

dx.doi.org/10.17488/RMIB.44.4.1

E-LOCATION ID: 1348

Comparison of Spectral and Sparse Feature Extraction Methods for Heart Sounds Classification

Comparación de Métodos de Extracción de Características Espectrales y Dispersas para Clasificación de Sonidos Cardíacos

Roilhi Frajo Ibarra-Hernández¹  , Miguel Ángel Alonso-Arévalo² , Eloísa del Carmen García-Canseco³ 

¹Universidad de Ensenada, Ensenada - México

²Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada, Baja California - México

³Universidad Autónoma de Baja California, Ensenada, Baja California - México

ABSTRACT

Cardiovascular diseases (CVDs) remain the leading cause of morbidity worldwide. The heart sound signal or phonocardiogram (PCG) is the most simple, low-cost, and effective tool to assist physicians in diagnosing CVDs. Advances in signal processing and machine learning have motivated the design of computer-aided systems for heart illness detection based only on the PCG. The objective of this work is to compare the effects of using spectral and sparse features for a classification scheme to detect the presence/absence of a pathological state in a heart sound signal, more specifically, sparse representations using Matching Pursuit with multiscale Gabor time-frequency dictionaries, linear prediction coding, and Mel-frequency cepstral coefficients. This work compares the performance of PCGs classification applying features as a result of averaging the samples or the features for each PCG sound event when feeding a random forest (RF) classifier. For data balancing, random under-sampling and synthetic minority oversampling (SMOTE) methods were applied. Furthermore, we compare the Correlation Feature Selection (CFS) and Information Gain (IG) for the dimensionality reduction. The findings show a SE=93.17 %, SP=84.32 % and ACC=85.9 % when joining MP+LPC+MFCC features set with an AUC=0.969 showing that these features are promising to be used in heart sounds anomaly detection schemes.

KEYWORDS: classification, heart sounds, matching pursuit, spectral features, time-frequency representation

RESUMEN

Las enfermedades cardiovasculares (ECVs) han persistido como la principal causa de mortalidad en el mundo. La señal de audio cardíaco o fonocardiograma (FCG) es la herramienta más simple, efectiva y de bajo costo para auxiliar a especialistas diagnosticando ECVs. Los avances en el procesamiento de señales y aprendizaje máquina han motivado el diseño de auscultación y detección computarizada. El objetivo de este trabajo es comparar el uso de características espectrales y *dispersas* para un sistema de clasificación que detecte la presencia/ausencia de una patología en un audio cardíaco mediante representaciones dispersas usando *Matching Pursuit* con diccionarios de Gabor tiempo-frecuencia, predicción lineal y coeficientes cepstrales Mel. Se crearon 5 conjuntos de características como resultado de combinar las características para cada FCG y se examinó su desempeño usando un clasificador de bosque aleatorio (RF). Se aplicaron métodos de balanceo de muestras basados en sobremuestreo (SMOTE) y submuestreo aleatorio. Se compararon métodos de selección de características por correlación (CFS) y ganancia de información (IG) para *reducir la dimensionalidad* del conjunto. Los resultados muestran métricas de SE=93.17 %, SP=84.32 % y ACC=85.9 % al juntar los parámetros MP+LPC+MFCC además de una AUC=0.969. El trabajo muestra el potencial de las características espectrales y escasas para la detección de patologías en señales de audio cardíaco.

PALABRAS CLAVE: características espectrales, clasificación, matching pursuit, representación tiempo-frecuencia, sonidos cardíacos

Corresponding author

TO: Roilhi Frajo Ibarra-Hernández

INSTITUTION: Universidad de Ensenada

ADDRESS: Mar 198, Tercer Ayuntamiento, CP 22830.

Ensenada BC.

CORREO ELECTRÓNICO:

roilhi.ibarra@universidaddeensenada.edu.mx

Received:

27 March 2023

Accepted:

21 June 2023

INTRODUCTION

Heart diseases remain the leading cause of death worldwide, according to a report from the World Health Organization [1]. An effective method that leads to the primary diagnosis of heart illness is automatic abnormal heart sound detection, which aims to identify the presence of a cardiac malfunction. This area has raised interest among researchers with the introduction of electronic stethoscopes and the advances in signal processing. In general, the methods for diagnosing pathological states of heart sounds consist of two stages: firstly, the feature extraction process to obtain the most representative parameters of cardiac sound, and secondly, the classification, which predicts the patient's condition from the patterns found in the extracted features. In healthy individuals (adults), the heart sound signal, also known as phonocardiogram (PCG), comprises two main components called *fundamental heart sounds* (FHS), which are denoted as s_1 and s_2 . Usually, a typical time duration and low-frequency spectral content characterize each FHS. For instance, the s_1 components dominate the region from 10 Hz to 140 Hz, while the s_2 components usually concentrate their energy around the 10 Hz to 200 Hz band [2].

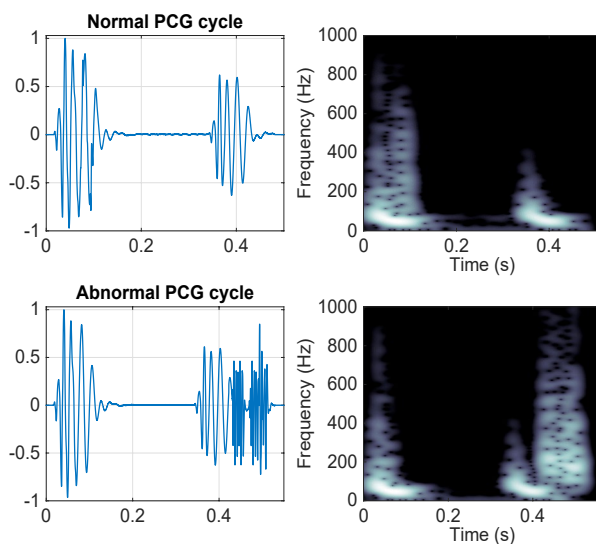


FIGURE 1. Top: time waveform and spectrogram of a normal PCG signal. Bottom: time waveform and spectrogram of an abnormal (pathological) PCG signal.

In pathological conditions, sounds named *murmurs* appear. Murmurs are sounds stemming from a *turbulent blood flow* due to a valve malfunction or an obstruction, denoting a pathological or abnormal state. The energy distribution of murmurs in frequency varies widely and, depending on their nature, can go above 800 Hz. Unfortunately, the frequency content of murmurs can overlap with the distribution of s_1 and s_2 , and thus, the correct identification of the sound is a difficult task that requires sophisticated methods to determine the type of sound. Figure 1 illustrates the waveform and the time-frequency content representative spectrogram of a PCG cardiac cycle in both normal and pathological states.

Review of PCGs classification schemes

A thorough review of existing methods to classify heart sounds is out of the scope of this work. However, the 2016 PhysioNet/Computing in Cardiology Challenge (CinC) [3] and the release of one of the more extensive public databases of PCG recordings are a milestone in the field. To provide a literature review for PCGs classification algorithms, we can organize them into two main categories according to a) the feature extraction methods and b) the classification schemes used by each research paper.

For the first category, the feature extraction methods aim to represent the cardiac sound signals in different domains (time, frequency, and joint time-frequency, mainly), revealing the main physiological and pathological PCG attributes to allow an effective feature extraction. Since the PCG signal is quasi-stationary, the features provided would be able to capture concurrent variations and the structural components in time, frequency, and joint time-frequency domains. For these reasons, selecting an adequate feature extraction method is crucial for classifying heart sound signals. For instance, in the time-frequency domain representation of the PCG, researchers have chosen the short-time Fourier Transform (STFT) [4][5][6], Wigner-Ville dis-

