PhD. In Electrical and Electronics Engineering

DOCTORAL THESIS

Synthesis of normal and abnormal heart sounds using Generative Adversarial Networks

Author: Pedro Juan Narváez Rosado
Thesis advisor: Dr. Winston Spencer Percybrooks Bolívar

A thesis submitted in fulfillment of the requirements for the degree of Doctor in Electrical and Electronics Engineering in the

Department of Electrical and Electronics Engineering

UNIVERSIDAD DEL NORTE

Barranquilla – Colombia, November /2022

TECHNICAL DATA	
PhD Thesis title	“Synthesis of normal and abnormal heart sounds using Generative Adversarial Networks”
Student ID	200111569
Author Name	Pedro Juan Narváez Rosado
E-mail	pjnarvaez@uninorte.edu.co
Thesis advisor	Dr. Winston Spencer Percybrooks Bolívar
Research group	BSPAI
Research line	Biomedical Signal Processing
Research sub line	Machine Learning
Delivery date of the proposal	June 2019
City / Country	Barranquilla / Colombia
VoBo. Thesis Advisor	

Acknowledgements

First of all, I thank GOD for allowing me to live this experience, obtain professional growth, receive many blessings and give me triumph and victory during many battles. All glory and honor be to GOD.

I thank my father Pedro Antonio (RIP) and my mother Celina who have always given me their unconditional support to be able to fulfill all my personal and academic goals. They are the ones who, with their love, have always encouraged me to pursue my goals and never abandon them in the face of adversity. I know that my father was always proud of my abilities and was convinced that I would achieve this professional title, although I will not be able to celebrate it with him in life, I will always remember every moment of happiness that I was able to generate for him throughout this period.

I deeply thank my tutor, Professor Winston Percybrooks, for his dedication and patience, without his words and precise corrections I would not have been able to reach this long-awaited stage. Thank you very much for the multiple words of encouragement from him, in those moments that I needed the most; for being there when my work hours got confusing. Thanks for his directions.

Thank all my colleagues, many of whom have become my friends, accomplices and brothers. Thank you for the hours shared, the work done together and the stories lived.

Finally, I would like to thank the university that has demanded so much of me, but at the same time has allowed me to obtain my long-awaited degree. I thank each manager for their work and for managing it, without which the foundations and conditions for learning knowledge would not exist.

List of Publications

1. Narváez, P.; Vera, K.; Bedoya, N.; Percybrooks, W. “Classification of Heart Sounds using Linear Prediction Coefficients and Mel-Frequency Cepstral Coefficients as Acoustic Features”. *In Proceedings of the IEEE Colombian Conference on Communications and Computing*, Cartagena, Colombia, 16 August 2017. (Chapter 4)

2. Narváez, P.; Gutierrez, S.; Percybrooks, W. “Automatic Segmentation and Classification of Heart Sounds using Modified Empirical Wavelet Transform and Power Features”. *Applied Science*. 2020, 10, 4791. (Chapter 4)

3. Narváez, P.; Percybrooks, W. “Synthesis of Normal Heart Sounds using Generative Adversarial Networks and Empirical Wavelet Transform”. *Applied Science*. 2020. (Chapter 5)

4. Narváez, P.; Percybrooks, W.; Giraldo, L. “This phonocardiogram does not exist: Adversarial model for the synthesis of heart sound and murmurs”. *Pending submission* (Chapter 5)

5. Narváez, P.; Percybrooks, W.; Giraldo, L. A Review of Biomedical Signals Synthesis using Generative Adversarial Networks. *Pending submission* (Chapter 2)

List of Awards:

- Google Research Awards for Latin America (LARA 2017), with the project: "Towards large scale, intelligent, computer-aided auscultation for remote primary-care settings" under the guidance of Prof. Winston Percybrooks.

- Google Research Awards for Latin America (LARA 2018), with the continuation of the project: "Towards large scale, intelligent, computer-aided auscultation for remote primary-care settings" under the guidance of Prof. Winston Percybrooks.

- Google Research Awards for Latin America (LARA 2019), with the continuation of the project: "Towards large scale, intelligent, computer-aided auscultation for remote primary-care settings" under the guidance of Prof. Winston Percybrooks.

Table of Contents

List of Figures --- 7

List of Tables -- 11

List of Abbreviations -- 13

Abstract -- 14

1. Introduction -- 15

2. State of the art and Contributions --- 18

2.1. Analysis of Heart sounds: State of the art -- 18

2.2. Synthesis of Heart sounds: State of the art --- 21

2.3. Synthesis of Biomedical Signals using GAN: State of the art -------------------------------- 24

2.3.1. Introduction to Generative Adversarial Network (GAN) --------------------------------------- 24

2.3.2. A review of GAN applied to biomedical signals --- 26

2.4. Contribution to the State of Art -- 27

3. Research hypothesis -- 28

4. Analysis of Heart sounds -- 28

4.1. Automatic segmentation of heart sounds -- 29

4.1.1. Proposed method --- 29

4.1.2. Results and discussion on the proposed automatic segmentation ------------------------------ 33

4.2. Automatic classification of normal and abnormal heart sounds ---------------------------- 36

4.2.1. Proposed method using LPC and MFCC as acoustic features --------------------------------- 36

4.2.1.1. Data -- 36

4.2.1.2. Feature extraction -- 36

4.2.1.3. Classification --- 37

4.2.1.4. Results --- 38

4.2.2. Proposed method using EWT and power features --- 39

4.2.2.1. Feature extraction -- 40

4.2.2.2. Experiments and results -- 42

5. Synthesis of Heart sounds -- 45

5.1. Synthesis of Normal Heart Sounds Using GAN and EWT ----------------------------------- 45

5.1.1. Proposed method -- 45

5.1.2. Experiments and Results --- 49

5.1.2.1. Results using Mel Cepstral Distortion (MCD) --- 51

5.1.2.2. Results using classification models --- 53

5.2. FeaturesGAN: adversarial model for the synthesis of heart sound and murmurs ------- 54

5.2.1. General description --- 55

5.2.2. Definition of mathematical models for the generation of normal heart sounds and types of murmurs -- 56

5.2.2.1. Cardiac murmurs -- 56

5.2.2.2. Mathematical model of cardiac murmurs --- 57

5.2.3. FeaturesGAN using MFCC features -- 60

5.2.4. FeaturesGAN using MFCC and FFT features --- 65

5.2.5. Analysis of results -- 68

5.2.5.1. Analysis of Generator and Discriminant models -- 68

5.2.5.2. PCA and t-SNE -- 72

5.2.5.3. MCD and SSIM --- 78

5.2.5.4. Classification models -- 80

5.2.5.5. MOS test -- 81

6. Conclusions and Future Work -- 82

7. Annexes --- 84

Bibliography --- 94

List of Figures

Figure 1. Cardiac cycle wave: A.) Normal; B.) Abnormal -- 15

Figure 2. A) Synthetic HS in the time domain; B) Natural signal in the time domain; C) Synthetic signal in the frequency domain; D) Natural signal in the frequency domain. --------------------------- 17

Figure 3. Synthetic heart sounds generated with state-of-the-art algorithms. The Y axis of each signal corresponds to the normalized amplitude. Signals taken from [27] ------------------------------- 22

Figure 4. General diagram of a GAN. --- 24

Figure 5. General diagram of a HS classification model using synthetic signals for training --------- 28

Figure 6. General diagram of the automatic heart sound classification system ------------------------- 28

Figure 7. Block diagram for the proposed segmentation system. --- 29

Figure 8. Decomposition of heart sound using EWT. (a) Heart sound recording; (b) FFT of heart sound with the frequency bands determined; (c) EWT components defined between 1–40 Hz. (d) EWT components defined between 40–80 Hz. (e) EWT components defined between 80–350 Hz. (f) EWT components defined at 350 Hz and above. --- 30

Figure 9. Stages of segmentation: a) EWT component: heart sound; b) NASE of the signal; c) identification of borders of each lobe; d) identification of S1 and S2 sounds. --------------------------- 32

Figure 10. Segmentation of heart sounds with S1 of low amplitude and short duration. a) NASE of heart sound; b) Identification of peaks; c) Heart sound with normalized intervals; d) Identification of S1 sounds. -- 33

Figure 11. Examples of segmentation using the proposed method. Heart sound signals for recordings (***A***) a031, (**B**) a0112, (**C**) a0284 and (**D**) a0352 are taken from the PhysioNet/CinC Challenge 2016, together with the successful segmentation using the proposed algorithm. ---------- 35

Figure 12: LPC+ MFCC feature distribution. A) for a normal HS; B) for an abnormal HS ---------- 37

Figure 13. Types of murmurs according to their location. --- 40

Figure 14. Distribution of power features for a normal heart sound. ------------------------------------- 41

Figure 15. General diagram of proposed method. -- 45

Figure 16. Proposed GAN diagram. --- 46

Figure 17. Architecture of the generator model. -- 47

Figure 18. Architecture of the discriminator model. -- 47

Figure 19. A) Real heart sound; B) Real heart sound with Gaussian noise; C) FFT of real heart sound; D) FFT of heart sound with Gaussian noise; E) EWT component of the noisy signal in the frequency range of 0 – 200 Hz; F) EWT component of the noisy signal in the frequency range greater than 200 Hz. -- 48

Figure 20. A) Synthetic signal with 100 epochs; B) synthetic signal with 500 epochs; C) synthetic signal with 1000 epochs; D) synthetic signal with 2000 epochs; E) synthetic signal with 2000 epochs + EWT; F) synthetic signal with 12000 epochs; G) natural signal of heart sound. ------------- 49

Figure 21. A) Synthetic heart sound generated by the model [21]; B) natural signal of the heart sound. -- 51

Figure 22. General diagram of the procedure to calculate the signal distortion. ------------------------ 52

Figure 23. Result of MCD distortions. --- 52

Figure 24. Result of MCD distortion using one synthetic signal with training dataset ---------------- 53

Figure 25. General diagram of FeaturesGAN -- 55

Figure 26. General architecture of FeaturesGAN -- 55

Figure 27. Examples of heart sounds with systolic murmurs. A) Pansystolic - Synthetic, B) Pansystolic - Real. C) Ejective - Synthetic; D) Ejective - Real; E) Protosystolic - Synthetic; F) Protosystolic - Real; G) Telesystolic - Synthetic; H) Telesystolic - Real. -------------------------------- 59

Figure 28. Examples of heart sounds with diastolic murmurs. A) Protodiastolic - Synthetic, B) Protodiastolic - Real. C) Mesodiastolic - Synthetic; D) Mesodiastolic - Real; E) Telediastolic - Synthetic; F) Telediastolic – Real. --- 59

Figure 29. General architecture of FeaturesGAN using MFCC features -------------------------------- 59

Figure 30. Architecture of the generator model for FeaturesGAN using MFCC features. ------------- 61

Figure 31. Architecture of the discriminant model for FeaturesGAN using MFCC features. --------- 61

Figure 32. Examples of cardiac signals concatenation in the time domain and their respective MFCC feature vector. A) Signal obtained from the mathematical model, B) Real signal obtained from a database, C) Signal used for discriminator training, D) Signal obtained after 5000 training epochs. -- 62

Figure 33. Result of cardiac signals in the frequency domain. A) Normal heart sound obtained from the mathematical model, B) Real signal, C) Synthetic signal obtained from the proposed FeaturesGAN model. -- 63

Figure 34. Examples of abnormal cardiac signals concatenation in the time domain and their respective MFCC feature vector. A) Signal obtained from the mathematical model, B) Real signal obtained from a database, C) Signal used for discriminator training, D) Signal obtained after 5000 training epochs. -- 64

Figure 35. Result of abnormal cardiac signals in the frequency domain. A) Normal heart sound obtained from the mathematical model, B) Real signal, C) Synthetic signal obtained from the proposed FeaturesGAN model. --- 65

Figure 36. General architecture of FeaturesGAN using MFCC and FFT features --------------------- 65

Figure 37. Architecture of the generator model for FeaturesGAN using MFCC and FFT features. -- 66

Figure 38. Architecture of the discriminant model for FeaturesGAN using MFCC and FFT features. --- 66

Figure 39. Examples of abnormal cardiac signals concatenation in the time domain and their respective FFT and MFCC feature vector. A) Signal used for discriminator training, B) Signal obtained after 10000 training epochs. -- 67

Figure 40. Result of abnormal signals in the time and frequency domain. A) Abnormal heart sound obtained from the mathematical model, B) Real signal, C) Synthetic signal obtained from the proposed FeaturesGAN model. -- 67

Figure 41. Features map of the generating model with 2000 training epochs. A) First convolutional layer with 128 filters (only the first 64 maps are shown); B) Second convolutional layer with 64 filters (Only the first 16 maps are displayed); C) Third convolutional layer with 1 filter; D) Dense layer with output of 2000 samples. -- 68

Figure 42. Bias visualization in the last dense layer in different training epochs. A) 10 epochs; B) 100 epochs; C) 500 epochs; D) 2000 epochs. --- 69

Figure 43. Discriminator output and losses. Green signal: Discriminator output; Red signal: Discriminator loss; Blue signal: Generator loss --- 70

Figure 44. Diagram of the modified generator model --- 70

Figure 45. Features map in the discriminator layers. A) Third convolutional layer with 64 filters; B) Fourth convolutional layer with 32 filters; C) Convolutional layer with 1 filter -------------------- 71

Figure 46. Discriminator outputs using 50 generated samples as the test data set in different training epochs. A) 10 epochs. B) 100 epochs. C) 500 epochs. D) 2000 epochs. ----------------------- 72

Figure 47. A) Clustering result using PCA; B) Clustering result using T-SNE; C) Signal obtained from the Mathematical Model; D) Real Cardiac Signal; E) Signal obtained from the FeaturesGAN model using MFCC. The Y-axis represents the normalized amplitude of the signal. ------------------ 73

