCG signal segments;
4. Classification of the images generated, using different classifiers, distinguishing between normal and abnormal images;
5. Aggregate the segment classification, to classify the original signal.

The methodology is described below in detail.

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. We only use the training set A because this is the only one that has ECG and PCG signals, which allows us to apply our methodology. The remaining training sets that have been made available only have PCG signals.

The heart sound recordings were divided into two types: normal and abnormal heart sound recordings. The former was 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. This fact will contribute to an abnormal class of signals, where we find signals that will originate very distinct features for the same class of signals. It is noteworthy that abnormal recordings are substantially more frequent than normal recordings, i.e., the dataset used is unbalanced in terms of classes. The distribution of the two classes in the dataset is approximately 70% of abnormal recordings and 30% of normal recordings. The platform that hosts the database used has developed a page where you can view and compare signals from different classes. Just by visualizing some of the signs studied, you may notice the complexity of the problem that we are trying to solve. There are some recordings being corrupted by various sources of noise, such as talking, dogs barking and children playing. Other noise sources included stethoscope motion, breathing and intestinal sounds [1]. More detailed information about the dataset can be found in [19].

Segmentation

The first step of our method is to segment the PCG. In this work, the segmentation of the PCG was performed with Springer's segmentation algorithm [29]. 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 a reference, as explained above) to the PCG signals, we were able to identify the beginning and the end of the four fundamental heart sound states (S1, Systole, S2, and Diastole).

The original PCG signals are divided into short segments. Using the information obtained with the segmentation algorithm, the beginning of each heartbeat (S1) is selected as a starting point for each segment. This was performed to ensure that sequences were aligned during the classification. Starting from the S1 state, each segment has a length of three seconds. A total of 13404 segments of three seconds are produced from the original 400 PCG signals. In Fig. 4 we can see the result of applying the segmentation algorithm.

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 collect a set of features which describe each segment of the PCG signals.

Totally, three types of features were extracted from the heart sound signals: time domain features, and from the frequency domain, the MFCC and motifs. These features are described in the following sections.

Time features

For each segment, eight time-domain features were extracted:

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

MFCC features

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 [17]. 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 [15]. 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 Fig. 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

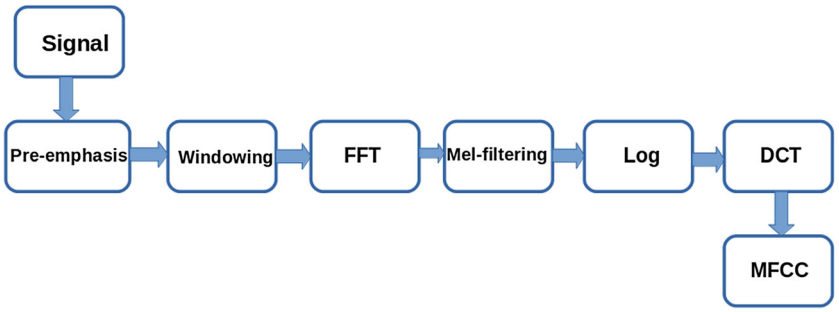

Fig. 2 MFCC feature extraction process

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 [20]:

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

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 [20]:

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

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 [20]:

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

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.

Motifs

Another set of features that we explore for this classification problem are based on motifs found in the cardiac audio time series [11, 12]. A motif in a time series is a frequent pattern, i.e. a particular segment of the series that is repeatedly observed. For the discovery of the motifs we use the algorithm MrMotif (Multiresolution Motif Discovery) [5]. It detects the top-K frequent motifs in a time series database D, given a window length m (number of points of the original series), a window overlap percentage (o), a word length w (number of symbols in the discretized series) and the value of K. Motifs are found in discretized versions of the time series that can be directly mapped to the original series. The discretization is done at different resolutions (number of discrete symbols used). The resolutions range in $g_{min}, g_{min} \times 2, \ldots, g_{max}$. In Fig. 3 we can see an example of a motif (20 - 20) with $m = 20$, $w = 8$ and $g = 4$. This algorithm is based on the iSAX methodology which discretizes continuous signals [18]. The minimum possible resolution g_{min} in iSAX is 2 and the maximum resolution g_{max} is typically 64.

To obtain the motif based features from the PCG segments, we have used window lengths of 20 and 40 ($m = 20$, $m = 40$), a word length of 8 ($w = 8$), a window overlap of 10% ($o = 10$) and a fixed resolution of 4 ($g = 4$). This resolution was the one that provided best results on similar data from our experience. The value of K was 20 and 40.