tributions^{[7][8]}, the empirical^[9], discrete^[10], and continuous Wavelet transform^{[11][12][13][14]}. Among other frequency domain features utilized for PCGs classification, the Mel-Frequency Cepstral Coefficients (MFCCs) have been widely used as classification input features^{[15][16][17][18][19][20][21]} since these parameters are the most popular to characterize the envelope information for audio signals successfully. The Linear Predictive Coefficients (LPCs)^[22] have also been used to capture PCG signal spectrum patterns. The second category comprises the selection of a classification scheme which is essential since it is the final step of a PCGs murmur detection algorithm. The classifier takes the extracted features and interprets them by extracting and recognizing the functional patterns to efficiently represent the murmurs associated with diseases of a PCG signal. In the state-of-the-art, the reported classification schemes used are Support Vector Machines^{[8][16][23][24][25][26][27][28][29]}, k-Nearest Neighbors^{[16][30][31][32][33][34]}, and Random Forests techniques^{[14][35][36][37]}, in terms of conventional Machine Learning techniques. On the other hand, reported Deep learning-based methods for PCGs classification are comprised of ensembles of neural networks^{[15][17][38][39][40]}, convolutional neural networks (CNN)^{[6][13][21][41][42][43][44][45][46]}, long short-term memory networks (LSTM)^{[47][48][49]}, and recurrent neural networks (RNN)^[50]. Although deep learning has emerged as a powerful approach that has shown promising advances in PCGs classification, there are still limitations due to the lack of data, carrying out training inefficiency, and insufficiently robust models^{[51][52][53]}. Deep learning algorithms require significantly more computational resources and may not be feasible for machines with embedded or limited hardware capabilities. Deep learning algorithms might also present a limitation called the exploding and vanishing gradient descent problem, which causes the classification error rate to increase after attaining a minimum value. This deficiency is also known to cause model overfitting. Another limitation of deep learning is the lack of interpretability of the features to the point where it is impossible to discern what they are and

have no physical meaning. While that may be a reasonable price for the theoretical performance gain in some applications, we consider it vital to understand the physiological phenomena in PCG analysis.

This work aims to leverage sparse representations to classify heart sounds. More specifically, Matching Pursuit (MP) coefficients combined with LPCs and MFCCs as features feed our proposed high-performance scheme that detects PCG abnormalities. We selected the Random Forest classifier as a classification algorithm due to its simplicity, low computational requirements, and excellent performance. The RF classifier is still used among researchers to detect pathological states from heart sounds. However, it is noteworthy that deep learning algorithms have significantly improved in recent years and are now the default go-to choice for many problems, especially in computer vision and natural language processing fields.

On the other hand, we used the Synthetic Minority Oversampling technique (SMOTE) to address the problem of unbalancing during the classification by creating synthetic samples for the minority class (abnormal or pathological PCG sound signals). The classification scheme's performance has been analyzed when the inputs are noisy PCG recordings. Finally, the work compared the performance of two feature selection techniques.

This paper presents our study's methodology, results, and conclusions, which aim to investigate the effectiveness of using sparse and spectral features for PCG signal classification. The methodology section describes the methods we used to conduct our experiments, including the selection of datasets, the choice of algorithms, and the evaluation metrics. The results section presents the findings of our investigations, including the performance of different algorithms and their comparison. Finally, in the conclusion section, we summarize our study's key insights and implications,

and the limitations and future research directions.

MATERIALS AND METHODS

The main goal of this research is to evaluate the classification performance of different sets of features as input parameters of a classifier to accurately detect pathological states in PCG signals. The Physionet/CinC 2016 is the largest database of PCG signals publicly available to the scientific community [54] in order to evaluate algorithms to segment and classify PCGs. It comprises the merge of six different research groups of recordings from subjects under normal and various pathological cardiac conditions. Specifically, the database includes 3,153 sounds recorded with a 2,000 Hz sampling frequency. Moreover, 2,488 samples come from cardiac sounds of subjects under normal conditions, while 665 represent an abnormal category.

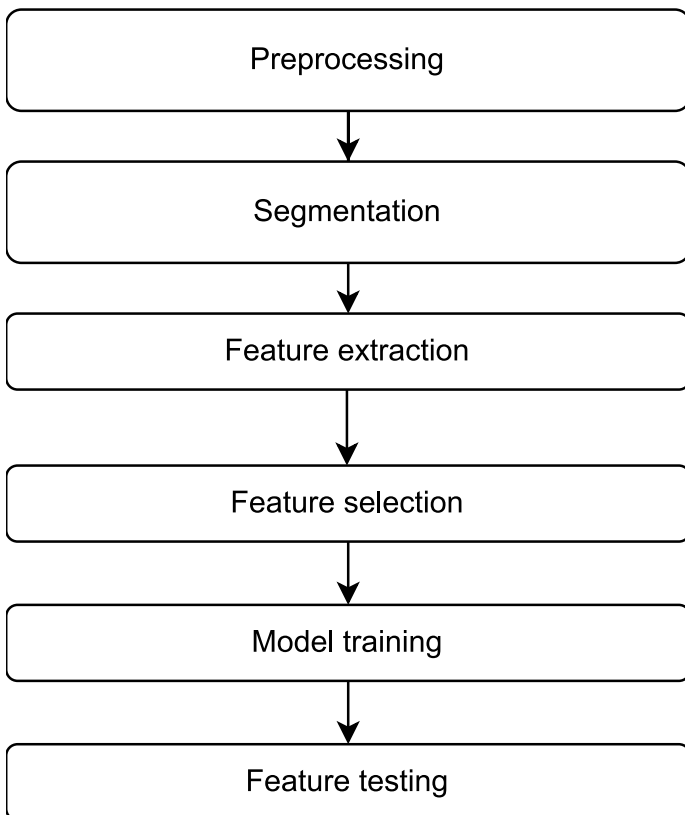


FIGURE 2. Block diagram to describe the sequence of methods to conduct the experiment of this paper.

In this paper, we conducted the methodology shown in the block diagram of Figure 2. For the preprocessing stage, the PCG signals were band-pass filtered between 25-600 Hz using a sixth-order Butterworth filter; then, we applied a normalization procedure in amplitude, which consists of dividing the recording samples by the maximum value. The second stage comprises the extraction of FHS, since for these events in the following step, different features will be extracted. The feature selection stage consists of reducing the number of features in order to know which of them are the most relevant and have the most information. Finally, in the training stage, we feed a classifier algorithm using the different sets of features to evaluate the classification performance of each one of them.

Matching Pursuit

The Matching Pursuit algorithm (MP), proposed by Mallat [55], is a greedy and iterative method that computes a *sparse* representation of a signal s as a linear combination of M_a *elementary waveforms* called *atoms* with *minimal error*. Each atom \hat{g}_m belongs to a redundant set of all possible predefined signals called dictionary \mathbf{D} . MP selects the best-correlated atom \hat{g}_m iteratively to provide a *sparse* decomposition in the following way:

$$s = \sum_{m=1}^{M_a} \alpha_m \cdot \hat{g}_m + r, \quad (1)$$

the atom that MP chooses at each iteration is the one that best matches the local signal structure of s by calculating the maximum inner product between the signal and the dictionary:

$$\hat{g}_m = \operatorname{argmax}_{\hat{g}_m \in \mathbf{D}} |\langle r, \hat{g}_m \rangle|, \quad (2)$$

the weighting factor α_m is a scalar that comes from the value of the inner product at each iteration:

$$\alpha_m = \langle \mathbf{r}, \hat{\mathbf{g}}_m \rangle, \quad (3)$$

\mathbf{r} is a signal called the *residual* term. It comes from the difference between the signal and the weighted-selected atom:

$$\mathbf{r} = \mathbf{r} - \alpha_m \cdot \hat{\mathbf{g}}_m, \quad (4)$$

notice that at the beginning of the algorithm $\mathbf{r}=\mathbf{s}$. MP is called a *greedy* method to reconstruct sparse signals because it stops until a desired number of iterations (or atoms) M_a or the ratio between the original signal energy and the residual has been reached. The dictionary selection is a crucial step for the MP decomposition into atoms. The dictionaries of Gabor functions have been widely used for the reconstruction of PCG signals due to the accurate signal representation in the time-frequency domain [56][57][58][59][60][61]; nonetheless, Gabor atoms are *well-concentrated* waveforms in both time and frequency. In this work, we use as a dictionary a set of predefined multiscale functions, which is a collection $\mathbf{D}=\mathbf{U}_{j=1}^J \mathbf{D}_j$ of blocks \mathbf{D}_j of time-frequency atoms at different scales. A Gabor atom in a multiscale dictionary is a waveform defined by the modulation, dilation, translation, and sampling of a continuous window $w_j(t)$ as:

$$\mathbf{g}_{j,n,k}(m) = w_j(mT_s - nT_j) \exp\left(\frac{2i\pi kmT_s}{K_j}\right) \text{ for } 1 \leq m \leq M, \quad (5)$$

where the time location or window shift is defined as nT_j , the window length or scale L_j and is modulated at a frequency k/K_j , where K_j is a predefined number of possible frequencies (according to the FFT size), T_s is the sampling period, and M the number of samples. Figure 3 shows the time waveform of a Gabor atom, which can be seen as a cosine-modulated Gaussian window. At the right panel, a couple of waveforms illustrate the effect of changing the modulation frequency. After the frequency of the signal has been warped into the Mel scale, each C_n MFCC coefficient is calculated as follows:

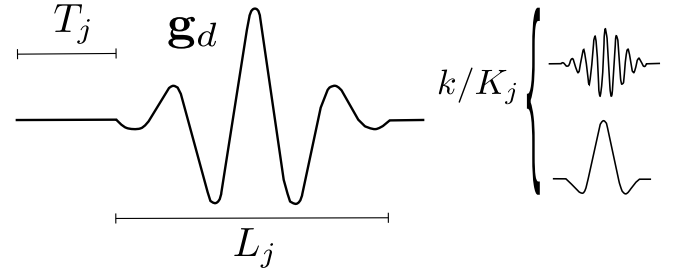


FIGURE 3. Time waveform of a Gabor atom and its defined parameters. In the right panel, the waveforms illustrate the effect of changing the frequency parameter k/K_j .