Figure 48. Sound S1 results using FeaturesGAN with MFCC and FFT features. Sound S1 obtained by the mathematical model: A) Signal in the time domain; B) Signal in the frequency domain. S1 sounds extracted from a real cardiac signal: C) Signal in the time domain; D) Signal in the frequency domain. Sound S1 generated by FeaturesGAN: E) Signal in the time domain; F) Signal in the frequency domain. --- 74

Figure 49. A) Clustering result using PCA; B) Clustering result using T-SNE; C) Signal obtained from the Mathematical Model; D) Real Cardiac Signal; E) Signal obtained from the FeaturesGAN model using MFCC and FFT. The Y-axis represents the normalized amplitude of the signal --------- 74

Figure 50. Concatenated signals used in the experiment. A) Mathematical model signal. B) Actual signal. C) Signal used in the discriminator. D) Signal obtained from FeaturesGAN after 5000 epochs. --- 75

Figure 51. Clustering result using concatenated signals. A) PCA; B) t-SNE --------------------------- 76

Figure 52. Clustering result using MFCC features. A) PCA; B) t-SNE --------------------------------- 76

Figure 53. Concatenated signals used in the experiment. A) Mathematical model signal. B) Real signal. C) Signal used in the discriminator. D) Signal obtained from FeaturesGAN after 10000 epochs. --- 77

Figure 54. Clustering result using concatenated signals. A) PCA; B) T-SNE -------------------------- 77

Figure 55. Clustering result using FFT features. A) PCA; B) T-SNE ------------------------------------ 78

Figure 56. Clustering result using MFCC features. A) PCA; B) T-SNE -------------------------------- 78

Figure 57. Examples of heart sounds. A) Normal signal, B) Aortic Stenosis, C) Mitral Regurgitation D) Mitral Valve Prolapse --- 79

Figure 58. Mel-Cepstral Distortion (MCD) and Structural Similarity Index Measure (SSIM) results. A) Normal heart sounds; B) Abnormal heart sound with aortic stenosis; C) Abnormal heart sound with Mitral Regurgitation; D) Abnormal heart sound with Mitral Valve Prolapse. The red, blue, and green dots represent the real heart sounds, those obtained by FeaturesGAN, and those obtained by the mathematical model, respectively. --- 80

List of Tables

Table 1. Heart sound (HS) databases available on the web 16

Table 2. Previous methods for generation of synthetic audio. 17

Table 3. Summary of previous work in the automatic segmentation of heart sounds. 19

Table 4. Summary of the state-of-the-art approaches. Ref: Reference, N: normal, A: abnormal, T: total, NN: does not specify the data, F: features, A: accuracy, E: specificity, S: sensibility, DWT: discrete wavelet transform, LPC: linear prediction coefficients, SVM: support vector machine, KNN: K-nearest neighbor, RF: random forest, LB: LogitBoost, CSS: cost-sensitive classification, DNN: dense neural network, ANN: artificial neural network. 20

Table 5. Timeline of comparison methods for generation of synthetic heart sound 21

Table 6. A review of GAN applied in biomedical signals 26

Table 7. Database of heart sounds: Pascal Challenge. HS: heart sound. 33

Table 8. Results of segmentation for dataset A. (HB: heartbeat). 34

Table 9. Results of segmentation for dataset B. 34

Table 10. Results general of segmentation. 34

Table 11. Results of classifiers with LPC-only features (6 LPC per segment) 39

Table 12. Results of classifiers with MFCC-only features (14 MFCC per segment) 39

Table 13. Results of classifiers with LPC+MFCC features (14 MFCC and 6 LPC per segment) 39

Table 14. Comparison of heart sound classification results 39

Table 15. Features used in [25]. 41

Table 16. Features used in [26]. 41

Table 17. Summary of 1D-CNN and 2D-CNN model configurations in [28]. 42

Table 18. Summary of LSTM model configurations in [29]. 42

Table 19. Results of accuracy between the methods in [25–27] and the proposed method. 43

Table 20. Results of specificity between the methods in [25–27] and the proposed. 43

Table 21. Results of sensibility between the methods [25–27] and the proposed. 43

Table 22. Results of AUC between the methods [25–27] and the proposed. 43

Table 23. Comparison of results between the methods in [28,29] and the proposed method. 44

Table 24. Confusion matrix of proposed method. 44

Table 25. Parameters used in [**21**] to generate normal heart sounds. 51

Table 26. Features extracted in models [**49**], [5**0**], [**144**] and [**151**]. -------------------------------------- 53

Table 27. Accuracy results of the methods proposed in [49], [50], [144] and [151], taken from article [151]. -- 54

Table 28. Accuracy results of synthetic signals, using the trained models proposed in articles [49], [50], [144] and [151]. --- 54

Table 29. Accuracy results of synthetic signals without applying EWT algorithm, using the trained models proposed in articles [49], [50], [144] and [151]. --- 54

Table 30. Description of the different types of cardiac murmurs -- 56

Table 31. Definition of the term $\Delta(t)$ in different types of cardiac murmurs -------------------------- 57

Table 32. Parameters of systolic murmurs -- 58

Table 33. Parameters of diastolic murmurs -- 58

Table 34. Accuracy results of synthetic signals using the trained models proposed in articles [49], [50], [144] and [151]. --- 80

Table 35. Scales defined in the MOS test. --- 81

Table 36. MOS test results using real signals --- 81

Table 37. MOS test results using real and synthetic signals -- 81

Table 38. MOS test results using synthetic signals -- 82

List of abbreviations

HS: Heart sound

AS: Aortic Stenosis

MR: Mitral Regurgitation

MS: Mitral Stenosis

MVP: Mitral Valve Prolapse

GAN: Generative Adversarial Network

MFCC: Mel-Cepstral Frequency Coefficients

LPC: Linear Prediction Coefficients

EWT: Empirical Wavelet Transform

NASE: Normalized Average Shannon Energy

ML: Machine Learning

DL: Deep Learning

SVM: Support Machine Vector

ANN: Artificial Neural Network

KNN: K-Nearest Neighbor

RF: Random Forest

MCD: Mel-Cepstral Distortion

SSIM: Structural Similarity Index Measure

PCA: Principal Component Analysis

t-SNE: t-Distributed Stochastic Neighbor embedding

MOS: Mean Opinion Score

Abstract

Currently there are many works in the literature focused on the analysis of heart sounds, specifically on the development of intelligent systems for the classification of normal and abnormal heart sounds, many of these works have reported good results for precision, specificity, and sensitivity. However, the available heart sound databases are not yet large enough to train generalized Machine Learning models. Therefore, there is interest in the development of algorithms capable of generating heart sounds that could augment current databases.

In this doctoral thesis different methods proposed for the analysis and synthesis of normal and abnormal heart sounds are presented. During the development of this research, the following contributions to the state of the art were achieved:

i) An algorithm based on the empirical wavelet transform (EWT) and Normalized Average Shannon Energy (NASE) was implemented to improve the automatic segmentation stage of heart sounds. The results of this method were favorable compared to the state of the art using the same set of test data.

ii) Different feature extraction techniques were implemented for cardiac signals using Mel-Frequency Cepstral Coefficients (MFCC), Linear Prediction Coefficients (LPC) and power values. In addition, several Machine Learning models were tested, such as Support Machine Vector (SVM), K-Nearest Neighbors (KNN), Random Forest and Artificial Neural Network (ANN) for the automatic classification of normal and abnormal heart sounds. The results obtained in each of the tests show that the proposed methods deliver better results of accuracy, specificity and sensitivity compared to works published in the state of the art.

iii) A model based on Generative Adversarial Networks (GAN) was designed to generate normal synthetic heart sounds. Additionally, a denoising algorithm is implemented using the Empirical Wavelet Transform (EWT), allowing a decrease in the number of epochs and the computational cost that the GAN model requires. A distortion metric (Mel-Cepstral Distortion) was used to objectively assess the quality of synthetic heart sounds. The proposed method was favorably compared with a mathematical model that is based on the morphology of the Phonocardiography (PCG) signal published in the state-of-the-art. Additionally, different heart sound classification models proposed in the state-of-the-art were also used to test the performance of such models when the GAN-generated synthetic signals are used as test data. In this experiment, good accuracy results were obtained with most of the implemented models, suggesting that the GAN-generated sounds correctly capture the characteristics of natural heart sounds.

iv) Finally, a model based on the GAN architecture is proposed, which consists of refining synthetic cardiac signals obtained by a mathematical model with characteristics of real cardiac signals. This model has been named FeaturesGAN and does not require a large database to generate different types of heart sounds. Different metrics were also used to evaluate the quality of cardiac signals, such as: MCD, Structural Similarity Index Measure (SSIM), Principal Component Analysis (PCA) and t-SNE. Similarly, classification tests were performed using the synthetic signals generated by FeaturesGAN as a test data set in different Machine Learning models published in the state of the art. Finally, subjective evaluations based on Mean Opinion Score (MOS) tests were carried out with expert doctors, in order to validate the quality of the audios. All the results obtained in the different tests and metrics were satisfactory, indicating that the normal and abnormal heart sounds generated by the FeaturesGAN model have very similar characteristics to real signals.

1. Introduction

According to the World Health Organization (WHO), cardiovascular diseases are among the leading causes of death worldwide **[1]**. An updated report from the American Heart Association on heart diseases statistics shows that 31% of deaths worldwide are generated by heart diseases (about 17.7 million each year) **[2]**. Typically, people who live in rural zones face an overall lower quality of life than their urban counterparts, for example having limited access to even primary-care programs for prevention and early detection of heart conditions **[2]**. Tobacco use, unhealthy eating, and lack of physical activity are the main causes of heart disease **[1]**.

Currently, there are sophisticated equipment and tests for diagnosing heart disease, such as: electrocardiogram, holter monitoring, echocardiogram, stress test, cardiac catheterization, computed tomography scan, and magnetic resonance imaging **[3]**. However, most of this equipment is very expensive, and must be used by specialized technicians and medical doctors, which limits its availability in rural and urban areas that do not have the necessary financial resources **[4]**. Therefore, even today, it is common in such scenarios for non-specialized medical personnel to rely on basic auscultation with an stethoscope as a primary screening tool for the detection of many cardiac abnormalities and heart diseases **[5]**. However, to be effective, this method requires a sufficiently trained ear to identify cardiac conditions. Unfortunately, the literature suggests that in recent years such auscultation training has been in decline **[6–8]**.

The genesis of heart sounds is closely related both to the vibration of the entire myocardial structure and to the vibration of the heart valves during closure and opening. A recording of heart sounds is composed of a sequence of cardiac cycles. A normal cardiac cycle is composed of the S1 section (generated by the closing of the atrioventricular valve), the S2 section (generated by the closing of the semilunar valve), the systole (located between S1 and S2) and diastole (located between S2 of the current cycle and S1 of the next one) **[9]**. Abnormalities are represented by murmurs that usually occur in the systolic or diastolic intervals **[10]**. In Figure 1, examples of a normal and an abnormal cardiac cycle are shown.

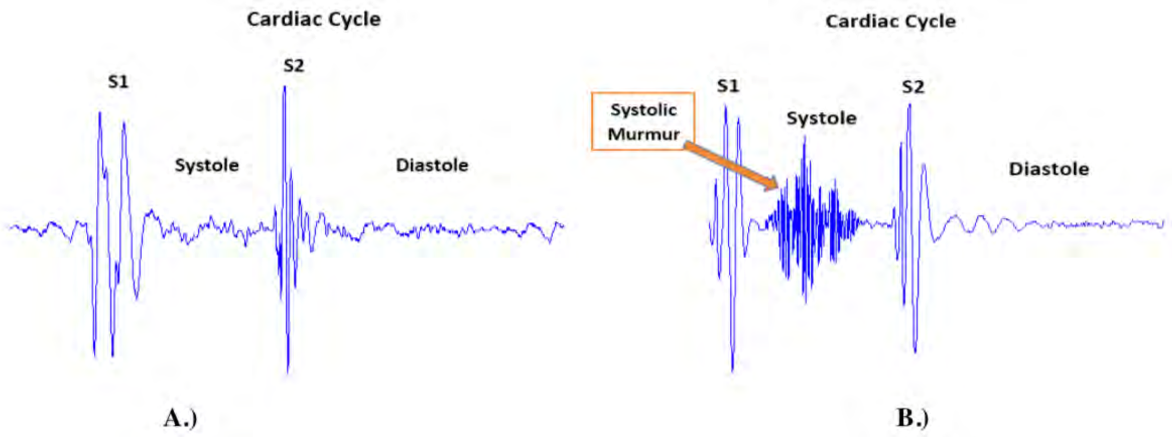


Figure 1. Cardiac cycle wave: A.) Normal; B.) Abnormal

Health professionals use different attributes of heart murmurs for their classification, the most common are: Timing, cadence, duration, pitch and shape of murmur **[11]** and **[12]**. They must have an ear sufficiently trained to identify each of these attributes.

Therefore, the development of a system that allows a classification of normal and abnormal heart sounds would be of great help for arriving to an accurate diagnosis by health professionals. In addition, it could generate a positive impact in those rural zones that have high mortality rates due to heart disease and do not have staff and equipment specialized in cardiology [2]. However, access to data for the construction of these intelligent systems is one of the main limitations today.

To date, many investigations related to the classification of normal and abnormal heart sounds have been published. Good accuracy, specificity and sensitivity results have been reported. However, in terms of predicting the type of abnormality, no outstanding advances have been found in the literature due to the lack of datasets labeled with their respective types of anomaly.