The process goes as follows. We apply MrMotif on the dataset of PCG segments. The top-K motifs $(M_1, M_2, \ldots, M_K)$ are identified and for each segment we obtain the frequency of occurrence f_i of each top motif M_i. The motif-based features are thus $M_1 = f_1, M_2 = f_2, \ldots, M_K = f_K$. On Table 1 we can see an example of six cases described with five motif based features ($K = 5$).

Data preparation

At this stage, we had a total of 13404 segments of three seconds of signal, resulting from the segmentation of the original 400 PCG signals. For each segment of three seconds of the signal, we have a total of eight features from the time domain, a collection of 3600 cepstral coefficients resulting from the 12 MFCC filterbanks and 300 time frames, and two groups of motif based features. One with 20 features (top-20) and window length of 20. The other has 40 features (top-40) and window length of 40. Both feature sets have resolution 4. Our set of features is in matrix form. The time features have 1 row and 8 columns, the MFCC has a total of 12 rows by 300 columns, and the two groups of motifs, one with 1 row by 20 columns, and another with 1 row and 40 columns. In some experiments, we join the

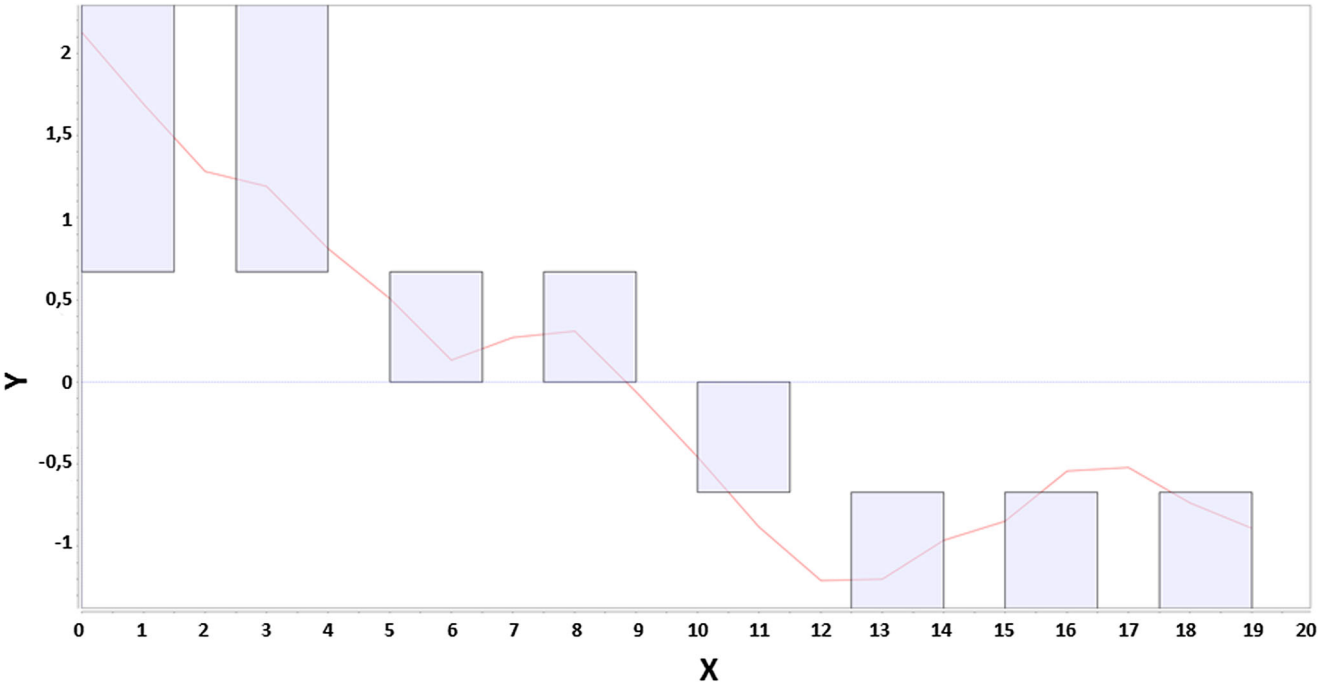

Fig. 3 Example of a discretized motif (20 - 20) with window length 20, resolution 4 and word length 8 overlayed on the original continuous time series segment