Linear Predictive Coding

As seen in equation (1), MP decomposes a signal in two main parts, a linear combination of Gabor atoms and a residual. The residual term \mathbf{r} is expected to be lowly correlated with the selected dictionary atoms. Thus, it must be expressed differently to be integrated as a feature representing the PCG signal. For this reason, instead of reconstructing the temporal waveform, we propose to represent \mathbf{r} using the Linear Predictive Coding technique [62], which approximates the signal's spectrum rather than the time domain waveform. The LPC representation is an all-pole filter where the residual \mathbf{r} can be predicted as a linear combination of the previous samples:

$$\mathbf{r}_n = - \sum_{i=1}^p h_i \mathbf{r}_{n-i} + \mathbf{e}_n, \quad (6)$$

where $n=0, \dots, N-1$, \mathbf{e}_n is the final residual, and p is the filter order. Filter coefficients h_i are added to the features set. Published works in the literature review have used the LPC coefficients as features for the automated detection of heart murmurs in PCG signals [22].

Mel-Frequency Cepstral Coefficients (MFCCs)

The Mel-Frequency Cepstral coefficients (MFCCs) are

the predominant features used for speech recognition [63], because they provide a compact and *smooth representation* of the magnitude spectrum. MFCCs are based on the human hearing physiological structure since the human perception of the frequency content of sounds does not follow a linear scale. Thus, having a signal with a fundamental frequency f and an estimated pitch should be measured on a ranking called the Mel Scale. The MFCC coefficients are calculated by taking the discrete cosine transform of a logarithmic spectrum after it was warped to the Mel scale as follows:

$$\text{Mel}(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right), \quad (7)$$

after the frequency of the signal has been warped into the Mel scale, each C_n MFCC coefficient is calculated as follows:

$$C_n = \sum_{m=1}^{M_c} D_m \cos \left(n \left(\frac{m-0.5}{M_c} \right) \pi \right), \quad (8)$$

where D_m is the output of the k -th triangular filter bank channel and M_c is the number of filter bank channels. In our implementation, we use $M_c=14$ to cover the range from 20 Hz to 900 Hz. Figure 4 shows the representation in the frequency domain (Hz and Mel scale) of the triangular filter bank used for the MFCC coefficients extraction.

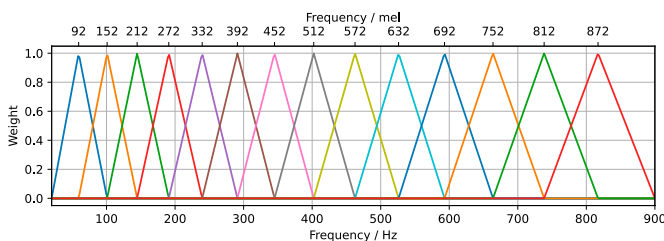


FIGURE 4. Triangular Mel Filter bank to extract MFCC coefficients used in this work. The frequency is shown at the top in the Mel Frequency and at the bottom in Hertz, respectively.

Reported research frameworks have used MFCCs as features for PCGs classification [15][16][17][18][19][20][21], since they provide meaningful representations in the spectral envelope rather than time features.

Random Forest Classifier

The random forest classifier (RF) comes from combining two or more *decision tree* classifiers. Each classifier uses a random vector sampled independently from the input vector and casts a *unit vote* for the most popular class. The features used are randomly selected to grow a tree. RF uses a *bagging* method to randomly replace the N examples of the original training set [64].

Let Θ be a random vector that chooses a random subset x from the training set X . Let N_T be the number of decision trees; each one has an additional parameter Θ_t and the ensemble of trees consists of the set $\{f_1(x, \Theta_1), f_2(x, \Theta_2), \dots, f_{N_T}(x, \Theta_{N_T})\}$. The RF algorithm attempts to reduce the variance of the model by averaging many trees estimates as follows:

$$f_{RF}(x) = \frac{1}{N_T} \sum_{t=1}^{N_T} \alpha_t f_t(x, \Theta_t), \quad (9)$$

Where α_t represents an associated weight. Because of its simplicity and promising results, the RF classifier has been widely used for PCG signals classification [14][65][66][67][68][69]. It is still a valuable method to detect heart murmurs accurately. For the experiments conducted in this research, we choose as hyperparameters a number of estimators $N_e=100$, and the Gini criterion to measure the quality of the splits.

The Synthetic minority oversampling technique (SMOTE)

Most PCG datasets around the reported research works contain more recordings from healthy people (commonly labeled as *normal sounds*) than people with

a heart pathology (commonly labeled as *abnormal sounds*). The training stage will be affected due to this unbalancing between class samples, causing overfitting and highly biased results. The Synthetic minority oversampling technique is an algorithm that addresses the unbalancing problem by creating synthetic samples of the minority class. These synthetic samples are generated over the feature space rather than the data space. Each minority sample is created by taking the difference between the feature vector (input sample) and its nearest neighbor. The difference is then multiplied by a random number between 0 and 1 and added to the feature vector under consideration. The SMOTE approach effectively forces the decision region of the minority class to be more general, causing a better performance in a classification that uses decision trees. It has been shown that SMOTE technique performs better in accuracy than under-sampling methods [70].

Feature extraction

We have previously evaluated several time-frequency dictionaries to decompose the PCG, showing that Gabor wavelets accurately represent this signal [71]. For the experiments conducted in this research, the selected number of atoms was $M_a=15$ in order to reach almost 99 % of the energy to reconstruct a PCG cycle. For the LPC analysis, the number of coefficients was $p=15$. For the MFCCs, we followed the suggestion proposed by some methods in the context of the Physionet Challenge [15][17][38], setting the number of coefficients as $M_c=14$. Figure 5 provides a block diagram which describes the feature sets used in our experiments. We generated five feature sets labeled as follows by combining the MP, MFCC, and LPC approaches. Set A contains 90 features (i.e., columns of A data frame are 90) by merging the MP+LPC parameters, while set B contains the same extracted features as set A; however, they are extracted after performing cycle averaging. Set C consists of 146 features after combining MP+MFCC+LPC, set D contains 131 features after joining MP+MFCC. Set E contains 56 features by considering only MFCC. The proce-

cedure of feature extraction was conducted in MATLAB ©.

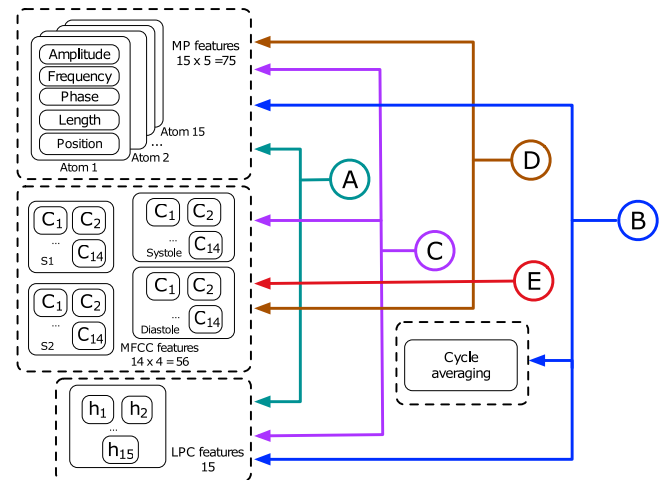


FIGURE 5. Block diagram of feature sets used in this work. In the first stage, we extracted $M_a=15$ atoms, with five parameters per atom yielding a total of 75 MP features. Using $M_c=14$ for each of the four states in the PCG cycle yields a total of 56 MFCC features. We defined $p=15$ as the number of LPC features.