Most of the published works are based on machine learning, using supervised learning. Therefore, a labeled database with numerous samples of normal and abnormal HS is needed to train the classification model. Table 1 describes the different databases that are available on the web and have been used in several research works.

Table 1. Heart sound (HS) databases available on the web

Data Base	**Total # of HS samples**	**Data Base**	**Total # of HS samples**
University of Michigan Health System [13]	23	University of Haute Alsace **[16]**	79
University of Washington [14]	16	Dalian University of Technology **[16]**	673
Thinklabs [15]	105	Shiraz University **[16]**	114
Massachusetts Institute of Technology [16]	409	Skejby Sygehus Hospital **[16]**	35
Aalborg University [16]	695	Shiraz University fetal HS **[16]**	211
Aristotle University of Thessaloniki [16]	45	iStethoscope Pro iPhone app **[17]**	176
Tossi University of Technology [16]	174	Digital stethoscope DigiScope **[17]**	656

The main limitations of these databases are: i.) They are not sufficiently labeled, for example, they do not describe the type of abnormality; ii.) There is no balance between the number of normal and abnormal samples; iii.) It is possible that the size of the database is not sufficient to obtain a generalized classification model and even less to classify specific types of abnormalities.

It should be mentioned, that a recording of heart sounds can be mixed with external noise (from the environment) or they can present innocuous murmurs that do not represent a cardiac condition, for this reason the classification system can be mistaken when selecting some type of cardiac anomaly.

Machine Learning models have been published on the classification of types of anomalies, such as: aortic stenosis, mitral stenosis, aortic regurgitation, mitral regurgitation and ventricular defect. These works do not use enough training samples to obtain good generalization and in several cases the cardiac signals are generated by a simulator. Then, these results do not guarantee a good performance of the algorithm if used in the field.

One solution to overcome these limitations is to build a database with abnormal HS types, well labeled with the help of cardiologists. However, this task requires a lot of time and dedication, since many samples of each type of abnormality are required. Another possible solution is the implementation of a model for the generation of synthetic sounds, capable of outputting varied synthetic heart sounds indistinguishable from natural ones by medical personnel. Such model could be used to augment existing databases for training robust machine learning models. However, heart sound signals are highly non-stationary, and their level of complexity makes obtaining good generative models very challenging.

In the literature, there are several publications related to the generation of synthetic heart sounds **[18-27]**. All these works are based on mathematical models to generate the S1 and S2 sections of a cardiac cycle, and to date there is no model that allows generating types of murmurs. On the other hand, the systolic and diastolic intervals of the cardiac cycle are not adequately modeled, and as a result, do not present the variability recorded in natural normal heart sounds. Therefore, these synthetic models are not suitable to train HS classification models. Additionally, a basic time-frequency analysis of these synthetic signals shows that they are very different from natural signals (see Figure 2).

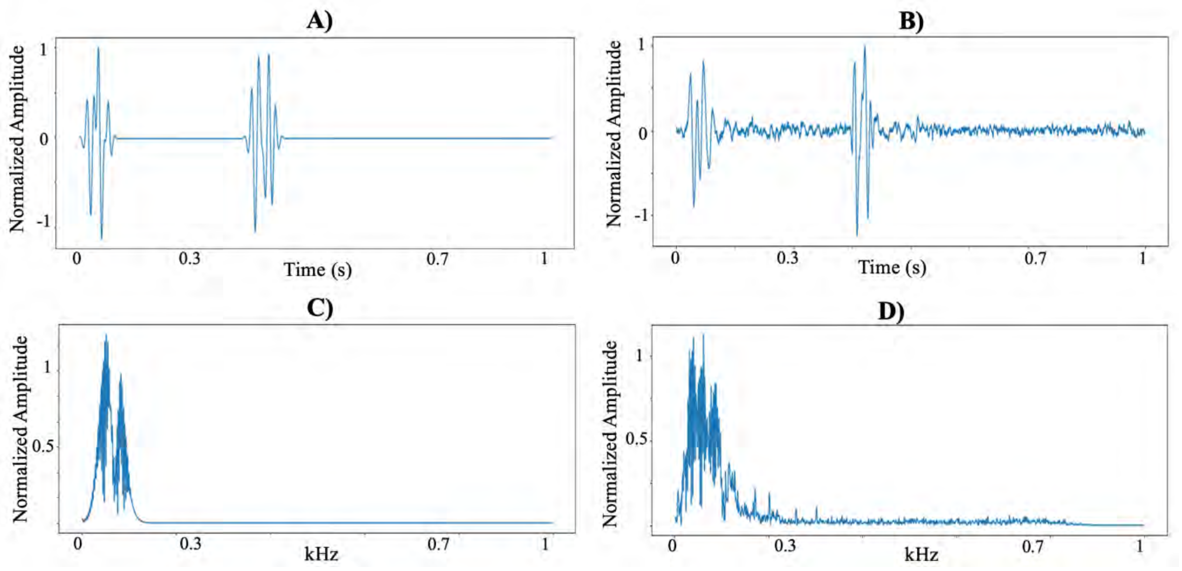


Figure 2. A) Synthetic HS in the time domain; B) Natural signal in the time domain; C) Synthetic signal in the frequency domain; D) Natural signal in the frequency domain.

On the other hand, in recent years there have been great advances in the synthesis of audio, mainly speech, using machine learning techniques, specifically with deep learning. In Table 2, several proposed methods to improve audio synthesis are presented. Therefore, taking into account that cardiac sounds are also audio signals and are perceived by human ears through the auscultation technique, it could be considered that these Deep Learning methods could also present good results in the synthesis of cardiac signals.

Table 2. Previous methods for generation of synthetic audio.

Year	Author	Proposed Method	Synthetic Signal
2015	Aaron van den Oord et al. **[28]**	WaveNet: Probabilistic and autoregressive model based on deep neural networks (DNN)	Music. Text to speech
2017	Jesse Engel et al. **[29]**	WaveNet and Autoencoders	Musical notes
2017	Bajibabu Bollepalli et al. **[30]**	Generative Adversarial Network (GAN)	Glottal waveform
2018	Giorgio Biagetti et al. **[31]**	Hidden markov model	Text to speech
2019	Chris Donahue et al. **[32]**	WaveGAN: GANs unsupervised	Intelligible words

According to the limitations presented in the current models for the generation of synthetic heart sounds and taking into account the significant advances in voice synthesis using deep learning methods, in this work we propose a model based on generative adversarial networks (GANs) to generate heart sounds that can be used to train machine learning models. In addition, we do an analysis of cardiac signals proposing models for the automatic segmentation and classification of heart sounds, which are compared with the state of the art. In this sense, having a model for the generation of synthetic sounds, capable of outputting varied synthetic heart sounds indistinguishable from natural ones by medical personnel, could be used to augment existing databases for training robust machine learning models.

This document is organized as follows: Section 2 presents the state of the art and contributions in the analysis and synthesis of heart sounds; Section 3 indicates the research hypothesis; in Section 4, the proposed methods and results obtained in the automatic segmentation and classification of heart sounds are described; section 5 details the methods proposed in the generation of heart sounds, and the results obtained in each experiment; and finally the document concludes in Chapter 6 with conclusions and further work.

2. State of the art and Contributions

In this chapter, we present published advances in the state of the art related to the analysis and synthesis of heart sounds. In the case of analysis, we emphasize the proposed methods or signal processing techniques for the segmentation and classification automatic of heart sounds. On the synthesis side, we divide the state of the art into two approaches: i) The generation of heart sounds based on mathematical models through systems of differential equations, in order to obtain a real representation of cardiac signals; and ii) progress in the implementation of Generative Adverse Networks (GAN) for the generation of biomedical signals, such as electrocardiography (ECG), electroencephalography (EEG), electromyography (EMG), and photoplethysmography (PPG).

2.1. Analysis of Heart sound: State of the art

A system for automatic classification of cardiac sounds has four main stages: i.) Pre-processing: It attenuates the unwanted noise that was acquired at the time of recording the sound. ii.) Segmentation: In this stage the boundaries for each cycle's segment (S1, S2, systole, diastole) are determined. iii.) Feature Extraction: Signal processing techniques are used to extract relevant characteristics from each type of cardiac signal (normal and abnormal), which helps to discriminate one class from another. iv.) Classification: Machine learning models are generally used in this stage, where the input corresponds to the characteristics extracted in the previous step.

Different methods have been proposed in the state-of-the-art approaches to create an intelligent system that allows for discrimination between normal and abnormal heart sounds. Below is a list of several works related to the pre-processing, segmentation, feature extraction and classification of these signals.

In [33], the envelope of the cardiac signal was calculated from the normalized average Shannon energy (NASE), specifying a threshold to identify the candidate peaks for S1 and S2; then, several criteria were used for the selection of definitive peaks for S1 and S2.

The method used in [34] calculates the Shannon energy of the local spectrum calculated by the S-transform for each sample of the heart sound signal. Finally, the authors extracted features based on the singular value decomposition (SVD) of the matrix S to distinguish between S1 and S2.

The decomposition and reconstruction of cardiac signals with fifth-level discrete wavelets using the frequency bands of the approximation coefficient a4 (0 to 69 Hz) and the detail coefficients d3 (138 to 275 Hz), d4 (69 to 138 Hz) and d5 (34 to 69 Hz) have been used as an alternative to extract normal heart sounds and facilitate the identification of S1 and S2 [35].

In several studies, the decomposition wavelet has been used to reduce signal noise and hidden Markov models (HMM) to segment the signal, considering each segment (i.e., S1, systole, S2 and diastole) as a state [36]. One of the most frequently used segmentation algorithms in the state-of-the-art models was proposed in [37]; this algorithm is based on logistic regression and hidden semi-Markov models and uses electrocardiographic signals (ECG) as a reference to make the annotations of the four segments of a heart cycle (S1, S2, systole and diastole). However, this method fails when the cardiac signal has significant murmurs that are longer than normal heart cycles and there is an irregular sinus rhythm. In [38], the authors present the performance of the algorithm by testing it with different databases of heart sounds and describe the limitations mentioned above.

In [39] a fourth-level, sixth-order Daubechies filter was used on the sound signal. The authors removed all the details at each level and reconstructed the signal using the approximation coefficients. Finally, they used the spectrogram to extract the signal below 195 Hz. In [40], a Butterworth band pass filter with an order of two and cutoff frequencies from 25 to 400 Hz was applied to the signal. The spikes were removed from the signal using a spike removal algorithm, as presented by Schmidt [41]. Subsequently, the authors used a homomorphic filter to extract the envelope of the cardiac signal. In [42], a fifth-order Chebyshev type I low-pass filter with cut-off frequencies of 100 Hz and 882 Hz was applied. Then, the Shannon envelope of the normalized signal was computed. Similarly, in [43], the authors used a sixth-order Butterworth bandpass filter with cut-off frequencies of 25 Hz and 900 Hz and then extracted the signal envelope using Shannon's average energy.

In general, although all the methods listed above propose different types of filters in their pre-processing stage, they are not yet sufficiently efficient to attenuate unwanted signals (external noise, murmurs) and amplify S1 and S2 sounds; therefore, the segmentation algorithm can easily make errors in the identification of each cardiac cycle. On the other hand, several segmentation algorithms use fixed amplitude thresholds to detect S1 and S2 sounds, but these can fail when these sounds have a low amplitude that does not exceed the stipulated threshold, as well as when some unwanted noise exceeds the threshold. Table 3 summarizes different processing methods proposed in the literature along with their year of publication.

Table 3. Summary of previous work in the automatic segmentation of heart sounds.

Author	Dataset	Method
[33]	37 recordings of heart sounds	Normalized Average Shannon Energy
[34]	80 recording of heart sounds	S-transform and Shannon Energy
[35]	77 recording of heart sounds	Wavelet decomposition and reconstruction (normalized average Shannon energy (NASE))
[36]	Physionet [44] and Pascal [45]	Hidden Markov model (HMM)
[37]	Physionet database [44]	Logistic regression and hidden semi-Markov model (HSMM)
[39]	Pascal database [45]	Wavelet Decomposition and Spectrogram
[40]		Butterworth band pass filter with order 2 and Homomorphic filter
[42]		Chebyshev type I low-pass filter and Shannon envelope
[43]		Band-pass Butterworth sixth-order filter and Shannon envelope
[46]	Physionet database [44]	HSMM–convolutional neural network (CNN)
[47]	Pascal database [45]	Discrete Wavelet Transform and Hilbert transformation
[48]	Physionet [44] and Pascal [45]	Adaptive sojourn time HSMM

The Physionet and Pascal databases are the most widely used databases for the analysis of heart sounds. Physionet is composed of eight data sets and contains 3153 recordings in total, lasting from 5 seconds to just over 120 seconds, of which 2488 recordings are labeled as normal and 665 recordings are labeled as abnormal, presenting an imbalance in the data set [44]. The Pascal database is composed of two data sets and contains the following categories: normal (351 recordings), murmur (129 recordings), extra heart sound (65 recordings) and artifact (40 recordings), the audio files are of varying lengths, between 1 second and 30 seconds. Unlike Physionet, the manual segmentation of a group of heart sound recordings is specified in this database [45].

On the other hand, different techniques for the extraction of characteristics and machine learning models have been proposed for the automatic classification of heart sounds in order to obtain good accuracy, specificity and sensitivity in the results. Many researchers have worked with features in the time and frequency domains, as well as perceptual-based features. Table 4 shows the results obtained from various authors, specifying the number of cardiac cycles used for the experiment, the segmentation and feature extraction methods and the machine learning (classifier) model.

In [49], 19 Mel-frequency cepstral coefficients (MFCCs) and 24 discrete wavelet transform (DWT) features were extracted from a cardiac cycle. These features were used as inputs to a Support Vector Machine (SVM) model to classify normal and abnormal heart sounds, obtaining an accuracy of 97%.