time features, with the two other types of features. When the join is made with the MFCC, the time features are placed in a new column, and to adjust the dimensions, zero padding is performed. From the combination of these two types of features, we get a matrix of 12 rows and 301 columns, in which the last column corresponds to the time features. In some tests, we use a different number of MFCC filterbanks. In our experiments we add as many columns as necessary to put the time features. For example, in the case where only 2 MFCC filterbanks are used, the time features are added in 4 new columns, resulting in a matrix with 2 rows and 304 columns, where the last 4 columns are filled by time features. In other experiments, like for example the case where we only use 5 MFCC filterbanks, the time features are added in 2 new columns, and we also need to add two zeros, to finish filling these two new columns. In these case we get a matrix with 5 rows and 302 columns, being the time features in the last two columns.

When we join the time features with the motifs, we simply concatenate the two matrices, resulting in the first case in a matrix of 1 row and 28 columns, and in the second case, in a matrix of 1 row and 48 columns. In both cases, the time features correspond to the eight final columns. The conversion of the features to a heatmap is done using the R package Plotly [14]. The features are normalized to values between 0 and 255, and the color domain of the heatmap, is defined, according to this range. We create different types of heatmap for each of the 13404 signal segments. For each signal segment we create one heatmap, resulting from the MFCC features and time features, and another two, resulting from the two groups of motifs, and the time features. Figure 4 illustrates one example, in which 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 heatmap resulting from the conversion of the MFCC together with the time features is also shown. The heatmap presented in the Fig. 4 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 represents the 12 filterbank frequencies that were used in the calculation of the MFCC. Several experiments were carried out, using the collected features and several classifiers. Given the computational

Table 1 Sample of a resulting dataset of resolution 4, mapping the Top-5 motifs

M_1	M_2	M_3	M_4	M_5	Class
1	0	5	2	0	N
2	2	7	4	3	N
4	0	2	2	1	N
0	1	3	2	1	A
6	4	5	3	3	A
2	0	2	2	0	A

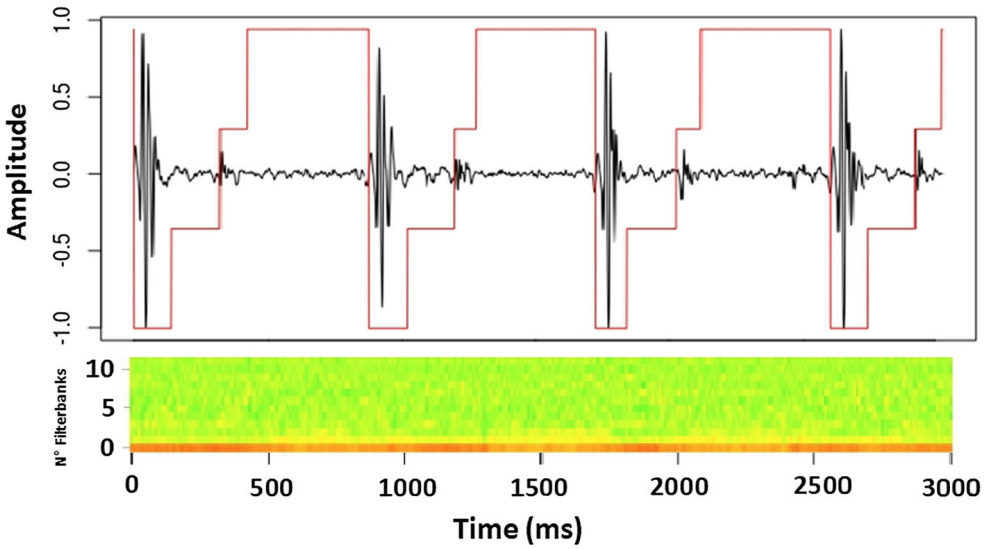

Fig. 4 Example of a PCG with the four states of the heart cycle (S1, systole, S2, diastole) identified (red line). MFCC heat map visualization of 3-second segment of heart sound data

power available, in the experiments in which we use MFCCs, instead of using the 12 filterbank frequencies, we use 2, 3, 4, 5 and 6 filterbank frequencies, in order to reduce the dimensionality of the features, and try to obtain the best possible performance. We have also done some experiments, using the feature sets separately, to evaluate the impact that their joint use has on the results. The tests in which the Motifs were used had a similar procedure. With regard to the heatmaps that are obtained, from the time features and the motifs, the procedure followed is the same, and we get two different heatmaps, one with 1 row and 28 columns, and the other with 1 row and 48 columns. In these heatmaps, the last 8 columns corresponds to the time features.