Processing of the low-quality recordings

The database described at the beginning of this section comes from the Physionet/CinC 2016 challenge. Data includes the PCG recordings, the FHS time segmentation boundaries, a label indicating the pathological condition (normal/abnormal), and a quality reference of the PCG signal. According to the noise level in the sound samples, the database is divided into High-Quality Recordings (HQR) and Low-Quality Recordings (LQR). Due to their highly noisy condition, there are 279 signals labeled as LQR, and their FHS time segmentation boundaries are not provided. Since most of the features are calculated per cardiac cycle, in the case of LQR the computation of the features was conducted by splitting the PCG in segments of $T_\mu + \sigma_T=1.15$ s, where T_μ is the average cardiac cycle duration for the recordings labeled as abnormal and σ_T is the respective standard deviation. However, to compute the MFCCs each PCG segment was sliced into 4 windows according

to the average duration of the FHS [2].

RF classifier settings

We changed the number of estimators for the RF method to 100, as recommended in the presence of unbalanced datasets [72][73]. Specific details and parameter settings used during the evaluation are provided in previous work [74], where the RF classifier outperformed the others. This evaluation and all the classification tests were conducted using the scikit-learn toolbox under Python [75]. The experiments presented in this paper were conducted on a workstation with an Intel i7-9750H processor (2.60 GHz) and NVIDIA GPU GTX 1660Ti.

The confusion matrix is a well-known method used to evaluate the performance of ML classification schemes. In our case, for a binary class problem (having *normal* and *abnormal* labels), the confusion matrix has four values:

-True positives (TP): number of correctly identified PCGs with a pathological condition.

-True negatives (TN): number of correctly classified PCGs that do not have a pathology.

-False positives (FP): number of PCG signals labeled as abnormal but classified as *normal*.

-False negatives (FN): number of PCG signals labeled as normal but classified as *abnormal*.

For the experiments conducted in this research, we considered these quantities in order to calculate the following classification metrics:

- Accuracy (ACC)

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \times 100 \in [0 \ 100]. \quad (10)$$

- Sensitivity (SE)

$$SE = \frac{TP}{TN + FP} \times 100 \in [0 \ 100]. \quad (11)$$

- Specificity (SP)

$$SP = \frac{TN}{TN + FP} \times 100 \in [0 \ 100]. \quad (12)$$

- Matthews Correlation Coefficient (MCC) (13)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TN + FP)(TP + FN)(TN + FP)}} \in [-1,1].$$

Nonetheless, to evaluate the classifier performance in terms of adding more data to the training set, we calculate learning curves using a size from 0 to 2500. We only made that computing from feature set C, since it has all types of parameters extracted. Figure 6 illustrates such a calculation.

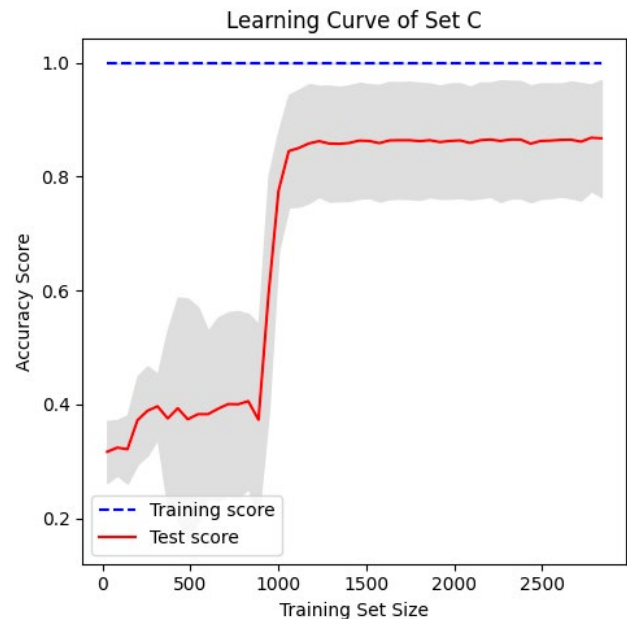


FIGURE 6. Learning curves generated to evaluate the performance of classification by RF and all parameters extracted (Feature set C).

RESULTS AND DISCUSSION

Under-sampling vs. oversampling

Since we have an unbalanced set of samples (78.9 % labeled as normal while 21.1 % as abnormal) it was

necessary to implement a strategy to equalize the number of samples (rows of a data frame) for each class (i.e., to have the same number of samples labeled as normal or abnormal in each data frame). Most of the algorithms tackle this issue by randomly under-sampling the majority class; however, the main drawback of this technique is that potentially useful information contained in the ignored samples is neglected. We address the unbalanced classes problem by adopting two strategies: dropping entities of the majority class and oversampling of the minority class using SMOTE [70]. In heart sound classification, obtaining new recordings labeled as abnormal is not a simple task (there will always be more healthy people). SMOTE allows us to use the already acquired data and create "new" samples in the feature space as if it were possible to access more pathological heart sounds. Table 1, section I shows the classification scores when using the input features from data frames A-E and comparing the abovementioned balancing techniques. When oversampling is applied, the *SP*, *ACC*, and *MCC* all increased, while *SE* has decreased.

TABLE 1. Results from the PCG sounds classification after splitting the recordings in High Quality (HQR) and Low Quality (LQR) labels.

Dataset	Balancing	SE	SP	ACC	MCC	Section
A	undersampling	88.06	79.03	81.3	0.61	I: HQR+LQR
B		74.22	70.13	71.16	0.4	
C		93.72	83.27	85.9	0.7	
D		94.34	82	84.32	0.67	
E		91.2	84.12	86.69	0.72	
A	oversampling	71.07	94.07	88.28	0.68	
B		27.05	95.56	78.29	0.33	
C		81.14	95.98	92.24	0.8	
D		79.25	95.56	91.45	0.77	
E		77.99	92.59	88.91	0.71	
A	undersampling	76.99	82.47	81.39	0.52	II: HQR
B		65.49	72.51	71.13	0.32	
C		87.61	82.25	83.3	0.6	
D		61.95	92.86	86.78	0.57	
E		87.61	82.25	83.3	0.6	
A	oversampling	61.95	92.86	86.78	0.57	
B		28.32	95.89	82.61	0.34	
C		76.11	93.94	90.43	0.7	
D		78.76	94.16	91.13	0.72	
E		78.76	91.13	88.7	0.66	
A	undersampling	70	69.44	69.64	0.38	III: LQR
B		60	58.33	58.93	0.18	
C		80	86.11	83.93	0.65	
D		80	86.11	83.93	0.65	
E		85	75	78.57	0.58	
A	oversampling	55	75	67.86	0.3	
B		45	83.33	69.94	0.31	
C		75	94.44	87.5	0.72	
D		75	91.67	85.71	0.68	
E		70	83.33	78.57	0.53	

Effects of signal quality on performance

We analyzed the influence of signal quality in the algorithm performance. According to the noise condition labels mentioned in the low-quality recordings subsection, we evaluated the signals tagged as HQR (2,874) and LQR (279) separately. The oversampling and under-sampling balancing procedures were also considered. Table 1 also presents the results for the HQR and LQR, respectively. As expected, the scores are generally higher for the HQR compared to the highly noisy recordings. A more detailed analysis of the results is provided in the following section.

Training time evaluation

To assess the practicality and efficiency of the proposed schemes, we evaluated the training time of the algorithm. The results are shown in Figure 7; it can be seen that for 2,500 features in the training set, the computation time is below 4 seconds. This result is unsurprising since the amount of data is relatively small to use more sophisticated and computationally expensive classification algorithms, such as Deep Learning methods.