Table 4. Summary of the state-of-the-art approaches. Ref: Reference, N: normal, A: abnormal, T: total, NN: does not specify the data, F: features, A: accuracy, E: specificity, S: sensibility, DWT: discrete wavelet transform, LPC: linear prediction coefficients, SVM: support vector machine, KNN: K-nearest neighbor, RF: random forest, LB: LogitBoost, CSS: cost-sensitive classification, DNN: dense neural network, ANN: artificial neural network.

Ref	Number of heart cycles	Segmentation	Feature Extraction (Number of features)	Classifier	Results
[49]	N:200; A:800	Manually	MFCC + DWT (43 F)	SVM	A: 97%
[50]	N: 2488; A: 665	Not applicable	Statistical domain, Frequency domain and MFCC (27 F)	XgBoost	S: 94.5%; E: 91.3%; A: 92.9%
[51]	N: 2488; A: 665	Manually	1D – Convolutional neural network (CNN)	DNN	S: 91.5%; E: 71.6%; A: 87.5%
			MFCC and 2D – CNN	DNN	S: 92.5%; E: 76.6%; A: 89.3%
[52]	N: 320; A: 141	Not applicable	Long short-term memory (LSTM)	DNN	A: 80.8%
[53]	N: 399; A: 399	[13]	Time, time-frequency and perceptual domain. (90 F)	ANN	S: 90,1%; E: 93,1%
[54]	N: 669; A: 722	[13]	MFSC and CNN	DNN	A: 93,7%
[55]	N: 2488; A: 665	[13]	Time and frequency domain, wavelet and statistics (29 F)	RF + LB + CSS	S: 79.6%; E:80.6%
[56]		Hilbert transformation	Statistical properties, Heart rate (53 F)	Logic rules	S: 91.3%; E:77%
[57]			CWT	SVM, KNN	A: 86%
[58]		[13]	Time features and MFCC (13 F)	SVM	S: 91.8%; E: 82%; A: 97%
[59]		[13]	Sparse coding and time domain features (25 F)	SVM	S: 84.3%; E: 77.2%; A: 80.7%
[60]		[13]	Time domain, frequency domain and entropy (40 F)	BP Neural networks	S: 68.3%; E: 94%; A: 88.5%
		[13]		Logistic Regression	S: 75.6%; E: 87.7%; A: 72.5%
[61]	N: 115; A: 287	Manually	Curve fitting, MFCC (25 F)	KNN	A: 92%
	N: 386; A: 103				A: 81%
	N: 7; A: 22				A: 98%
[62]	N: 40; A: 80	Manually	DWT (50 F)	KNN, Fuzzy C-Means	A: 86.67%
[63]	N: 336; A: 171	DWT and Shannon energy	Spectrogram and tensor decomposition (100 F)	SVM	A: 83%
	N: 2575; A: 665				A: 88%

In [50], the authors extracted a total of 27 features: 13 MFCC features, 10 statistical features and four frequency features. In that work, they used an XgBoost algorithm, obtaining an accuracy of 92.9%. In [53], a total of 90 features were extracted in the time domain, time–frequency and perceptual dimensions. These were used as the input to a two-layer feed forward neural network, achieving 90.1% sensitivity and 93.1% specificity in the validation process, using samples that present high sound quality. In [54], deep convolutional neural networks and MFCCs were used for the recognition of normal and abnormal heart sounds, achieving an overall accuracy of 84.15%.

The experiments in [50,51,55–60,63] were performed with the Physionet database [44]. In [55], features were extracted in time, frequency, wavelet and statistics, obtaining a total of 29 features. In the classification stage, a nested set of ensemble algorithms consisting of random forest (RF), LogitBoost (LB) and cost-sensitive classification (CSS) were used, obtaining an overall accuracy of an 80.1%, specificity of 80.6% and sensibility of 79.6%.

Tables 3 and 4 show different methods used for automatic segmentation and classification of heart sounds, respectively. However, the techniques used in segmentation tend to fail when the sound signal contains murmurs or external noise with amplitude peaks equal or greater than the peaks of the S1 or S2 sounds. As for automatic classification, it tends to require very complex feature extraction and classification models that limit its use in real-time applications due to their high computational cost.

2.2. Synthesis of Heart sound: State of the art

Despite important research progress in the analysis of PCG signals, with good results for accuracy, specificity and sensitivity, the number of samples used for training does not guarantee a generalizable model, and even less the suitability for classification of types of abnormalities. Therefore, the size of the databases is a great limitation for the advancement of this research.

An alternative could be the generation of synthetic heart sounds that serve to train the classification model. However, in the state of the art we have not found studies that use this type of synthetic signals to train machine learning models and compare the performance of the classifier using real signals in the tests. Table 5 presents some published works on the synthesis of heart sounds, this information was obtained from the Web of Science, Scopus and IEEE databases. Figure 3 shows the results of the signals obtained by several proposed models and a real normal cardiac signal. Several of these mathematical models are described below.

Table 5. Timeline of comparison methods for generation of synthetic heart sound

Year	Author	Proposed method
1992	Y. Tang et al. [18]	The exponentially damped sinusoid model
1995	T. Trang et al. [19]	S1 and S2 are modeled as transient-linear-chirp signals
1998	X. Zhang et al. [20]	The model based on a sum of Gaussian modulated sinusoids
2000	J. Xu et al. [21]	S1 and S2 are modeled as transient-nonlinear-chirp signal
2009	C. Toncharoen et al. [22]	A heart-sound-like chaotic attractor
2011	Almasi et al. [23]	A Dynamical Model for Generating Synthetic Phonocardiogram Signals.
2012	Tao et al. [24]	Modify the amplitude and width of S1 and S2 sounds from real heart sound and combine it with noise.
2013	Jablouna et al. [25]	A model based on three coupled ordinary differential equations
2018	Saederup [26]	Estimation of the second heart sound split using windowed sinusoidal models
2018	Joseph et al. [27]	A sum of almost periodically recurring deterministic “wavelets”. S1 and S2 are modeled by two sinusoidal pulses of Gaussian modulation.

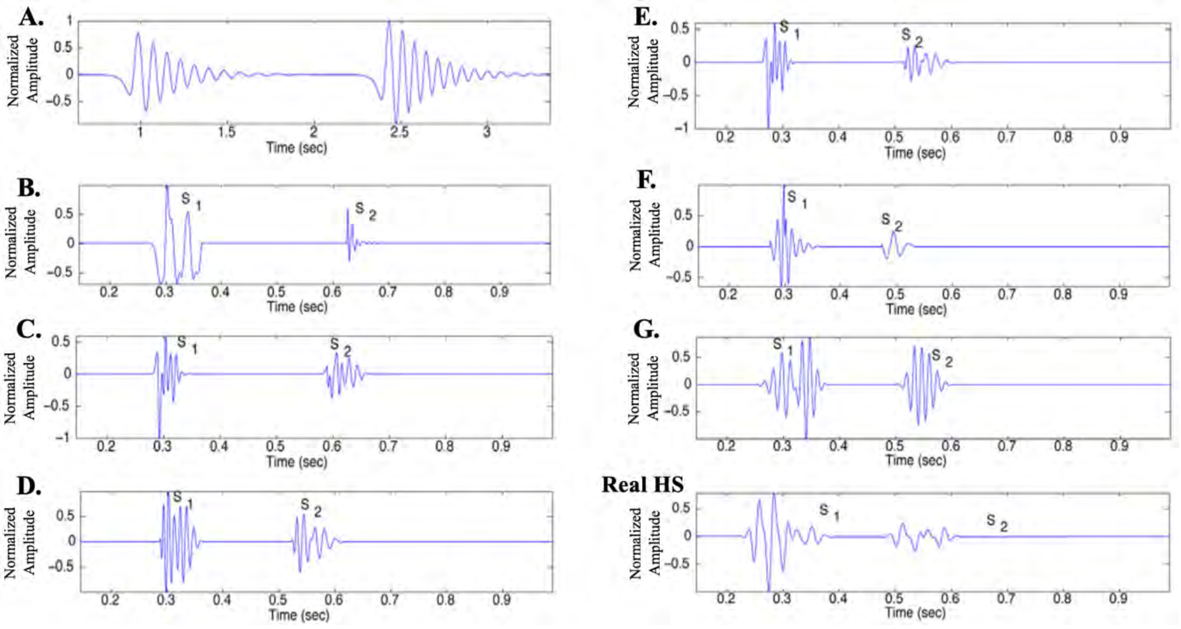


Figure 3. Synthetic heart sounds generated with state-of-the-art algorithms. Signals taken from [27]. Results obtained in: A) [21]; B) [23]; C) [25]; D) [27]; E) [22], F) [24]; G) [26]

Authors Chen, Duran and Lee performed a Time-Frequency analysis of the first heart sound (S1). This sound was modeled using two components, the first one represents the mitral valve and the second one refers to the muscular component of the myocardium **[19]**. The valvular component was modeled as a transient deterministic signal represented by two sinusoids that decay exponentially (see equation 1).

$$S_v(t) = \sum_{i=1}^{N} A_i e^{-k_i t} sin(2\pi f_i t + \phi_i) \quad (\mathbf{1})$$

Where N is the number of sinusoids, A_i is the amplitude, k_i is the damping factor, f_i is the frequency and ϕ_i is the phase of the i-th sinusoid.

The myocardial component is modeled with a frequency modulated wave, represented by:

$$S_m(t) = A_m(t) sin\left(2\pi\big(f_0 + f_m(t)\big)t + \phi_m(t)\right) \quad (\mathbf{2})$$

Where $A_m(t)$ is the amplitude of the modulated signal, f_0 is the carrier frequency, $f_m(t)$ is the frequency of the modulated wave, and ϕ_m is the phase of the function. In short, the heart sound S1 is represented by:

$$S1(t) = S_m(t) + \begin{cases} 00 \le t \le t_0 \\ S_v(t - t_0) t \ge t_0 \end{cases} \quad (\mathbf{3})$$

Where t_0 is the delay between the two components, since the mitral valve closes after muscle contraction.

In **[21]**, authors Xu, Durand, and Pibarot worked on the S2 heart sound model. They also used two components, represented by the aortic and pulmonary valves. In this work, two non-linear transient chirp components were proposed, as described in equation 4.

$$S2(t) = A_a(t) sin\big(\varphi_a(t)\big) + A_p(t - t_0) sin\left(\varphi_p(t)\right) \quad (\mathbf{4})$$

Where $\varphi_a(t)$ and $\varphi_p(t)$ are the instantaneous phases, in equation (5) they are represented as a non-linear function of time.

$$\varphi_a(t) = \sum_{m=0}^{M} a_m t^m \quad \text{and} \quad \varphi_p(t) = \sum_{m=0}^{M} P_m t^m \qquad \mathbf{(5)}$$

Where M is the order of the polynomial function a_m and P_m are the real coefficients.

Other authors have tried to model the heart sound from an exponentially damped sinusoidal function that describes the closure and vibrations of the heart valves [**18**]. The model is represented by the following expression:

$$S[n] = \sum_{m=0}^{M} a_m e^{-n\alpha_m} sin(2\pi f_m n + \varphi_m) \qquad paran = 0,1,\ldots,N-1. \quad (6)$$

Where M is the number of damped sinusoids, a_m is the amplitude, α_m is the damping factor, and φ_m is the phase.

All these attempts to propose an accurate model of PCG signals have a common drawback, they do not take into account specific characteristics such as heart rate variability (HRV), respiration rate and time division in S1 and/or S2 during inspiration and expiration. For its part, in [**25**] the development of a dynamic model of PCG (phonocardiographic) signals is presented considering these characteristics, in order to generate more realistic cardiac sounds. The authors hypothesize that the PCG signals are the action of the right side of the heart (T1 and P2) superimposed on the left side (M1, A2), represented by:

$$Z_{PCG}(t) = \sum_{S\epsilon[r,l]} Z_s(t) \qquad (\mathbf{7})$$

Where r and l correspond to the right and left sides respectively. The Z_s function is generated by three ordinary differential equations (8, 9, 10) derived from an ECG signal generation model [**23**], since these signals are correlated.

$$\dot{x} = \epsilon x - \omega y. \qquad (\mathbf{8})$$

$$\dot{y} = \epsilon y + \omega x \qquad (\mathbf{9})$$

$$\sum_{k\epsilon[P,Q,R,S,T]} a_{k_s} \Delta\theta_{k_s} sin\left(\phi\left(\Delta\theta_{k_s}\right)\right) e^{\left(\frac{-\Delta\theta_{k_s}{}^2}{2b_{k_s}{}^2}\right)} - (Z_s - Z_0) \qquad (\mathbf{10})$$

Where ω is the angular velocity of the trajectory, $\epsilon = 1 - \sqrt{x^2 - y^2}, z_0 = A_0 sin(2\pi f_0 t)$ where f_0 is the respiratory rate. a_{k_s} and b_{k_s} are parameters used for the amplitude.

$\Delta\theta_{k_s} \cong \left(\theta - \theta_{k_s}\right)[2\pi]$, where $\theta_{k_s} = \theta_k + \delta\theta_k \forall k \in \{P,Q,R,S,T\}$

θ_k are the fixed angles along the unit circle for distinct points (P, Q, R, S, T) in the z plane; $\theta = tan^{-1}(y,x) \in [-\pi,\pi]$; $\delta\theta_k$ is a positive deviation.

A modulated sine is found in this model, which represents the deflections to the PCG signal exhibiting a linear transient chirp morphology: $sin\left(\phi\left(\Delta\theta_{k_s}\right)\right)$

Where $\phi(\Delta\theta_{k_s}) = \Delta\theta_{k_s} f_{k_s}(\Delta\theta_{k_s})$, being f_{k_s} the frequency of the vibration.