Classification of heart sound images

We have tested different learning algorithms to explore the dataset obtained in the previous step to discriminate between normal and abnormal heart sounds. In our setup, these algorithms will first assign a class to each three seconds segment. In a second step, we group the predictions for the various segments of the same patient to classify each original PCG signal. Here, the classification is done by considering the percentage of segments classified as normal. If the percentage is above a pre-defined threshold then the patient's heart condition is classified as normal. Otherwise as abnormal.

The value of the threshold was found empirically through exploration of the decision boundary on the posterior probability given by the classifiers. With the analysis of the results, we verified that a feature that greatly influences the obtained results is the heart rate (HR). Because of this, we decided to divide the PCG signals into eight different subgroups, taking into account the HR of the signals. In this way, it was possible to create groups of more similar signals, and to carry out their study individually. After that, we optimize the decision boundary, in each of the subgroups created, instead of making this optimization in the total set of signs, which have very different characteristics. The subgroups created were:

- The first group (G1) includes the signals with an HR less than 30;
- The second group (G2) includes the signals with an HR greater or equal to 30 and less than 40;
- The third (G3) includes the signals with an HR greater or equal to 40 and less than 50;
- The fourth (G4) includes the signals with an HR greater or equal to 50 and less than 60;
- The fifth (G5) includes the signals with an HR greater or equal to 60 and less than 70;
- The sixth (G6) includes the signals with an HR greater or equal to 70 and less than 80;
- The seventh (G7) includes the signals with an HR greater or equal to 80 and less than 90;
- The eighth group (G8) includes the signals with an HR greater or equal to 90.

We have observed that the decision boundary on the posterior probability, which presents the best results Overall, is around 25%, i.e., the original signals which has at least 25% of the segments classified as Normal (minority class of the dataset) would be classified as Normal. However, better results can be obtained by assigning different values of threshold according to the heart rate. In Table 2, we show the threshold values used for the different groups in different experiments.

Table 2 Threshold values used, for the several signals subgroups

Type of features	Model	G1	G2	G3	G4	G5	G6	G7	G8
5 MFCC + TF	SVM	0.3	0.3	0.25	0.17	0.3	0.21	0.4	0.21
6 MFCC + TF	SVM	0.3	0.3	0.37	0.25	0.24	0.21	0.24	0.2
5 MFCC + TF	RF	0.3	0.3	0.17	0.23	0.16	0.16	0.14	0.22
6 MFCC + TF	RF	0.3	0.3	0.13	0.18	0.13	0.22	0.16	0.18
Motifs (20-20) + TF	SVM	0.28	0.32	0.31	0.17	0.14	0.19	0.24	0.17
Motifs (40-40) + TF	SVM	0.36	0.32	0.21	0.17	0.16	0.18	0.23	0.17
Motifs (20-20) + TF	RF	0.3	0.3	0.24	0.14	0.13	0.16	0.32	0.18
Motifs (40-40) + TF	RF	0.3	0.3	0.19	0.15	0.15	0.23	0.29	0.18
TF	SVM	0.3	0.3	0.38	0.2	0.24	0.26	0.15	0.17
TF	RF	0.3	0.3	0.28	0.2	0.22	0.18	0.29	0.2

There is a wide variety of learning methods applied in the study of cardiac signals (ECG and PCG). Some of the classifiers used in this area are SVM [31], RF [2], and Neuronal Networks (NN) [6]. In our methodology, we study and evaluate the following algorithms: SVM, RF and Convolutional Neural Network (CNN). While the CNN allows the use of images directly as an input parameter, for the other classifiers, it is necessary to convert the heatmap to a vector line, so that it can be used as an input parameter. In the conversion of the heatmaps obtained with the motifs and with the time features, we simply linearize the pixel values of the images, being the last eight values, referring to time features. In the conversion of heatmaps obtained with MFCC and the time features, we first read and remove the column 301, where the temporal features are represented. Then, we linearize the pixel values of the images, by placing them sequentially in a single one-dimensional vector. In the end, we add the values from the temporal features, at the end of the line vector obtained.