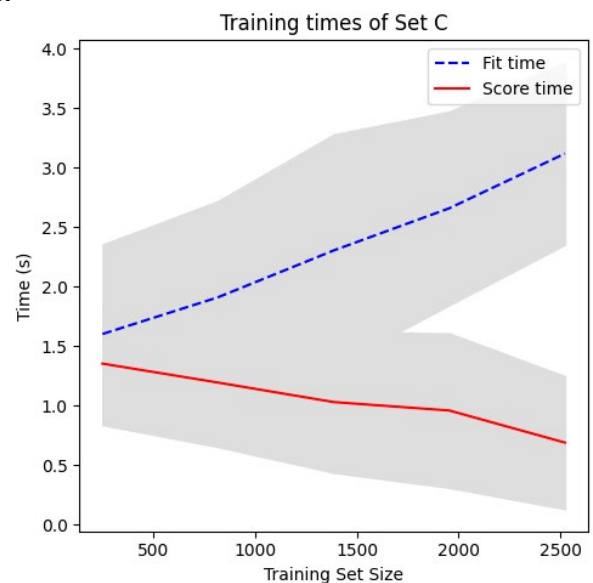


FIGURE 7. Time to fit/train the proposed algorithm and score or evaluate new instances. Results are shown for set C, since it contains the highest number of features.

Feature selection

In order to keep the best parameters for classification and reduce overfitting and computational complexity in the proposed classification scheme, we implemented feature selection. There are features or variables in our data that are the most relevant, i.e., those that contribute the most to the output prediction. In the present work, we applied the Correlation Feature Selection (CFS) [76] and Information Gain (IG) methods [77] for this task. The reduced subset was constructed for the IG method, neglecting features that provided a null information gain (zero). Table 2 shows the feature selection results, presenting the number of features before and after. The CFS method works significantly better regarding dimensionality reduction, keeping only between 12 and 30 features, while IG varies between 50 and 116 attributes.

TABLE 2. Number of features in the datasets A-E originally produced, then number of reduced features after applying CFS and IG feature selection.

Set	Original	Reduced CFS	Reduced IG
A	90	22	66
B	90	20	65
C	146	30	116
D	131	22	101
E	56	12	50

This section presents the classification performance evaluation when using feature selection by comparing the Receiver Operating Characteristics (ROC) curves and the calculation of the Area Under the Curve (AUC) for each feature set. Figure 8 shows the ROC curves for feature sets A-E. In this case, features set D exhibits the best performance since it has an AUC=0.97, the highest score obtained. However, set C is relatively close, showing AUC=0.969. This result suggests improving classification performance when adding MP and LPC param-

eters rather than only using MFCC or MP+LPC features. For cycle averaging, set B got the worst AUC score (0.79).

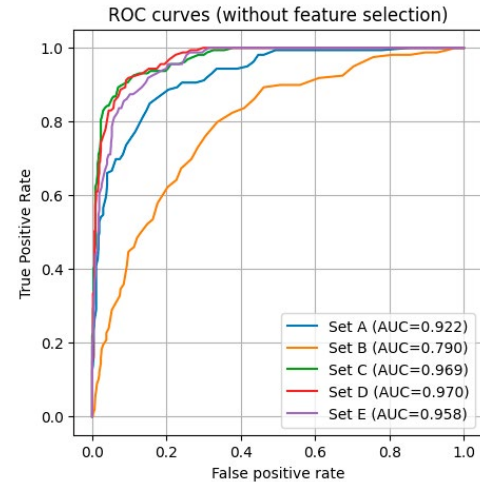


FIGURE 8. ROC curves and AUC for feature sets A-D without feature selection.

In another experiment, the ROC curves and AUC calculation was conducted when applying the CFS feature selection algorithm for each feature set, see Figure 9. There is an improvement in ACU scores since this metric increases for all feature sets. However, features set E now shows the best performance with an AUC=0.967. Feature sets C and D present a close AUC score of 0.961 and 0.967, respectively. On the other hand, although there is an increase from 0.76 to 0.867 for the AUC score of set B it is still the lowest obtained.

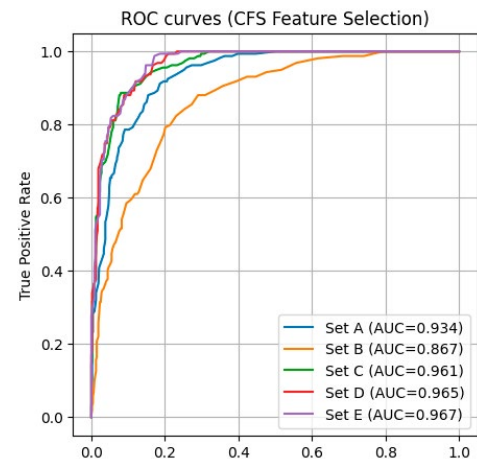


FIGURE 9. ROC curves and AUC for feature sets A-D after applying Correlation Feature Selection (CFS).

Finally, the same experiment was conducted but now using the features selected by the IG algorithm. Figure 10 shows the result. There is an improvement compared with results shown in Figure 9, when using CFS. However, feature sets C and D now show the best AUC score of 0.971, while in set D the AUC score is close (0.969). Set B is still presenting the worst ACU score (0.971).

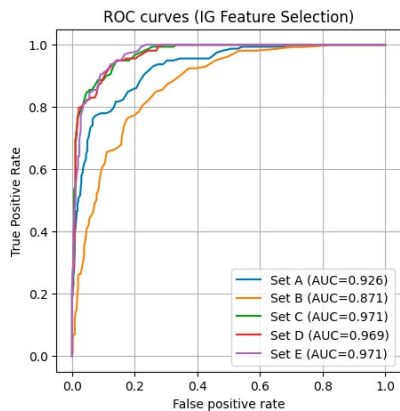


FIGURE 10. ROC curves and AUC for feature sets A-D after applying Information Gain Feature Selection (IG).

Discussion

This work aimed to compare different feature extraction schemes based on spectral and sparse representations for the automated classification of heart sounds. A low-cost system with accurate automatic analysis could prove very useful in assisting early diagnosis and improving the prognosis of patients with cardiovascular diseases. The Physionet/CinC 2016 Challenge provides the research community with the largest open database of annotated heart sounds; the research presented in this paper employed this dataset. An algorithm performance comparison with a universally standardized database contributes to promoting advances in the field of automated heart sound analysis. Our key objective is to evaluate the performance of different feature extraction, balancing, and feature selection techniques that can be relevant for effectively detecting heart murmurs. After thoroughly examining various classification schemes,

we selected the RF method for its simplicity and high performance [69]. The main reason to use MP+LPC as features stems from the heart sound reconstruction model we have previously proposed [71]. This sparse time-frequency model accurately represents the non-stationary behavior of the FHS and murmurs. The *de facto standard* feature for sound recognition are the MFCCs; they have been extensively used in PCG classification. In this work, we analyzed the classification performance for MP+LPC+MFCC. Comparing datasets A and B, all the output scores for set A are always higher than B's. In our tests, the feature averaging approach outperforms cardiac cycle averaging. We suppose that this fact results from the higher diversity produced when taking the mean value of the features rather than the direct calculation of the features from a single averaged cycle. The atomic decompositions in MP were performed using MPTK, the Matching Pursuit Toolkit [78].

Although dataset A displayed a good score (with best sensitivity $SE=88.5\%$) using undersampling, dataset E presents better results (best sensitivity $SE=91.19\%$). That is, classification based only on MFCC features outperforms the MP+LPC approach. Nevertheless, a combination of features improves the results as shown by the scores of datasets C and D. The merger of MP+LPC+MFCC ranked second in sensitivity $SE=93.17\%$, and dataset D (MP+MFCC) obtained the highest score ($SE=94.34\%$), both in case of using undersampling. In terms of feature selection, performance improved in terms of the AUC and ROC curve scores, the best obtained when applying the IG method. Feature sets D and E got an AUC=0.971, while set C obtained an AUC of 0.969, which is closer than the other sets. Although CFS produces a higher reduction in the number of features than the IG method and better results, the AUC scores are lower than when applying IG.

CONCLUSIONS

In this work, we analyzed the effects of detecting

cardiac murmurs when applying the SMOTE oversampling and random-under sampling methods for class balancing. Classification algorithms work best in cases where the number of samples is balanced. The reason is that they are designed to maximize accuracy and reduce error. SMOTE creates new abnormal PCG synthetic instances using the existing (real) ones. Although highly desirable, increasing the number of abnormal PCG sounds is a challenging task.