An inconvenience presented in this model is that the parameters depend on the characteristics obtained from an ECG signal. Considering that the samples to be used will be real heart sounds and not ECG signals, the authors Almasi, Shamsollahi and Senhadji propose a mathematical model that is also based on an ODE system to generate dynamic heart sounds but does not depend on parameters of the ECG signal (P, Q, R, S, T) **[23]**. The equation of this model is described as follows:

$$\dot{z} = \sum_{i\in\{S1^- S1^+ S2^- S2^+\}} \left(\frac{a_i}{\sigma_i}(\theta-\mu_i)e^{\left(-\frac{(\theta-\mu_i)^2}{2\sigma_i^2}\right)}\right)\cos(2\pi f_i\theta-\varphi_i)+2\pi\alpha_i f_i e^{\left(-\frac{(\theta-\mu_i)^2}{2\sigma_i^2}\right)}\sin(2\pi f_i\theta-\varphi_i) \quad (\mathbf{11})$$

Where a_i, μ_i and σ_i are the amplitude, center and width parameters of the Gaussian terms, and f_i, φ_i are the frequency and phase shift of the sinusoidal terms, respectively. θ is the independent parameter in radians that varies in the range -π, π for each beat. The superscripts -/+ indicate the two Gabor nuclei used to model each heart sound.

Although advances have been made in the synthesis of the S1 and S2 sounds, the proposed synthesis models still present limitations due to the complexity and highly non-stationary nature of PCG signals. A basic time-frequency analysis of these synthetic signals shows that they are far from the real PCG signals. Moreover, the signals generated by these models do not have a physical meaning linked to the mechanisms related to the genesis of heart sound [21, 23 and 26].

Additionally, all these proposed methods present a very ideal behavior in the systolic and diastolic segments, since they only focus on the generation of the S1 and S2 sounds. Also, none of these models can generate any kind of murmur. Therefore, these synthetic signals are not viable for the training of machine learning models for normal versus abnormal HS classification.

2.3. Synthesis of Biomedical Signal using GAN: State of the art

2.3.1. Introduction to Generative Adversarial Network (GAN)

Generative adversarial networks (GANs) are architectures of deep neural networks widely used in the generation of synthetic images **[64]**. This architecture is composed of two neural networks that face each other, called the generator and discriminator **[65]**. In Figure 4, a general diagram of a GAN architecture is presented. The Generator (counterfeiter) needs to learn to create data in such a way that the Discriminator can no longer distinguish it as false. The competition between these two networks is what improves their knowledge, until the Generator manages to create realistic data.

As a result, the Discriminator must be able to correctly classify the data generated by the Generator as real or false. This means that their weights are updated to minimize the probability that any false data will be classified as belonging to the actual data set. On the other hand, the Generator is trained to trick the Discriminator by generating data as realistic as possible, which means that the weights of the Generator are optimized to maximize the probability that the false data it generates will be classified as real by the Discriminator **[65]**.

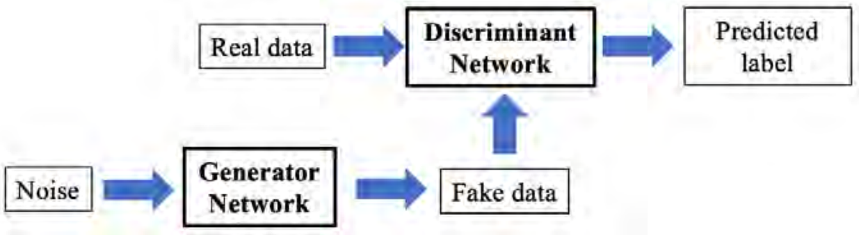


Figure 4. General diagram of a GAN.

The Generator is an inverse convolutional network, that is, it takes a random noise vector and converts it into an image, unlike the Discriminator who takes an image and samples it to produce a probability. In other words, the Discriminator (D) and Generator (G) play the following two-player minimax game with value function L(G, D), as described in equation (12).

$$\min_{G}\max_{D} L(D,G) = E_{x\sim\rho_{data}(x)}[log D(x)] + E_{z\sim\rho_{data}(z)} \quad \mathbf{(12)}$$

Where D(x) represents the probability that x estimated by the discriminator, z represents the input of random variables from the generator, $\rho_{data}(x)$ and $\rho_{data}(z)$ denote the data distribution and the distribution of samples from the generator respectively.

After several training steps, if the Generator and the Discriminator have sufficient capacity (if the networks can approach the objective functions), they will reach a point where both can no longer improve. At this point, the Generator generates realistic synthetic data, and the Discriminator cannot differentiate between the two input types.

GAN training is a very complex challenge that is currently an active research area to understand and improve performance. There are several very common failure modes such as: Mode Collapse, Lack of Convergence, and Leakage Gradient, which have become a motivation for researchers to design and implement new architectures, cost functions, and hyperparameter selection **[66 – 67]**. However, none of these problems have been fully resolved. Next, we mention each of them:

Mode Collapse: The main goal of a GAN is to generate a wide variety of signals in its output, for example, if it was trained on images of people's faces, the model's output is expected to produce a different face for each random input to its image generator. Thus, a mode collapse refers to a generator model that is only capable of producing one or a small subset of different outputs or modes, in this case the discriminator will always try to learn to reject that output, but if the discriminator gets stuck at a local minimum during training and doesn't find a way to reject that output, then it's too easy for the next iteration of the generator to find the most plausible output for the current discriminator. In this way, the generator overfits each iteration for a particular discriminator.

In summary, a mode collapse can be identified in two ways: i) the generator produces similar or very low diversity outputs regardless of the different points in the latent space that were used in the model input; ii) An oscillation of the generator and discriminator losses is expected with time [68-69].

Failure to Converge: It is one of the most common problems when training a GAN. Generally, a neural network fails to converge when the model loss is not stabilized during the training process. In the case of a GAN, lack of convergence refers to not finding a balance between generator and discriminator. Here are the different ways to identify this type of fault:

- When the discriminator loss has reached zero or close to zero. In some cases, the generator loss increases as the training time progresses.
- When the generator produces very low-quality signals that the discriminator easily identifies as false.

This type of failure can occur at the beginning of the run and continue throughout the training, it can also occur over several batch updates, or even over several epochs, and then recover.
Various forms of regularization have been tried to improve this type of failure, such as: Add noise to the input of the discriminator or penalize the weights of the discriminators **[70 - 71].**

Vanishing Gradients: This type of failure occurs when the discriminator is too good so the generator gradient fades away and does not learn anything. Therefore, an optimal discriminator does not provide enough information for the generator to move forward. Faced with this situation, the researchers try to modify the cost function of the GAN or the hyperparameters of each model in order to avoid overfitting and imbalance between the generator and the discriminator **[72, 73]**.

2.3.2. A review of GAN applied in biomedical signals

Generating training samples for supervised tasks is a long-sought goal in Artificial Intelligence. For the case of physiological signals, the challenge is even greater, due to the dynamic nature in which various parts of the system interact in complex ways. Although studies have already been carried out to understand these dynamics from mathematical processes, it is still not possible to obtain many diverse examples that contribute to the training of classification models, without falling into overfitting. Therefore, as mentioned in the introduction to this document, considering that in recent years there have been great advances in audio synthesis using Deep Learning techniques such as Generative Adversarial Networks (GANs), the motivation of many researchers to use these models for the generation of biomedical signals has increased, with the main objective of increasing the training data set to improve the performance of the classification models. Table 6 shows several proposed works of GANs for the synthesis of ECG, EEG, EMG and PPG signals.

Table 6. A review of GAN applied in biomedical signals

Ref	Year	Synthetic Signal	Proposed Method	Dataset
[74]	2021	ECG	GAN	PhysioNet Database [96]. (549 samples).
[75]	2020		Attention-based Generators and CycleGAN	BIDMC [97] (53 recordings of 8 min each), CAPNO [98] (42 recordings of 8 min each), DALIA [99] (15 recordings of 2 hours each), and WESAD [100] (15 recordings of 1 hour each).
[76]	2020		GAN	PTB-XL database [101]. (21k samples, 10s length each, sampled at 500 Hz)
[77]	2020		WaveNet; SpectroGAN; WaveletGAN	MIT-BIH Arrhythmia Database [102]. (10000 samples)
[78]	2020		Multivariate GAN	
[79]	2020		SimGAN-ECG	
[80]	2019		GAN (LSTM -CNN)	
[81]	2019		RPSeqGAN	
[82]	2019		PGAN	
[83]	2019		GAN (BiLSTM -CNN)	
[84]	2019		GAN	
[85]	2019	ECG and EEG	RGAN	Three real-world biosignal datasets from The UEA & UCR Time Series Classification Repository [103] (1362 samples ECG and 11500 samples EEG)
[86]	2021	EEG	CVAE-GAN	Private dataset: 12 samples. Public dataset: Competition IV Data sets 2a [104]: 36 samples.
[87]	2020		RNN; GAN; MFCC	Dataset used by authors in [105] (1120 samples)
[88]	2020		WGAN	1) Action Observation dataset [106]: 346 samples 2) Grasp and Lift dataset [107]: 576 samples 3) Motor Imagery dataset [108]: 1560 samples
[89]	2019		DCGAN; WGAN; VAE	Private dataset: 30 samples per class
[90]	2019		GAN	A public dataset collected by authors of [109]: 600 samples.
[91]	2018		GAN	A private dataset: 438 samples
[92]	2020	PPG	CGAN	1) PPG from Patients in Vietnam with Hand-Foot-Mouth Disease: 1980 samples 2) PPG from Patients in Vietnam with Tetanus: 5978 samples 3) PPG from Patients in China with Cardiovascular Disease [110]: 219 samples 4) PPG from Physionet 2015 Challenge [111]: 2202 samples
[93]	2020	EMG	SinGAN	A private dataset: 240 recording, each of which is of 3 s

[94]	2020		DCGAN	1) NinaPro is an open EMG database [112] (240 recording, each of which is of 8 s) 2) Parkinson’s Disease EMG Dataset [113] (90 recording, each of which is of 60 s)
[95]	2020	ECG; EMG; EEG; PPG	GAN (BiLSTM - CNN)	ECG data: MIT-BIH Arrhythmia Database [102] (1000 Samples). EEG data: Siena Scalp EEG Database [114,115] (recording of 128 h). EMG data: Sleep-EDF Database [116] (recordings of 197 full-nights). PPG data: BIDMC PPG and respiration Dataset [117] (53 recording, each of which is of 8 min duration).

The ECG signal is one of the most used in the implementation of GAN models, perhaps one of the main reasons is the availability of large databases in the web [74-84]. The MIT-BIH database is one of the most used in the case of ECG signals, it contains many cardiac signals (about 10,000 samples) with different types of abnormalities and has been widely used in training different models of ECG signals. machine learning classification. Another widely used database is Physionet, with more than 500 normal and abnormal cardiac signals. In the case of EEG signals, many of the repositories are private, generally there are few samples with a long time duration. The dataset described in [103] contains more than 10,000 EEG samples that could be very useful in different types of applications. EMG and PPG signal databases have also been explored for GAN model training as described in [92-95]. The interest of researchers in working on the synthesis of biomedical signals with GAN is increasing in recent years, without a doubt reflecting the low availability of this type of signals on current datasets.

Regarding the generator and discriminator models, different types of neural networks, cost functions and a hyperparameter selection analysis have been proposed in order to combat the common issues of convergence and mode collapse. In most of the proposed methods, convolutional neural networks (CNN) are used in both the generator and the discriminator. However, different types of networks have been explored in the generator, such as LSTM, Bi-LSTM and even Recurrent Networks [80], [83], [87], [95]. Techniques such as Wavelet and MFCC calculations have also been used to improve the performance of the GAN on ECG and EEG signals, respectively [77], [87]. The DCGAN architecture is one of the most used in the generation of images, and it was also implemented for this type of signals [89], [94].

Without a doubt, there are many GAN architectures published in the state of the art related to biomedical signals. However, although many training signals are used in several experiments, the result is not as expected since there are still limitations of mode collapse and convergence [84-85]. Additionally, the training of these models requires a high computational cost and training time. In other works, the synthetic signals produced by the GANs are used as training samples for a classification model, the results show an increase in accuracy [80], [91]. However, it is very likely that the classification model is overfitted by being trained with many similar signals. In summary, the advancement of GANs in biomedical signals is very promising, but many aspects still need to be addressed to reduce the main limitations that prevent achieving the goal.

2.4. CONTRIBUTION TO THE STATE OF ART:

The development of a system to generate normal and abnormal synthetic heart sounds in a reliable way and validated by health specialists would advance the state of the art in HS processing and analysis in the following aspects:

i. Train ML models to discriminate types of cardiac abnormalities. See figure 5.
ii. Test new features extraction techniques that facilitate classification.
iii. Validate the performance of existing Deep Learning based audio synthesis methods used in cardiac signal.
iv. Improving the training of medical students, since they will not need to look for patients with different anomalies to perform auscultations.

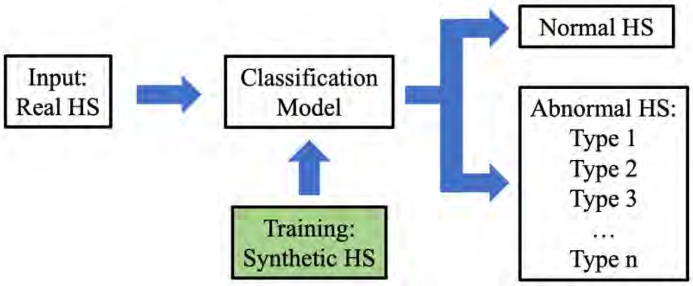


Figure 5. General diagram of a HS classification model using synthetic signals for training

In this way, we could move forward in the construction of a large-scale, computer-assisted intelligent tool for the diagnosis of cardiac conditions. In addition, this work is expected to be a great contribution to the generation of other types of biological signals, such as: Lung sounds, digestive sounds and even electrophysiological signals.