For SVM we have used a radial kernel, with a cost value of 1.69 and a gamma value of $0.6618 * 10^{-3}$. For RF we use 400 trees and 200 variables randomly sampled as candidates at each split. These parameter values were found by a combination of initial manual exploration by the authors, followed by employing a random search over a limited range of parameters. With regard to CNN, the architecture and the parameters selected were based on the work of Rubin et al. [26], who built a PCG signal classifier using deep convolutional neural networks. The architecture is feasible using relatively limited computational resources. The classifiers were trained on a PC-type platform having a Intel® Core i7 CPU (3,40GHZ) and with 32 GB of RAM.

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% abnormal segments and 30% normal segments, we performed a 10-fold stratified cross validation (as shown in Fig. 5. The result is that the class distribution in each fold is similar to that in the original dataset. With this approach, it is guaranteed that all the k test sets are disjoint, and thus each case in the original training set is tested once and only once. In addition, it has been ensured that signal segments from the same patient are not placed in more than one fold. With this restriction, it is ensured that signal segments from the same patient are not split in train and test folds.

Equations 4, 5 and 6 show the Sensitivity (Se), Specificity (Sp) 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):

$$Se = \frac{TP}{TP + FN} \tag{4}$$

$$Sp = \frac{TN}{TN + FP} \tag{5}$$

$$Overall = \frac{Se + Sp}{2} \tag{6}$$

The Overall metrics was the evaluation metric used in the 2016 Physionet Challenge, and we will use it, in order to compare our results with those obtained in the challenge.

In order to try to overcome the problem of the unbalanced dataset, we performed two oversampling experiments, for distributions of classes 60% of abnormal and 40% of normal and another one to 50% of abnormal and 50% of normal. Since the results obtained with this methodology were poor, we decided not to include them in the article.

Experimental results

During classification, the number of filterbank frequencies of the MFCC features to be used was treated as a hyper-parameter. Several tests were performed, with a different

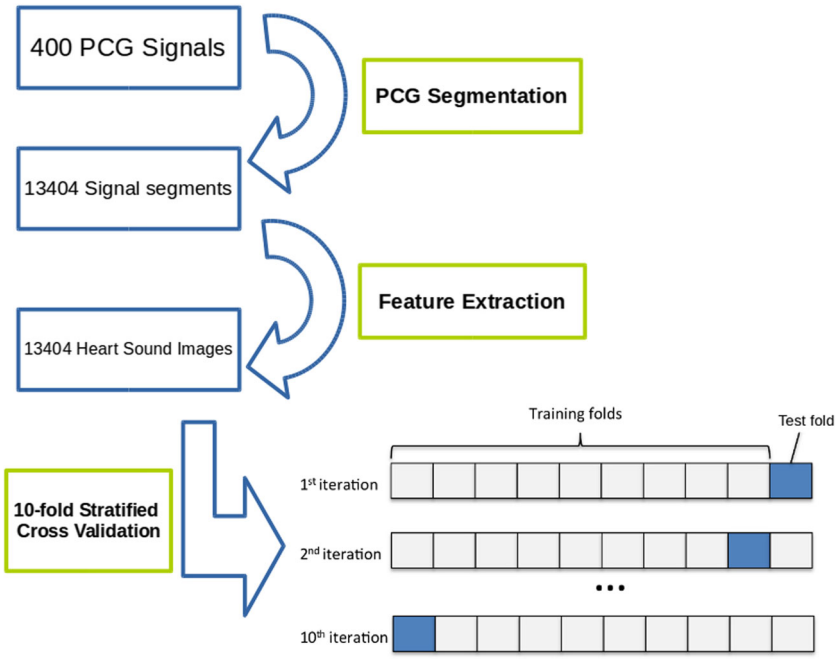

Fig. 5 Experimental setup diagram, with the 10-fold stratified cross validation

number of filterbank frequencies, in order to optimize the classification results. Table 3 shows the results obtained with the different approaches, using 5 and 6 MFCC features, together with TF, performed with the SVM, RF and CNN. We also present the results achieved with the TF, and the Motifs together with TF, obtained with SVM and RF. 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 and CNN, in this order, as can be seen in Table 3. Regarding the results obtained with the motifs, the Motifs (40-40) presents an Overall metrics slightly higher than that obtained with the Motifs (20-20). The TF, present intermediate results, with a higher Overall metrics than that obtained with the Motifs, but lower than that obtained with the MFCC.

Table 4 shows some results of the experiments where we studied the contribution of mapping temporal features to improve the results obtained with the MFCC, and Motifs features alone. In this table we present the results obtained with the classifiers SVM and RF, for two pairs of features, one pair composed by 4 MFCC only and by 4 MFCC together with TF, and the other is composed by Motifs (40-40) only, and by Motifs (40-40) together with TF.