For this reason, we opted to use SMOTE. However, it is essential to be aware that this procedure might increase the likelihood of overfitting since it replicates the minority class events. The obtained SE scores were higher for the case of undersampling. However, the remaining scores *SP*, *ACC*, and *MCC* were improved when applying SMOTE.

The selection of sparse and spectral features helps classify PCG recordings under a high noise level without using FHS segmentation. For the group of LQR samples, the algorithm reached a $SE=85\%$ in feature set E applying under-sampling. On the other hand, the SMOTE oversampling effects produced lower scores in LQR when applying cycle averaging (data frame B). The best results were achieved when using all LQR+HQR samples of the dataset.

Finally, we compared the effectiveness of our method when using information gain (IG) and correlation feature selection (CFS). The results after applying dimensionality reduction were slightly higher using less features than when using the whole set of attributes. The highest sensitivity obtained was $SE=96.23\%$ when using under-sampling and feature set C for both CFS and IG, and for feature set D when using CFS. After inspecting the discarded features, we noticed that the phase of the time-frequency atoms selected by MP is not a relevant feature. This parameter was never chosen when using CFS, and it obtained a value of $IG=0$. This result is not surprising since, in general, the phase is considered a random variable uniformly

distributed in the $[-\pi, \pi]$ interval. In contrast, most of the MFCC were selected by both IG and CFS, in this case the first coefficients were ranked higher than the last. Regarding the other time-frequency atom parameters, the frequency, length, and position play a significant role in the classification. On the other hand, the amplitude has a low relevance. Only about a third of the coefficients were selected for the LPC without apparent order.

This research provides a complete assessment of feature selection methods for a classification algorithm to detect a pathological state from heart sounds. Different methods were evaluated, such as the balancing of samples, the comparison of MP+LPC vs. MFCC, and feature averaging vs. cycle time averaging as feature extraction methods. We also analyzed the effect of PCG signal quality and feature selection on classification performance. We selected the Random Forest technique algorithm to generate the classification model for PCG signals because the amount of data available is still small. Classical machine learning algorithms can often perform better than deep learning algorithms since they require a large amount of data for training. Nonetheless, classical machine learning algorithms are preferred when the interpretability of the model is essential since they are simpler and easier to understand [79].

The source code to reproduce the results of this paper can be downloaded free from; <https://github.com/roilhi/ABMEPaperPCGClassif.git/>.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

AUTHOR CONTRIBUTIONS

R.F.I.H. conceptualized the project, performed data curation, contributed to the research and methodol-

ogy, participated in the use of specialized software, and oversaw the project, obtained resources, and participated in the writing of the original draft of the manuscript. M.A.A.A. performed formal analyses, validated analyses, visualized results and supervised the development of the project, participated in the writing review and the editing of the manuscript. E.C.G.C. Obtained funding and economic resources, visualized results and supervised the development of the project, participated in the writing review and the editing of the manuscript. All authors reviewed and approved the final version of the manuscript.

REFERENCES

- [1] D. I. Macht, "On the absorption of drugs and poisons through the [1] World Health Organization (WHO), "Cardiovascular diseases (CVDs)," WHO. Available: <http://www.who.int/mediacentre/factsheets/fs317/en/> (accessed 2022).
- [2] A. K. Abbas, R. Bassam, "Phonocardiography signal processing," Springer Cham, 2009, pp. 194. [Online]. Available: <https://doi.org/10.2200/S00187ED1V01Y200904BME031>
- [3] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, R. G. Mark, "Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016", in: 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 609-612. [Online]. Available: <https://ieeexplore.ieee.org/document/7868816>
- [4] W. Zhang, J. Han, S. Deng, "Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation," Biomed. Signal Process. Control, vol. 53, art. no. 101560, Aug. 2019, doi: <https://doi.org/10.1016/j.bspc.2019.101560>
- [5] Y. Soeta, Y. Bito, Y. "Detection of features of prosthetic cardiac valve sound by spectrogram analysis," Appl. Acoust., vol. 89, pp. 28-33, Mar. 2015, doi: <https://doi.org/10.1016/j.apacoust.2014.09.003>
- [6] L. Orozco-Reyes, M.-Á. A. Arévalo, E. García-Canseco, R. F. Ibarra-Hernández, "Clasificación de la señal de audio cardíaco mediante la transformada de Fourier de tiempo corto y aprendizaje profundo," Res. Comput. Sci., vol. 151, no. 7, pp. 141-155, 2022. [Online]. Available: http://148.204.65.169/2022_151_7/Clasificacion%20de%20la%20senal%20de%20audio%20cardiaco%20mediante%20la%20transformada%20de%20Fourier%20de%20tiempo%20corto%20y.pdf
- [7] M. S. Obaidat, "Phonocardiogram signal analysis: techniques and performance comparison," J. Med. Eng. Technol., vol. 17 no. 6, pp. 221-227, 1993, doi: <https://doi.org/10.3109/03091909309006329>
- [8] N. E., Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Ten, C. C. Tung, H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. Lond., vol. 454, no. 1971, pp. 903-995, Mar. 1998, doi: <https://doi.org/10.1098/rspa.1998.0193>
- [9] V. Nivitha Varghees, K. I. Ramachandran, "Effective heart sound segmentation and murmur classification using empirical Wavelet transform and instantaneous phase for electronic stethoscope," IEEE Sens. J., vol. 17, no. 12, pp. 3861-3872, Jun. 2017, doi: <https://doi.org/10.1109/JSEN.2017.2694970>
- [10] S. Yuenyong, A. Nishihara, W. Kongprawechnon, K. Tungpimolrut, "A framework for automatic heart sound analysis without segmentation," BioMed. Eng. OnLine, no. 10, art. no. 13, Feb. 2011, doi: <https://doi.org/10.1186/1475-925X-10-13>
- [11] S. M. Debbal, F. Bereksi-Reguig, "Detection of Differences of the Phonocardiogram Signals by Using the Continuous Wavelet Transform Method," Int. J. Biomedical Soft Computing Hum. Sci., vol. 18 no.2, pp. 73-81, 2013, doi: https://doi.org/10.24466/ijb-schs.18.2_73
- [12] B. Ergen, Y. Tatar, H. O. Gulcur, "Time-frequency analysis of phonocardiogram signals using Wavelet transform: a comparative study," Comput. Methods Biomech. Biomed. Eng., vol. 15, no. 4, pp. 371-381, Jan. 2011 doi: <https://doi.org/10.1080/10255842.2010.538386>
- [13] S. A. Singh, T. G. Meitei, S. Majumder, "Short PCG classification based on deep learning," in Deep Learning Techniques for Biomedical and Health Informatics, B. Agarwal, V. E. Balas, L. C. Jain, R. C. Poonia, Manisha, Eds. London, United Kingdom: Academic Press, 2020, ch. 6, pp. 141-164. [Online]. Doi: <https://doi.org/10.1016/B978-0-12-819061-6.00006-9>
- [14] P. Qiao, Z. Yu, Z. Jingyi, C. Zhuo, "A method for diagnosing heart sounds in adolescents based on wavelet analysis and random forest," in 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Fuzhou, China, 2020, pp. 69-74, doi: <https://doi.org/10.1109/ICBAIE49996.2020.00021>
- [15] M. Abdollahpur, S. Ghiasi, M. J. Mollakazemi, A. Ghaffari, "Cycle selection and neuro-voting system for classifying heart sound recordings," in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 1-4. [Online] Available: <https://ieeexplore.ieee.org/document/7868814>
- [16] P. Lubaib, K.V. Ahammed Muneer, "The Heart Defect Analysis Based on PCG Signals Using Pattern Recognition Techniques," Proc. Technol., vol. 24, pp. 1024-1031, 2016, doi: <https://doi.org/10.1016/j.protcy.2016.05.225>
- [17] M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj, A. K. Katsaggelos, "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation, in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 613-616. [Online]. Available: <https://ieeexplore.ieee.org/document/7868817>
- [18] P. Wang, C. S. Lim, S. Chauhan, J. Y. A. Foo, V. Anantharaman, "Phonocardiographic signal analysis method using a modified hidden Markov model," Ann. Biomed. Eng., vol. 35, pp. 367-374, 2007, doi: <https://doi.org/10.1007/s10439-006-9232-3>
- [19] Y. Zheng, X. Guo, X. Ding, "A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification," Expert Syst. Appl., vol. 42, no. 5, pp. 2710-2721, Apr. 2015, doi: <https://doi.org/10.1016/j.eswa.2014.10.051>
- [20] S. Chauhan, P. Wang, C. Sing Lim, V. Anantharaman, "A computer-aided MFCC-based HMM system for automatic auscultation," Comput. Biol. Med., vol. 38, no. 2, pp. 221-233, Feb. 2008, doi: <https://doi.org/10.1016/j.combiomed.2007.10.006>
- [21] V. Maknickas, A. Maknickas, "Recognition of normal-abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients," Physiol. Meas., vol. 38, no. 8, art. no. 1671, Jul. 2017, doi: <https://doi.org/10.1088/1361-6579/aa7841>
- [22] G. Redlarski, D. Gradolewski, A. Palkowski, "A System for Heart Sounds Classification," PLoS One, vol. 9, no. 11, art. no. e112673, Nov. 2014, doi: <https://doi.org/10.1371/journal.pone.0112673>
- [23] B. M. Whitaker, P. B. Suresha, C. Liu, G. D. Clifford, D. V. Anderson, "Combining sparse coding and time-domain features for heart sound classification," Physiol. Meas., vol. 38, no. 8, art. no. 1701, Jul. 2017, doi: <https://doi.org/10.1088/1361-6579/aa7623>
- [24] I. J. Diaz Bobillo, "A tensor approach to heart sound classification," in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 629-632. [Online]. Available: <https://ieeexplore.ieee.org/document/7868821>
- [25] W. Zhang, J. Han, S. Deng, "Heart sound classification based on scaled spectrogram and tensor decomposition," Expert Syst. Appl., vol. 84, pp. 220-231, Oct. 2017, doi: <https://doi.org/10.1016/j.eswa.2017.05.014>