3. Research hypothesis

The following hypothesis is stated:

"It is possible to implement a Machine Learning algorithm to generate synthetic heart sounds that can serve to improve performance in heart sound classification models."

4. Analysis of Heart sounds

In this chapter we present the advances obtained during the investigation in the analysis of heart sounds. A block diagram of the main stages of an automatic heart sound classification system is shown in Figure 6, as mentioned at the beginning of the previous section. We have explored different signal processing techniques and machine learning models to obtain a good performance on each of these stages. The results have been compared favorably with the state of the art. Each of these developments are described below.

This section begins with the description of an algorithm capable of segmenting the phonocardiography recordings (PCG), in order to identify the sounds S1, S2 and the systolic and diastolic intervals. This algorithm presents a better performance compared to the methods proposed in the state of the art. Subsequently, different signal processing techniques are described to be used in the feature extraction stage, then, these characteristics are used as input to several Machine Learning models.

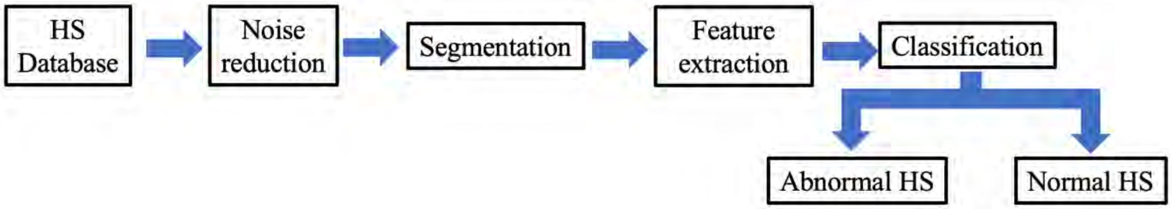


Figure 6. General diagram of the automatic heart sound classification system

4.1. Automatic segmentation of heart sounds

This is the initial stage of the classification system and consists of identifying the S1 and S2 sounds of a cardiac cycle. To achieve this, a pre-processing stage is required to help attenuate unwanted noise. Subsequently, signal processing techniques are used to identify the corresponding peaks of the S1 and S2 sounds. Below we present our contribution to the state of the art with a robust algorithm that allows to automatically segment normal and abnormal heart sounds, identifying systolic and diastolic intervals, even in the presence of high amplitude murmurs.

4.1.1. Proposed method

In this work, the algorithm proposed in [118], called empirical wavelet transform (EWT), is studied as a reference in the pre-processing stage of the cardiac signal. The EWT allows the extraction of different components of a signal when designing an appropriate filter bank. The theory of the EWT algorithm is described in detail in [118]. This method has given better results than empirical mode decomposition (EMD), since the latter decomposes the signal in many ways that are difficult to interpret or do not give relevant information [118].

The EWT algorithm has been used as a signal filtering stage in different applications, such as in the processing of voice and movement signals for the classification of Parkinson’s disease severity [119], detection of mechanical failures from vibration signals [120] and in the analysis of geomagnetic signals for the detection of seismic activity [**121**]. However, this technique demands a high computational cost, since the model can decompose the signal into many components according to the criteria set and then chooses the optimal component, and is therefore a non-viable method for a real-time system. Therefore, many researchers have sought ways to apply improvements to the method in order to obtain the desired components in a more effective way, as described in [118–121]. In our case, we need to improve the EWT algorithm to extract the S1 and S2 components of the cardiac signal, while attenuating the murmurs or external noises that are present in the original signal. To achieve this, we consider that the frequency range of the S1 and S2 sounds is between 20–200 Hz [122]; therefore, we decided to modify the edge selection method that determines the number of components. Figure 7 shows the block diagram for the proposed segmentation system. Initially, all the signals taken from the databases are decimated to a sampling frequency of 4 kHz, and amplitude-normalized using the equation (13), where X(n) and N denote the original signal and length in samples of the signal.

$$X_{Norm}(n) = \frac{X(n)}{Max_{n=1}^{N}[X(n)]} \qquad (13)$$

The first stage of the system decomposes the signal into different frequency bands using the modified EWT (mEWT) method. To achieve this, the first step is to identify the maximum value of the Fast Fourier Transform (FFT) of the signal in the range of 20 Hz to 150 Hz. After identifying which frequency belongs to the maximum value, this is taken as the center frequency for a filter with a bandwidth 40 Hz; that is, if the center frequency is 70 Hz, a bandwidth between 50-90 Hz is established using the EWT method. It is worth mentioning that tests were carried out with different ranges of bandwidth and it was found that sounds S1 and S2 have a more pronounced amplitude with a bandwidth of 40 Hz. Thus, we reduce the computational cost by not computing many filters for many different frequency bands and efficiently eliminate unwanted noise.

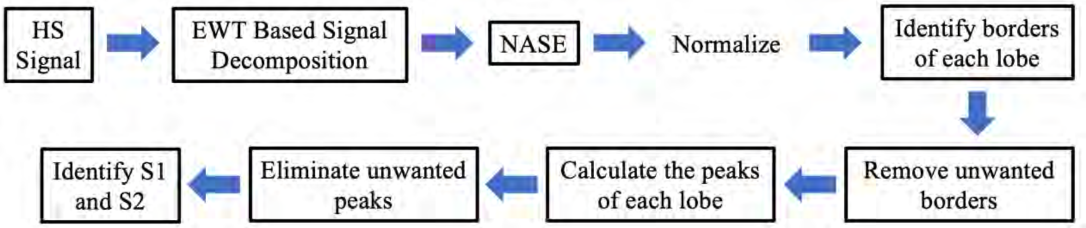

HS Signal
EWT Based Signal Decomposition
NASE
Normalize
Identify borders of each lobe
Remove unwanted borders
Calculate the peaks of each lobe
Eliminate unwanted peaks
Identify S1 and S2

Figure 7. Block diagram for the proposed segmentation system.

In Figure 8, a heart sound and its respective FFT are shown. In this example, the maximum amplitude is approximately 60 Hz; therefore, the chosen frequency band is 40–80 Hz, as seen in the green segment (see Figure 8b). Figure 8 also shows four components in order to illustrate the shape of the signal on different frequency ranges, using the adaptive filters of the EWT method. For this example, the low-frequency segment is defined between 1–40 Hz (blue segment, Figure 8c); in the segment from 80 Hz to 350 Hz some kind of murmur is expected (red segment, Figure 8e); and in the segment of 350 Hz and above it is expected that high-frequency noises that intervene in the recording can be observed (cyan segment, Figure 8f). It can be seen that in the segment of 40–80 Hz (green segment, Figure 8d), the S1 and S2 sounds have a considerable amplitude that can facilitate their identification.

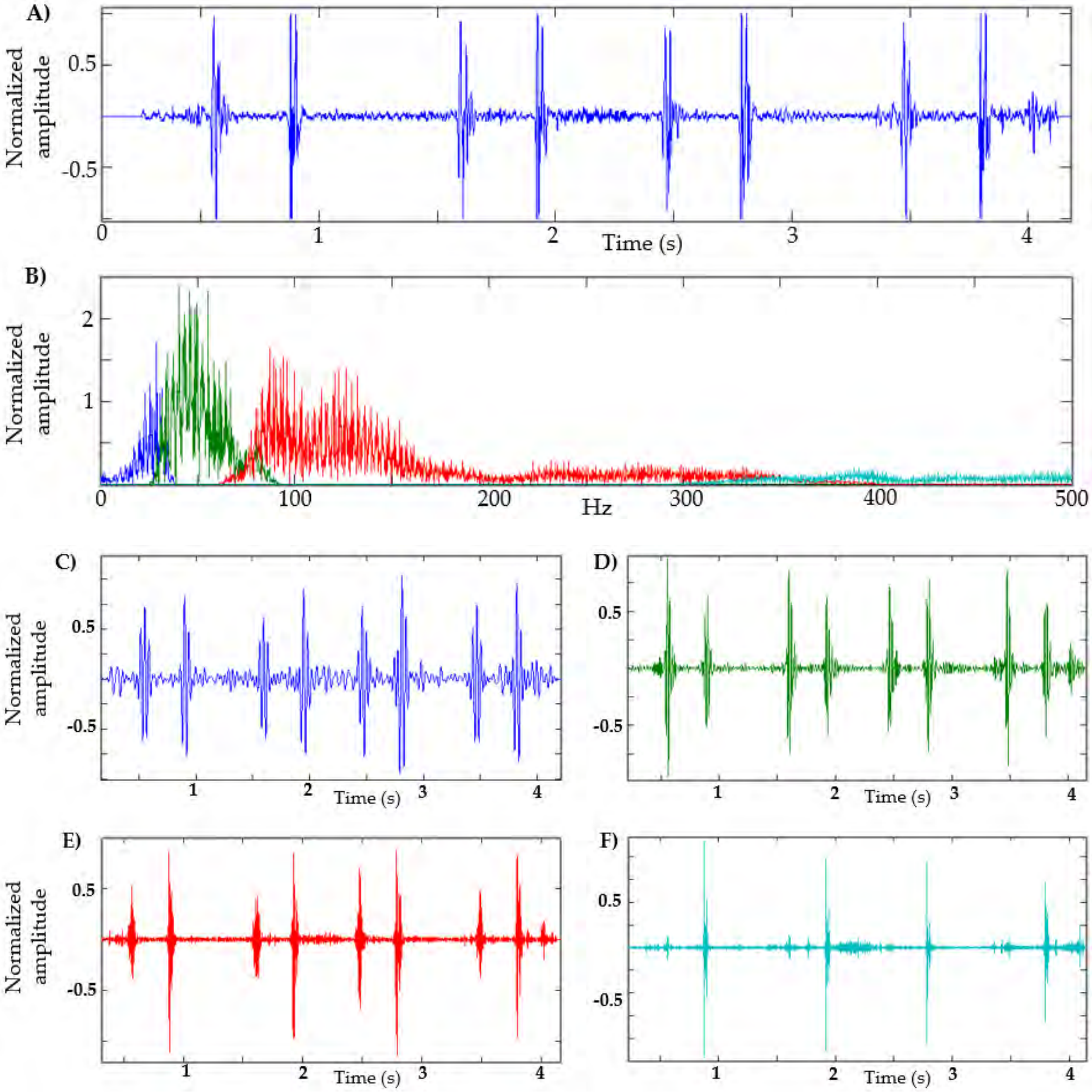


Figure 8. Decomposition of heart sound using EWT. (a) Heart sound recording; (b) FFT of heart sound with the frequency bands determined; (c) EWT components defined between 1–40 Hz. (d) EWT components defined between 40–80 Hz. (e) EWT components defined between 80–350 Hz. (f) EWT components defined at 350 Hz and above.

A main stage in automatic segmentation is the detection of the envelope of the S1 and S2 segments. Various techniques have been attempted in the literature, such as the absolute value of the signal, quadratic energy, Shannon entropy, Shannon energy and Hilbert transform, among others [123]. In [123], the authors make a comparison of different methods used to calculate the envelope of the PCG signal, with the Shannon energy being the most effective for identifying the S1 and S2 peaks. In [33], [34] and [35], this method is used, obtaining good results in the identification of S1 and S2 segments. The Shannon energy equation is defined as

$$E = -x(i)^2 log(x(i)^2) \quad \textbf{(14)}$$

where x(i) represents the samples of the signal and E is Shannon's energy.

When calculating the quadratic value of a sample, large oscillations can be generated; the amplitude of the sample increases when the absolute value of the original amplitude is greater than 1 and decreases otherwise. Therefore, it is convenient to use normalized energy, as in the case of the normalized average Shannon energy (NASE) [124], defined as follows:

$$En = \frac{E - \mu}{\sigma} \quad \textbf{(15)}$$

where μ is the average value of energy E of the signal, σ is the standard deviation of energy E of the signal and En is the normalized average Shannon Energy.

After calculating the NASE in the HS signal decomposed by the EWT—that is, the signal that contains the S1 and S2 sounds—the negative values are equaled to zero and the signal is normalized. In Figure 9b, the result of an NASE signal of a heart sound recording can be observed.
Analyzing the NASE signal, we will proceed to identify the edges of each of its lobes; this helps to determine the beginning and the end of the S1 or S2 sounds. The limits obtained from each lobe are shown in Figure 9c. A common problem occurs when there are lobes that are very close to each other; in this case, the lobe that has less energy is eliminated. This process is repeated three times. Lobes of short duration and low amplitude are also eliminated.

Subsequently, the peaks in each lobe are calculated as shown in Figure 9d. Unwanted peaks are removed using the following steps:

1. Calculate the average of the intervals between peak (i) and peak (i + 1).
2. Eliminate the peaks that belong to an interval less than 0.25 * on average.
3. Eliminate the peaks that belong to an interval less than 0.3 * on average.
4. Eliminate the peaks that belong to an interval less than 0.4 * on average.
5. Eliminate the peaks that belong to an interval less than 0.55 * on average.

Taking into account that the systolic and diastolic intervals do not vary drastically in a recording, these thresholds (0.25, 0.3, 0.4 and 0.55) were established empirically to eliminate those intervals with a very short duration that were detected due to unwanted peaks. The threshold is increased step by step (ascending) to avoid removing a peak that could actually be an S1 or S2 sound.