In the Fig. 6, a graph is presented with the evolution of the Overall metrics, depending on the number of filterbank frequencies of the MFCC, which are used. This study was performed with the SVM and RF classifiers.

Table 3 Results obtained in the several tests performed

Type of features	Model	Sensitivity	Specificity	Overall
5 MFCC + TF	SVM	0.8737	0.7907	0.8322
6 MFCC + TF	SVM	0.8529	0.7792	0.8161
5 MFCC + TF	RF	0.9171	0.5314	0.7243
6 MFCC + TF	RF	0.9061	0.5224	0.7143
5 MFCC + TF	CNN	0.7529	0.3348	0.5439
6 MFCC + TF	CNN	0.8134	0.3386	0.5760
Motifs (20-20) + TF	SVM	0.8779	0.4459	0.6619
Motifs (40-40) + TF	SVM	0.7742	0.5656	0.6699
Motifs (20-20) + TF	RF	0.8242	0.6339	0.7291
Motifs (40-40) + TF	RF	0.86	0.5997	0.7299
TF	SVM	0.835	0.6767	0.7559
TF	RF	0.7671	0.7194	0.7433

Table 4 Results obtained, with or without inclusion of time features

Type of features	Model	Sensitivity	Specificity	Overall
4 MFCC + TF	SVM	0.8836	0.7692	0.8264
4 MFCC	SVM	0.88	0.7650	0.8225
4 MFCC + TF	RF	0.9385	0.5570	0.7478
4 MFCC	RF	0.91	0.5488	0.7294
Motifs (40-40) + TF	SVM	0.7742	0.5656	0.6699
Motifs (40-40)	SVM	0.835	0.4288	0.6319
Motifs (40-40) + TF	RF	0.86	0.5997	0.7299
Motifs (40-40)	RF	0.7921	0.3433	0.5677

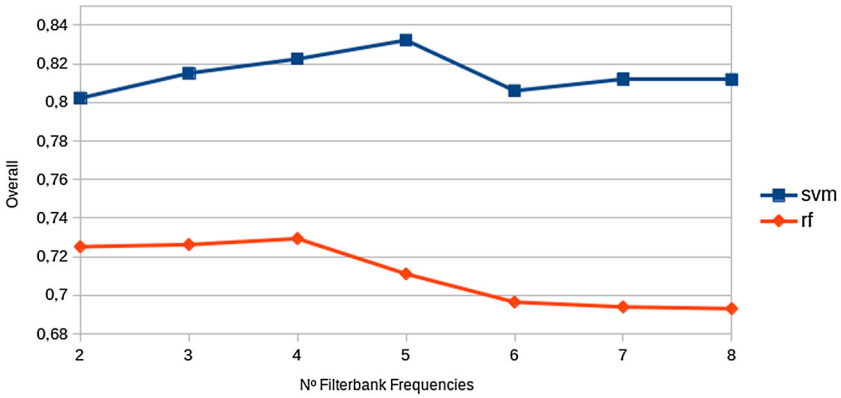

Fig. 6 Evolution of the Overall metrics obtained for the several MFCC filterbank used, with the SVM and RF

The Overall scores for the top entries of the PhysioNet Computing in Cardiology Challenge were very close(see scores in [7]). The difference between the top place finisher Overall (0.8602) and the 10th place (0.8263) was just approximately 0.04.

Discussion

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 3 and 4. Analyzing the Table 3, the best results were obtained with the SVM radial basis function kernel. The best result was obtained with a set of five MFCC features, together with the temporal features, with which a sensitivity of 0.8737, a specificity of 0.7907 and an Overall of 0.8322 were obtained. With the other 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. [26]. 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 [28]. In our work, we only used a small portion of the dataset due to the computational power that was available.

The misclassification of data may be due to the heterogeneity of the two signal classes. Pathologic heart sounds consists of several categories, ranging from valvular problems to congenital disorders, each possessing a very different feature. Besides that, how these heart sounds are labeled is also an issue. Once the signals are labeled 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 misclassified signals, which will also contribute to the heterogeneity of the signal classes.

Another important conclusion, that can be deduced from the analysis of the results, is the contribution of mapping TF to improve the results obtained with the other features alone. As can be seen in Table 4, in all the experiments where the TF were used together with the other features we get a higher classification accuracy. Analyzing the results, it is possible to identify one case (Motifs (40-40) + TF) which presents better results than using the Motifs (40-40) alone, where the Overall metrics gain is approximately 0.16.