- [26] W. Zhang, J. Han, S. Deng, "Heart sound classification based on scaled spectrogram and partial least squares regression," *Biomed. Signal Process. Control*, vol. 32, pp. 20-28, Feb. 2017, doi: <https://doi.org/10.1016/j.bspc.2016.10.004>
- [27] P. Banerjee, A. Mondal, "An Irregularity Measurement Based Cardiac Status Recognition Using Support Vector Machine," *J. Med. Eng.*, vol. 2015, art. no. 327534, Oct. 2015 doi: <http://dx.doi.org/10.1155/2015/327534>
- [28] Y. Zheng, X. Guo, J. Qin, S. Xiao, "Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics," *Comput. Methods Programs Biomed.*, vol. 122, no. 3, pp. 372-383, Dec. 2015, doi: <https://doi.org/10.1016/j.cmpb.2015.09.001>
- [29] I. Maglogiannis, E. Loukis, E. Zafiroopoulos, A. Stasis, "Support vectors machine-based identification of heart valve diseases using heart sounds," *Comput. Methods Programs Biomed.*, vol. 95, no. 1, pp. 47-61, Jul. 2009, doi: <https://doi.org/10.1016/j.cmpb.2009.01.003>
- [30] A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, G. Castellanos-Dominguez, "Selection of Dynamic Features Based on Time-Frequency Representations for Heart Murmur Detection from Phonocardiographic Signals," *Ann. Biomed. Eng.*, vol. 38, pp. 118-137, 2010, doi: <https://doi.org/10.1007/s10439-009-9838-3>
- [31] L. D. Avendaño-Valencia, J. I. Godino-Llorente, M. Blanco-Velasco, G. Castellanos-Dominguez, "Feature Extraction From Parametric Time-Frequency Representations for Heart Murmur Detection," *Ann. Biomed. Eng.*, vol. 38, pp. 2716-2732, Jun. 2010, doi: <https://doi.org/10.1007/s10439-010-0077-4>
- [32] Singh, S. A., & Majumder, S. (2019). Classification of unsegmented heart sound recording using KNN classifier," *J. Mech. Med. Biol.*, vol. 19, no. 04, art. no. 1950025, doi: <https://doi.org/10.1142/S0219519419500258>
- [33] A. Sofwan, I. Santoso, H. Pradipta, M. Arfan A. A. Zahra M, "Normal and Murmur Heart Sound Classification Using Linear Predictive Coding and k-Nearest Neighbor Methods," in 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2019, pp. 1-5, doi: <https://doi.org/10.1109/ICICoS48119.2019.8982393>
- [34] P. Narváez, S. Gutierrez, W. S. Percybrooks, "Automatic Segmentation and Classification of Heart Sounds Using Modified Empirical Wavelet Transform and Power Features," *Appl. Sci.*, vol. 10 no.14, art. no. 4791, doi: <https://doi.org/10.3390/app10144791>
- [35] M. N. Homsí, N. Medina, M. Hernandez, N. Quintero, G. Perpiñan, A. Quintana, P. Warrick, "Automatic heart sound recording classification using a nested set of ensemble algorithms," in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 817-820. [Online]. Available: <https://ieeexplore.ieee.org/document/7868868>
- [36] C. H. Antink, J. Becker, S. Leonhardt, M. Walter, "Nonnegative matrix factorization and random forest for classification of heart sound recordings in the spectral domain," in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 809-812. [Online]. Available: <https://ieeexplore.ieee.org/document/7868866>
- [37] N. E. Singh-Miller, N. Singh-Miller, "Using spectral acoustic features to identify abnormal heart sounds," in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 557-560. [Online]. Available: <https://ieeexplore.ieee.org/document/7868803>
- [38] C. Potes, S. Parvaneh, A. Rahman, B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 621-624. [Online]. Available: <https://ieeexplore.ieee.org/document/7868819>
- [39] T.-E. Chen, S.-I. Yang, L.-T. Ho, K.-H. Tsai, Y.-H. Chen, Y.-F. Chang, Y.-H. Lai, S.-S. Wang, Y. Tsao, C.-C. Wu, "S1 and S2 Heart Sound Recognition Using Deep Neural Networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 372-380, Feb. 2017, doi: <https://doi.org/10.1109/TBME.2016.2559800>
- [40] F. Beritelli, G. Capizzi, G. Lo Sciuto, C. Napoli, F. Scaglione, "Automatic heart activity diagnosis based on Gram polynomials and probabilistic neural networks," *Biomed. Eng. Lett.*, no. 8, pp. 77-85, Feb. 2018, doi: <https://doi.org/10.1007/s13534-017-0046-z>
- [41] F. Li, H. Tang, S. Shang, K. Mathiak, F. Cong, "Classification of Heart Sounds Using Convolutional Neural Network," *Appl. Sci.*, vol. 10, no. 11, art. no. 3956, Jun. 2020, doi: <https://doi.org/10.3390/app10113956>
- [42] F. Demir, A. Şengür, V. Bajaj, K. Polat, "Towards the classification of heart sounds based on convolutional deep neural network," *Health Inf. Sci. Syst.*, vol. 7, art. no. 16, Aug. 2019, doi: <https://doi.org/10.1007/s13755-019-0078-0>
- [43] Y. Chen, S. Wei, Y. Zhang, "Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network," *Med. Biol. Eng. Comput.*, vol. 58, pp. 2039-2047, Jul. 2020, doi: <https://doi.org/10.1007/s11517-020-02218-5>
- [44] F. Renna, J. Oliveira, M. T. Coimbra, "Deep Convolutional Neural Networks for Heart Sound Segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2435-2445, Nov. 2019, doi: <https://doi.org/10.1109/JBHI.2019.2894222>
- [45] F. Li, M. Liu, Y. Zhao, L. Kong, L. Dong, X. Liu, M. Hui, "Feature extraction and classification of heart sound using 1D convolutional neural networks," *EURASIP J. Adv. Signal Process.*, vol. 2019, art. no. 59, Dec. 2019, doi: <https://doi.org/10.1186/s13634-019-0651-3>
- [46] B. Xiao, Y. Xu, X. Bi, J. Zhang, X. Ma, "Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption," *Neurocomputing*, vol. 392, pp. 153-159, Jun. 2020, doi: <https://doi.org/10.1016/j.neucom.2018.09.101>
- [47] S. Latif, M. Usman, R. Rana, J. Qadir, "Phonocardiographic Sensing Using Deep Learning for Abnormal Heartbeat Detection," *IEEE Sens. J.*, vol. 18, no. 22, pp. 9393-9400, Nov. 2018, doi: <https://doi.org/10.1109/JSEN.2018.2870759>
- [48] F. A. Khan, A. Abid, M. S. Khan, "Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features," *Physiol. Meas.*, vol. 41, no. 5, art. no. 055006, Jun. 2020, doi: <https://doi.org/10.1088/1361-6579/ab8770>
- [49] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, B.-W. On, "Heartbeat Sound Signal Classification Using Deep Learning," *Sensors*, vol. 19, no. 21, art. no. 4819, Nov. 2019, doi: <https://doi.org/10.3390/s19214819>
- [50] E. Messner, M. Zöhrer, F. Pernkopf, "Heart Sound Segmentation—An Event Detection Approach Using Deep Recurrent Neural Networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1964-1974, Sep. 2018, doi: <https://doi.org/10.1109/TBME.2018.2843258>
- [51] A. K. Dwivedi, S. A. Imtiaz, E. Rodriguez-Villegas, "Algorithms for Automatic Analysis and Classification of Heart Sounds—A Systematic Review," *IEEE Access*, vol. 7, pp. 8316-8345, 2019, doi: <https://doi.org/10.1109/ACCESS.2018.2889437>
- [52] S. Ismail, I. Siddiqi, U. Akram, "Localization and classification of heart beats in phonocardiography signals—a comprehensive review," *EURASIP J. Adv. Signal Process.*, vol. 2018, art. no. 26, 2018, doi: <https://doi.org/10.1186/s13634-018-0545-9>