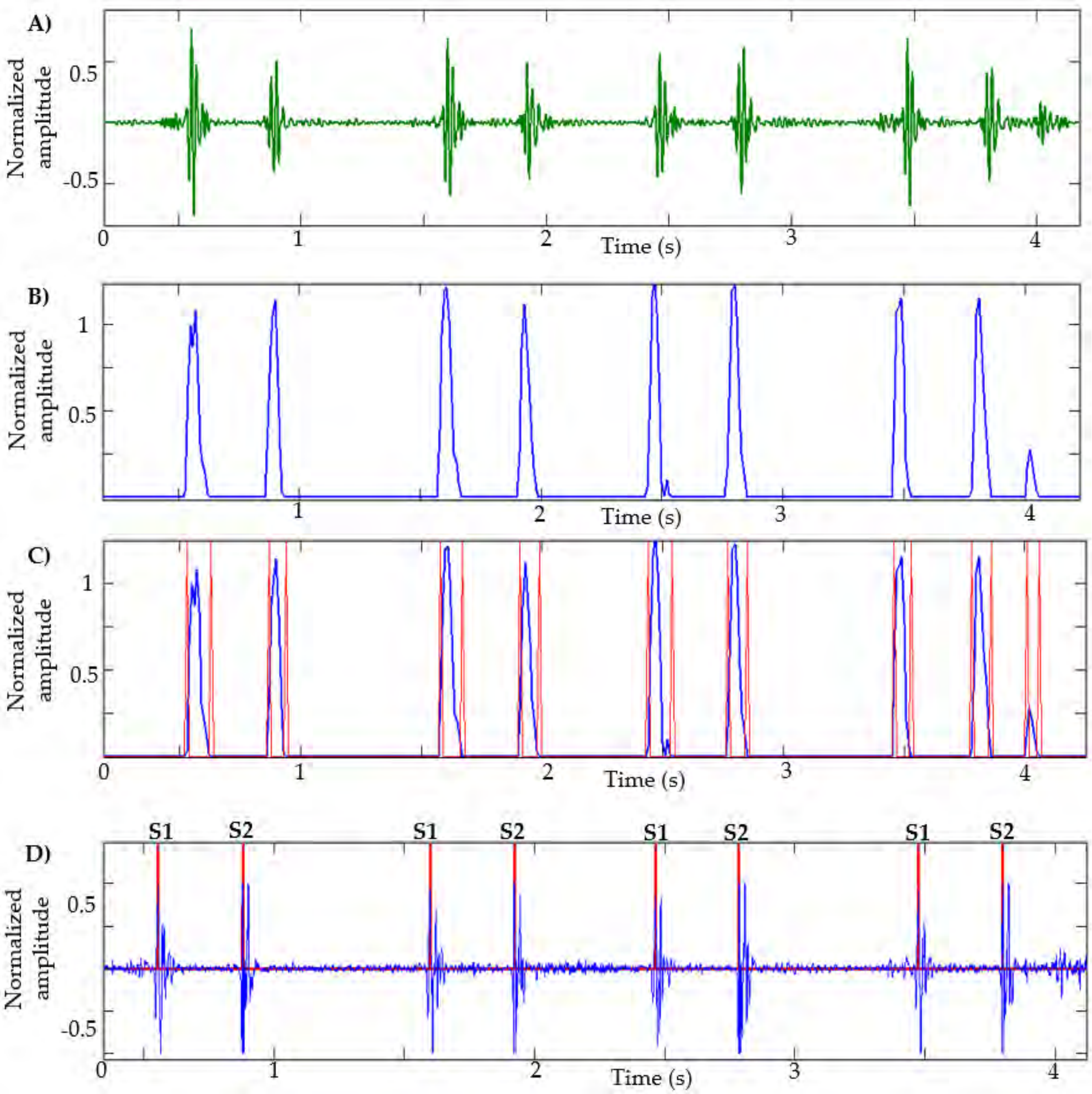


Figure 9. Stages of segmentation: a) EWT component: heart sound; b) NASE of the signal; c) identification of borders of each lobe; d) identification of S1 and S2 sounds.

There are cases in which the sound of S1 or S2 has a low amplitude and short duration, as seen in Figure 10a. Generally, the methods presented in the state-of-the-art approaches fail in this case. Taking into account the fact that the duration of the cardiac cycle is approximately 0.8 s and the diastolic interval is approximately 0.6 s [125], a condition is established in which the interval between each peak is evaluated, and if in most cases the interval is greater than 650 milliseconds, it can be said that there is a sound (S1 or S2) in that interval. Then, each interval is normalized as shown in Figure 10c. Subsequently, all the previous stages are performed to find S1 or S2.

To identify the S1 and S2 sounds in a recording, it is enough to know that the systolic interval is located between S1 and S2, while the diastolic is located between S2 and S1 of the next cycle. In addition, it is known that the systolic interval is always shorter than the diastolic [125]. With these criteria, it is easy to use the resulting time marks to discriminate between S1and S2 sounds, systole and diastole.

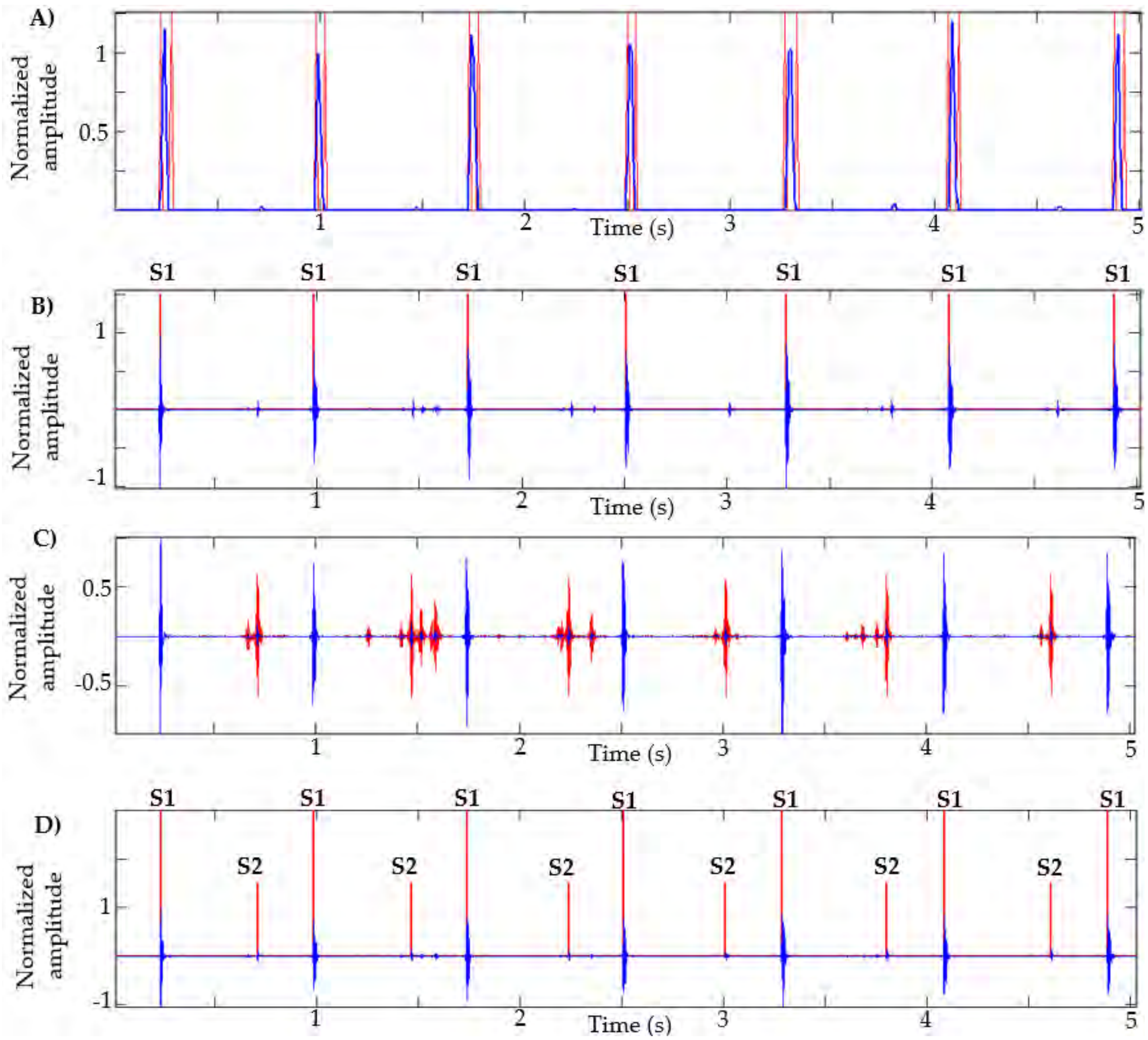


Figure 10. Segmentation of heart sounds with S1 of low amplitude and short duration. a) NASE of heart sound; b) Identification of peaks; c) Heart sound with normalized intervals; d) Identification of S1 sounds.

4.1.2. Results and discussion on the proposed automatic segmentation

The heart sound database can be downloaded from the Pascal Classifying Heart Sounds Challenge website [17]. Data have been gathered from two sources: A) samples acquired through a smartphone using the iStethoscope app and B) samples acquired in a hospital setting using a DigiScope electronic stethoscope. The database contains the following categories: normal, murmur, extra heart sound and artifact. Table 7 presents in detail the number of recordings and cardiac cycles for the two datasets.

Table 7. Database of heart sounds: Pascal Challenge. HS: heart sound.

Dataset A	Number of recordings	Dataset B	Number of recordings
Normal	31	Normal	200
Murmur	34	Normal noisy	120
Extra HS	19	Murmur	95
Artifact	40	Extra systole	46

In [17], the results of manual segmentation carried out by experts are presented that serve to evaluate the performance of the segmentation algorithms. The objective of this challenge is to calculate the number of cardiac cycles and identify the S1 and S2 sounds of a recording. Table 8 and Table 9 show the results obtained using dataset A and B respectively, where the evaluation metric is the error that exists between manual segmentation labels provided by the database and those obtained by the proposed method; that is, the difference between the samples corresponding to the S1 and S2 sounds with those detected by the segmentation algorithm. In [17], a spreadsheet is presented to evaluate the error of each sound.

Table 8. Results of segmentation for dataset A. (HB: heartbeat).

HS (file name)	HB	Average Error (Samples)	HS (file name)	HB	Average Error (Samples)
201101070538	11.5	15,792.91	201103101140	9	58,920.83
201101151127	10	177,625.15	201103140135	9.5	24,891.94
201102081152	9.5	159,024.94	201103170121	10	343.15
201102201230	11.5	17,384.91	201104122156	11.5	173,664.78
201102270940	8.5	159,194.17	201106151236	9.5	56,598.00

Table 9. Results of segmentation for dataset B.

HS (file name)	HB	Average Error (Samples)	HS (file name)	HB	Average Error (Samples)
103_1305031931979_B	12.5	35.04	147_1306523973811_A	4	226.5
103_1305031931979_D2	10	33.05	148_1306768801551_D2	8	31.75
106_1306776721273_B1	4	14.62	151_1306779785624_D	4.5	2543.77
106_1306776721273_C2	3	12.16	154_1306935608852_B1	4.5	2096.66
106_1306776721273_D1	3.5	109.00	159_1307018640315_B1	6	19.33
106_1306776721273_D2	7.5	1613.4	159_1307018640315_B2	3	28.66
107_1305654946865_C1	7.5	1524.6	167_1307111318050_A	13	58.96
126_1306777102824_B	8.5	2260.82	167_1307111318050_C	3	26.5
126_1306777102824_C	5.5	42.72	172_1307971284351_B1	3.5	12.14
133_1306759619127_A	4	32.75	175_1307987962616_B1	2.5	10.00
134_1306428161797_C2	2.5	4.6	175_1307987962616_D	7	36.71
137_1306764999211_C	15	1615.5	179_1307990076841_B	16.5	51.45
140_1306519735121_B	11	45.86	181_1308052613891_D	3	19.5
146_1306778707532_B	18	2115.97	184_1308073010307_D	26.5	57.60
146_1306778707532_D3	3	8.33	190_1308076920011_D	3.5	2386.14

Table 10, the results of the total errors shown in [39], [40], [42], [43], [45] and the proposed method for datasets A and B are shown. Taking into account that the results presented in [39], [42] and [45] were the best for the Pascal Challenge. Our method obtained a total error of 843,440.8 for dataset A and 17,074.1 for dataset B. These results are the best compared to the state-of-the-art approaches.

Table 10. Results general of segmentation.

Method	Dataset A	Dataset B
[42]	4,219,736.5	72,242.8
[39]	3,394,378.8	75,569.8
[45]	1,243,640.7	76,444.4
[43]	873,577.9	29,269.6
[40]		47,804.4
Proposed	843,440.8	17,074.1

This algorithm was also tested with recordings obtained from Physionet, specifically those signals in which the method in [37] failed, as described in [38]. Figure 7 presents the result of the segmentation in recordings a031, a0112, a0284 and a0352 using the proposed method. In [38] (Figure 4, page 14 of that article), the result of segmentation with these same signals is presented using the method in [37] based on logistic regression and HSMM, with this being one of the most used algorithms in the literature. As shown in Figure 11, the proposed algorithm correctly detects the S1 and S2 sounds in each of the recordings.