As was already mentioned above, the Overall scores for the top entries of the PhysioNet Challenge were very close, being the difference between the ten best submissions, just 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 achieved an Overall metrics of 0.8322, between the top entries of the Challenge. The use of time features, along with the other features, had a fundamental role in the obtained results, as it was demonstrated in the analysis of Table 4. Their presence led to an improvement of the results in all the tests performed.

Furthermore, our best performance was achieved using a single SVM with a Radial basis function kernel, whereas other top place finishers of the challenge achieved strong classification accuracies with an ensemble of classifiers [7]. In practical terms, a system that relies on only a single classifier, as opposed to a large ensemble, has the advantage of limiting the number of computational resources required for classification and be easier to understand.

Lastly, analyzing the graph of the Fig. 6, regarding the results obtained with the RF (in orange), the biggest Overall metrics was achieved using five MFCCs filterbank frequencies, and for cases where more than five MFCCs filterbank frequencies were used, the Overall metrics decreased. About SVM, the biggest Overall metrics was achieved with the use of six MFCCs filterbank frequencies,

and just like with RF, for the cases in which a larger number of MFCCs filterbank frequencies was used, the Overall metrics is smaller.

Conclusions

In this work we explore the dataset that was made available at Physionet databases, whose the main goal is to discriminate between normal and abnormal hearts using PCG and ECG signals. Our 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 set of time domain features and two types of frequency domain features, the MFCC and the Motifs. We calculate the features of each segment of PCG signal, being these features used as the input to our classifiers. We perform a binary classification that will assign a classification to all signal segments. Thereafter, it was necessary to group the predictions of the various segments to classify the original PCG signals. We have used a SVM radial basis algorithm in the classification of heart sounds as normal or abnormal, obtaining an accuracy of approximately 83.22%. The performance of the model was evaluated and compared to other classifiers. With the proposed approach we get an Overall metrics among the top entries of the Challenge. 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. One of the classifiers used was the CNN, which had worse results than the other classifiers. One possible cause for this result 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 (spectrograms). Another important procedure, which we will use in a future work, is the implementation of a dimension reduction process, in order to extract the most relevant information from the calculated features. Relatively to the threshold, based on the percentage of segments classified as normal, which is used to classify the original PCG signals, we intend to use the ROC curves, to finding the optimal threshold. Furthermore, we intend to use CNN in larger datasets, in order to take full advantage of its ability.

Acknowledgements 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)*.

Funding Information This study was funded 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)*.

Compliance with Ethical Standards

Conflict of interests None.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Physiobank atm. https://physionet.org/cgi-bin/atm/ATM.
2. Balili, C. C., Sobrepena, M. C. C., and 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, 2015.
3. Barschdorff, D., Bothe, A., and Rengshausen, U., Heart sound analysis using neural and statistical classifiers: a comparison. In: *[1989] Proceedings. Computers in Cardiology*, pp. 415–418, 1989.
4. Boussaa, M., Atouf, I., Atibi, M., and Bennis, A., Ecg signals classification using mfcc coefficients and ann classifier. In: *2016 International Conference on Electrical and Information Technologies*, pp. 480–484, 2016.
5. Castro, N., and Azevedo, P., Multiresolution Motif Discovery in Time Series, pp. 665–676. https://doi.org/10.1137/1.9781611972801.73.
6. Chen, T. E., Yang, S. I., Ho, L. T. et al., S1 and s2 heart sound recognition using deep neural networks. *IEEE Trans. Biomed. Eng.* 64(2):372–380, 2017.
7. Clifford, G. D., Liu, C., Moody, B., Springer, D., Silva, I., Li, Q., and Mark, R. G., Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016. In: *2016 Computing in Cardiology Conference (cinc)*, pp. 609–612, 2016.
8. Colonna, J., Peet, T., Ferreira, C. A., Jorge, A. M., Gomes, E. F., and 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, C3S2E '16*, pp. 73–78. ACM, 2016.
9. Ergen, B., Tatar, Y., and Gulcur, H. O., Time–frequency analysis of phonocardiogram signals using wavelet transform: a comparative study. *Comput. Methods Biomech. Biomed. Engin.* 15(4):371–381, 2012.
10. Gomes, E., Bentley, P., Coimbra, M., Pereira, E., and Deng, Y., Classifying heart sounds: Approaches to the pascal challenge pp 337–340, 2013.
11. Gomes, E. F., Jorge, A. M., and Azevedo, P. J., Classifying heart sounds using multiresolution time series motifs: an exploratory study. In: *Proceedings of the International C* Conference on*