- [53] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, C. Xu, "Deep Learning Methods for Heart Sounds Classification: A Systematic Review," *Entropy*, vol. 23, no. 6, art. no. 667, 2021, doi: <https://doi.org/10.3390/e23060667>
- [54] C. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. W. Johnson, et al., "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, art. no. 2181, Nov. 2016 doi: <https://doi.org/10.1088/0967-3334/37/12/2181>
- [55] S. G. Mallat, Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397-3415, Dec. 1993, doi: <https://doi.org/10.1109/78.258082>
- [56] R. F. Ibarra, M. A. Alonso, S. Villarreal, C. I. Nieblas, "A parametric model for heart sounds," in 2015 49th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2015, pp. 765-769, doi: <https://doi.org/10.1109/ACSSC.2015.7421237>
- [57] C.I. Nieblas, M. A. Alonso, R. Conte, S. Villarreal, "High performance heart sound segmentation algorithm based on Matching Pursuit," 2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), Napa, CA, USA, 2013, pp. 96-100, doi: <https://doi.org/10.1109/DSP-SPE.2013.6642572>
- [58] W. Wang, Z. Guo, J. Yang, Y. Zhang, L.-G. Durand, M. Loew, "Analysis of the first heart sound using the matching pursuit method," *Med. Biol. Eng. Comput.*, vol. 39, pp. 644-648, Nov. 2001, doi: <https://doi.org/10.1007/BF02345436>
- [59] Xuan Zhang, L. Durand, L. Senhadji, H. C. Lee, J.-L. Coatrieux, "Analysis-synthesis of the phonocardiogram based on the matching pursuit method," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 8, pp. 962-971, Aug. 1998, doi: <https://doi.org/10.1109/10.704865>
- [60] S. Jabbari, H. Ghasseman, "Modeling of heart systolic murmurs based on multivariate matching pursuit for diagnosis of valvular disorders," *Comput. Biol. Med.*, vol. 41, no. 9, pp. 802-811, Sep. 2011, doi: <https://doi.org/10.1016/j.compbiomed.2011.06.016>
- [61] W. Wang, J. Pan, H. Lian, "Decomposition and analysis of the second heart sound based on the matching pursuit method," in Proceedings 7th International Conference on Signal Processing, 2004, Beijing, China, 2004, pp. 2229-2232, vol. 3, doi: <https://doi.org/10.1109/ICOSP.2004.1442222>
- [62] P. P. Vaidyanathan, "The Theory of Linear Prediction," in *Synthesis Lectures on Signal Processing*, Switzerland: Springer Nature, 2022, pp. 1-184, doi: <https://doi.org/10.2200/S00086ED1V01Y200712SPR003>
- [63] T. Virtanen, M. D. Plumbley, D. Ellis, *Computational analysis of sound scenes and events*. Springer Cham, 2018, pp. 1-422, doi: <https://doi.org/10.1007/978-3-319-63450-0>
- [64] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5-32, 2001, doi: <https://doi.org/10.1023/A:1010933404324>
- [65] R. Gonzalez-Landaeta, B. Ramirez, J. Mejia, "Estimation of systolic blood pressure by Random Forest using heart sounds and a ballistocardiogram," *Sci. Rep.*, vol. 12, art. no. 17196, doi: <https://doi.org/10.1038/s41598-022-22205-0>
- [66] C. C. Balili, M. C. C. Sobrepena, P. C. Naval, "Classification of heart sounds using discrete and continuous wavelet transform and random forests," in 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 2015, pp. 655-659, doi: <https://doi.org/10.1109/ACPR.2015.7486584>
- [67] W. Xu, K. Yu, J. Ye, H. Li, J. Chen, F. Yin, et al., "Automatic pediatric congenital heart disease classification based on heart sound signal," *Artif. Intell. Med.*, vol. 126, art. no. 102257, Apr. 2022, doi: <https://doi.org/10.1016/j.artmed.2022.102257>
- [68] J. Oliveira, D. Nogueira, C. Ferreira, A. M. Jorge, M. Coimbra, "The robustness of Random Forest and Support Vector Machine Algorithms to a Faulty Heart Sound Segmentation," in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 1989-1992, doi: <https://doi.org/10.1109/EMBC48229.2022.9871111>
- [69] M. Y. Esmail, D. H. Ahmed, M. Eltayeb, "Classification System for Heart Sounds Based on Random Forests," *J. Clin. Eng.*, vol. 44, no. 2, pp. 76-80, 2019, doi: <https://doi.org/10.1097/JCE.0000000000000335>
- [70] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "Smote: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321-357, Jun. 2002, doi: <https://doi.org/10.1613/jair.953>
- [71] R. F. Ibarra-Hernández, M. A. Alonso-Arévalo, A. Cruz-Gutiérrez, A. L. Licona-Chávez, S. Villarreal-Reyes, "Design and evaluation of a parametric model for cardiac sounds," *Comput. Biol. Med.*, vol. 89, pp. 170-180, Oct. 2017, doi: <https://doi.org/10.1016/j.compbiomed.2017.08.007>
- [72] L. L. Vercio, M. Del Fresno, I. Larrabide, "Detection of morphological structures for vessel wall segmentation in ivus using random forests," in 12th International Symposium on Medical Information Processing and Analysis, Tandil, Argentina, 2017, art. no. 1016012, doi: <https://doi.org/10.1117/12.2255748>
- [73] A. Mellor, S. Boukir, A. Haywood, S. Jones, "Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 155-168, Jul. 2015, doi: <https://doi.org/10.1016/j.isprsjprs.2015.03.014>
- [74] R. F. Ibarra-Hernández, N. Bertin, M. A. Alonso-Arévalo, H. A. Guillén-Ramírez, "A benchmark of heart sound classification systems based on sparse decompositions," in 14th International Symposium on Medical Information Processing and Analysis, Maztlán, México, 2018, art. no. 1097505, doi: <https://doi.org/10.1117/12.2506758>
- [75] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825-2830, 2011. [Online] Available: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://>
- [76] M. A. Hall, "Correlation-based feature selection for machine learning," PhD dissertation, University of Waikato, Hamilton, New Zealand, 1999. [Online]. Available: <https://hdl.handle.net/10289/15043>
- [77] B. Azhagusundari, A. S. Thanamani, "Feature selection based on information gain," *IJITEE*, vol. 2, no. 2, pp. 18-21, Jan. 2013. [Online]. Available: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e17df473c25cccd8435839c9b6150ee61bec146a>
- [78] S. Krstulovic, R. Gribonval, "Mptk: Matching Pursuit Made Tractable," 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, Francia, 2006, pp. III-III, doi: <https://doi.org/10.1109/ICASSP.2006.1660699>
- [79] G. D. Clifford, C. Liu, B. Moody, J. Millet, S. Schmidt, Q. Li, et al., "Recent advances in heart sound analysis," *Physiol. Meas.*, vol. 38, no. 8, art. no. E10, Aug. 2017, doi: <https://doi.org/10.1088/1361-6579/aa7ec8>