Computer Science and Software Engineering, pp. 23–30. ACM, 2013.
12. Gomes, E. F., Jorge, A. M., and Azevedo, P. J., Classifying heart sounds using sax motifs, random forests and text mining techniques. In: *Proceedings of the 18th International Database Engineering & Applications Symposium*, pp. 334–337. ACM, 2014.
13. Huiying, L., Sakari, L., and 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, 1997.
14. Inc., P. T., Collaborative data science. https://plot.ly, 2015.
15. Kishore, K. V. K., and Satish, P. K., Emotion recognition in speech using mfcc and wavelet features. In: *2013 3rd IEEE International Advance Computing Conference (IACC)*, pp. 842–847, 2013.
16. Kumar, D., Carvalho, P., Antunes, M., Paiva, R. P., and Henriques, J., Heart murmur classification with feature selection. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 4566–4569, 2010.
17. Lalitha, S., Geyasruti, D., Narayanan, R., and Shravani, M., Emotion detection using mfcc and cepstrum features. *Prog. Comput. Sci.* 70:29–35, 2015.
18. Lin, J., Keogh, E., Lonardi, S., and Patel, P., Finding motifs in time series. In: *Proceedings of the 2nd Workshop on Temporal Data Mining*, pp. 53–68, 2002.
19. Liu, C., Springer, D., and Li, Q., Et. al.: An open access database for the evaluation of heart sound algorithms. *Physiol. Meas.* 37(12):2181, 2016.
20. Lubaib, P., and Muneer, K. A., The heart defect analysis based on pcg signals using pattern recognition techniques. *Procedia Technol.* 24:1024–1031, 2016.
21. Malarvili, M. B., Kamarulafizam, I., Hussain, S., and Helmi, D., Heart sound segmentation algorithm based on instantaneous energy of electrocardiogram. In: *Computers in Cardiology*, pp. 327–330, 2003.
22. Mozaffarian, D., Benjamin, E. J., and Go Alan, S., E.a.: Heart disease and stroke statistics–2016 update. Circulation, 2015.
23. Nogueira, D. M., Ferreira, C. A., and Jorge, A. M., Classifying heart sounds using images of MFCC and temporal features. In: *Progress in Artificial Intelligence - 18th EPIA Conference on Artificial Intelligence, EPIA 2017*, pp. 186–203, 2017.
24. Obaidat, M. S., Phonocardiogram signal analysis: techniques and performance comparison. *J. Med. Eng. Technol.* 17(6):221–7, 1993.
25. Rangayyan, R., and Lehner, R., Phonocardiogram signal analysis: a review. *Crit Rev Biomed Eng.* 15(3):211–236, 1987.
26. Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., and 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.
27. Segal, B. L., Phonocardiology: Integrated study of heart sounds and murmurs. *JAMA* 224(11):1536–1536, 1973.
28. Shi, W., Gong, Y., and Wang, J., Improving cnn performance with min-max objective. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI'16*, pp. 2004–2010, 2016.
29. Springer, D. B., Tarassenko, L., and Clifford, G. D., Logistic regression-hsmm-based heart sound segmentation. *IEEE Trans. Biomed. Eng.* 63(4):822–832, 2016.
30. White, P. R., Collis, W. B., and 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.
31. Wu, J., Zhou, S., Wu, Z. M., and Wu, X., Research on the method of characteristic extraction and classification of phonocardiogram. In: *2012 International Conference on Systems and Informatics*, pp. 1732–1735, 2012.

Affiliations

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

Carlos Abreu Ferreira
cgf@isep.ipp.pt

Elsa Ferreira Gomes
efg@isep.ipp.pt

Alípio M. Jorge
amjorge@fc.up.pt

1 INESC TEC, Campus da FEUP, Rua Dr. Roberto Frias, 4200 – 465, Porto, Portugal

2 Instituto Superior de Engenharia do Porto, Rua Dr. Bernardino de Almeida, 431, 4200-072, Porto, Portugal

3 Faculdade de Ciências da Universidade do Porto, Rua Campo Alegre 1021/1055, 4169-007, Porto, Portugal