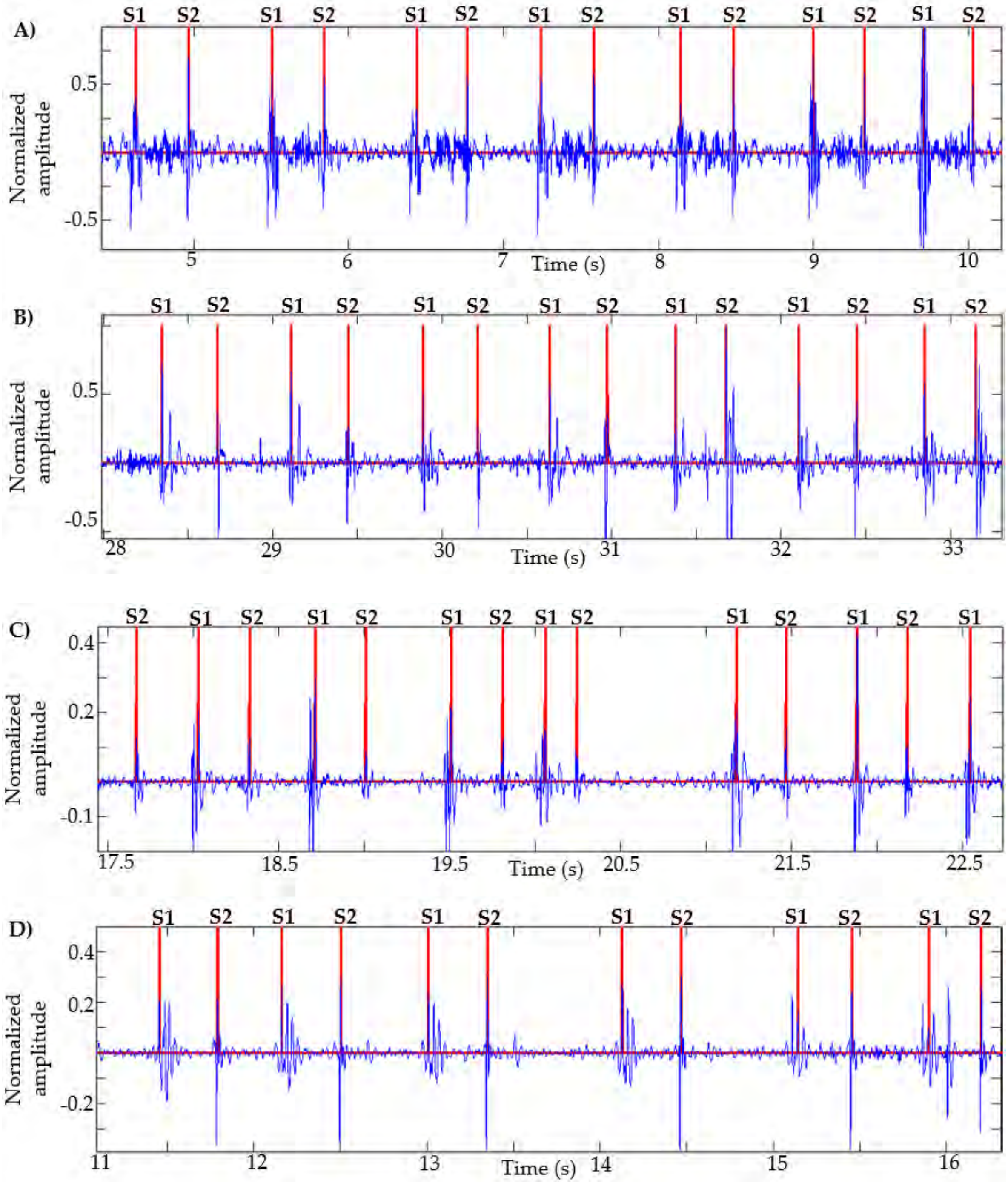


Figure 11. Examples of segmentation using the proposed method. Heart sound signals for recordings (***A***) a031, (**B**) a0112, (**C**) a0284 and (**D**) a0352 are taken from the PhysioNet/CinC Challenge 2016, together with the successful segmentations using the proposed algorithm.

4.2. Automatic classification of normal and abnormal heart sounds

This section presents different proposed approaches to classifying normal and abnormal heart sounds. The first experiment is mainly based on feature extraction using Linear Prediction Coefficients (LPC) and Mel Frequency Cepstral Coefficients (MFCC). Subsequently, a second approach is described applying a pre-processing to the signal using the EWT algorithm, then power values in the systolic and diastolic intervals are calculated. In both approaches, different Machine Learning models were used in order to evaluate the accuracy performance provided by each combination of characteristics. These experiments were compared favorably with the state of the art and published in scientific articles [49, 50].

4.2.1. Proposed method using LPC and MFCC as acoustic features

This first method proposes an analysis and classification procedure for discriminating between normal and abnormal cardiac sounds, based on Linear Prediction Coefficients (LPCs) and Mel-Frequency Cepstral Coefficients (MFCCs) as features for three different classifiers: Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Random Forest. Despite being widely used in speech and audio processing, LPCs and MFCCs have so far not being well studied as possible feature for acoustic analysis of auscultation sounds.

4.2.1.1. Data

A total of 805 heart sounds (415 normal and 390 abnormal) from six databases were selected:

- Samples taken from PhysioNet/Computing in Cardiology Challenge 2016 [16].
- Pascal challenge database [17].
- Database of the University of Michigan [13].
- Database of the University of Washington [14].
- Thinklabs database (digital stethoscope) [15].
- 3M database (digital stethoscope) [145].

4.2.1.2. Feature extraction

LPCs and MFCCs are widely used in audio processing, especially for speech signals [126]. Given that the signal obtained through auscultation is acoustic, and humans typically use their ears to analyze it [127], it seems plausible that features used successfully to model other acoustic signals should also be able to model auscultation sounds.

A. Linear prediction coefficients (LPCs):

LPC is a model for auto-regressive random variables that captures the spectral envelope of the signal of interest as the frequency response of an IIR, all-pole, filter, as shown in equation (16).

$$H(z) = \frac{G}{1 + a_1 Z^{-1} + a_2 Z^{-2} + \cdots + a_n Z^{-n}} \qquad (\mathbf{16})$$

The LPC model is based on the assumption that current signal samples can be generated by a linear combination of p previous samples. The model can be described by equation (17):

$$S(n) = \sum_{k=1}^{p} s(n-k) a_k \qquad (\mathbf{17})$$

Where p is the model order, a_k are the LPCs and s(n) is the reconstructed signal.

LPCs can be used for the processing of biological signals, such as heart sounds. These signals are characterized by a large variation in time and frequency domains, and are also classified as non-stationary signals [34]. In the case of the heart sounds, it seems natural to divide the signal based on the characteristic segments of the cardiac cycle (S1, S2, systole, diastole). The idea is then to use LPCs to capture the spectral characteristics of each differentiated segment of a heart cycle, and then use such spectral characteristics as the basis for a classifier.

B. Mel-Frequency Cepstral coefficients (MFCCs)

While LPCs model sounds from the point of view of its production, MFCCs attempt to model the perception of such sounds by the human auditory system. This is achieved using a filter bank with a nonlinear frequency scale, called Mel-scale [128]. Since auscultation sounds are interpreted by health professionals using their ears, it can be argued that MFCCs provide a feature representation of cardiac sounds close to that of a human, therefore suitable for automatic analysis and interpretation.

The computation of MFCCs is done by the following steps:

i) Divide the signal into frames, using a suitable window function. The hamming window is commonly used in speech signal spectral analysis due to its spectral characteristics.

ii) Estimate the Power Spectral Density (PSD) for each frame using the Discrete Fourier Transform (DFT). The PSD is then passed through the Mel filter bank, obtaining a vector of coefficients.

iii) Finally, the vector is represented in decibels and a Discrete Cosine Transform (DCT) is applied on it.

In this first experiment of heart sound classification, it was proposed to calculate the LPCs and MFCCs in each segment of the cardiac cycle, that is, for each cardiac cycle segment (S1, systole, S2 and diastole) there were computed two feature vectors: the first with 6 LPC and the second with 14 MFCC, Figure 12 shows the distribution of MFCC + LPC features in a normal and abnormal cardiac cycle, respectively. These features were computed using the Voicebox Matlab toolbox [129]. It was decided to calculate 6 LPCs considering that the sampling frequency of the signals is 2 KHz. In the case of the MFCCs, it was decided to establish 14 coefficients due to the good results shown in the state of the art.

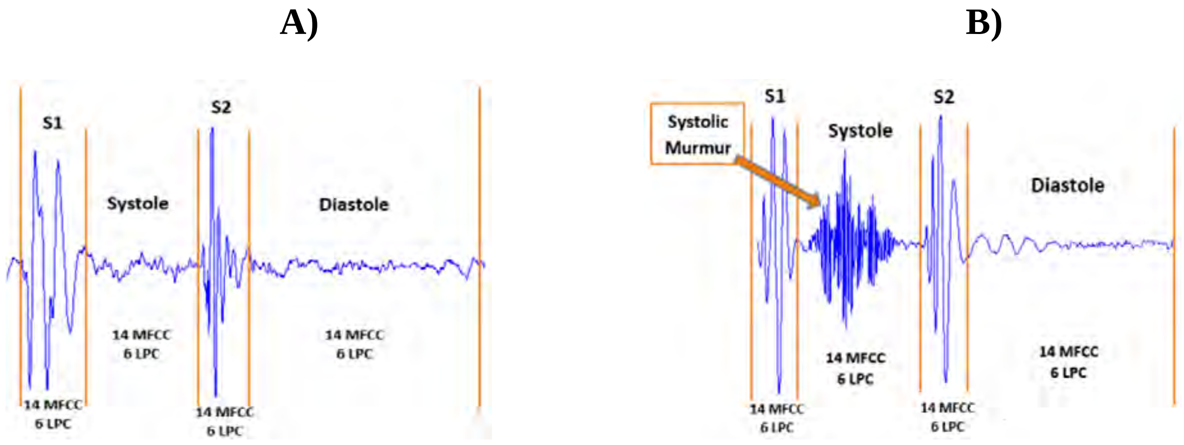


Figure 12: LPC+ MFCC feature distribution. A) for a normal HS; B) for an abnormal HS

4.2.1.3. Classification

At this stage we used the well-known, machine learning and data mining tool Weka (Knowledge Environment of Waikato University) [130] for building classification models. The LPC and MFCC features were organized into three different sets of characteristic vectors for classification: LPC-only characteristics, MFCC-only characteristics and LPC+MFCC characteristics. Each vector of characteristics (LPC, MFCC and LPC+MFCC) was used as input for four different classifiers: Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest and Multilayer Perceptron (MLP); 10-fold cross-validation was used to evaluate the performance of each classifier.

The literature shows plenty of similar work performed using SVM, KNN, MLP and Random Forest [131], [132], since the main objective of our work is to test the feature vectors, we decided to use those four common classification algorithms without any emphasis on algorithm optimization, the following paragraphs give a short introduction to each classifier.

A. Support Vector Machine (SVM)

This classification method, developed by Vladimir Vapnik and his team [133], is one of the most used in pattern recognition and binary classification problems, like the discriminating between normal and abnormal heart sounds. This model is based on constructing a hyperplane that allows to separate the data in two classes [134].

B. K-Nearest Neighbors (KNN)

A KNN classifier computes the distance between the data that we want to classify and labeled training data, selecting the nearest neighbors. Based on the neighbor's category, it is determined to which class the test data belongs [135]. Euclidean distance is commonly used to determine nearby neighbors for the test data. This model is also suitable for binary classification.

C. Random Forest

Random Forest is a combination of tree structured classifiers, where each tree depends on the values of a random vector. To classify an input vector, each tree that belongs to the forest gives a "vote" for classification. Subsequently, the model determines the classification according to the highest number of votes. Random forest are very efficient when used for binary classification on large number of samples, guaranteeing good specificity and sensitivity in the results. However, it is confusing in its interpretation compared to decision trees [136].

D. Multilayer Perceptron (MLP): The multilayer perceptron (MLP) is a direct-feed artificial neural network and is the most widely used neural network classifier. The MLP has an input layer that takes the characteristics/patterns of the training data, a hidden layer and an output layer with one node per class. The inverse propagation algorithm is used to calculate the weights transported by the network connections. The number of nodes in the hidden layer is determined experimentally [149].

4.2.1.4. Results

Accuracy, specificity and sensitivity are the metrics used to evaluate the performance of each classifier. These values were calculated from equations (5), (6) and (7), respectively, where the values TP (true positive), TN (true negative), FP (false positive) and FN (false negative) are taken from a confusion matrix [137].

TP: Number of normal heart sounds that were classified as normal.
TN: number of abnormal heart sounds that were classified as abnormal.
FP: Number of abnormal heart sounds that were classified as normal.
FN: Number of normal heart sounds that were classified as abnormal.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100. \qquad (\mathbf{18})$$

$$Specificity = \frac{TN}{FP + TN} * 100 \qquad (\mathbf{19})$$

$$Sensitivity = \frac{TP}{FN + TP} * 100 \qquad (\mathbf{20})$$

Tables 11, 12 and 13 show the accuracy, specificity and sensitivity of the three classifiers using the LPC-only, MFCC- only and LPC+MFCC characteristics vectors respectively. It can be seen that the best results were obtained with the LPC+MFCC characteristics, using the SVM and MLP classifiers.

Table 11. Results of classifiers with LPC-only features (6 LPC per segment)

Machine Learning	Accuracy	Specificity	Sensitivity
SVM	88.19%	91.54%	85.55%
KNN	92.91%	96.63%	89.95%
Random Forest	92.91%	91.31%	94.52%
MLP	94.16%	94.31%	94.01%

Table 12. Results of classifiers with MFCC-only features (14 MFCC per segment)

Machine Learning	Accuracy	Specificity	Sensitivity
SVM	94.90%	94.85%	94.85%
KNN	94.65%	96.76%	96.76%
Random Forest	96.52%	95.47%	95.47%
MLP	96.52%	97.13%	97.13%

Table 13. Results of classifiers with LPC+MFCC features (14 MFCC and 6 LPC per segment)

Machine Learning	Accuracy	Specificity	Sensitivity
SVM	96.27%	97.61%	97.61%
KNN	95.52%	98.09%	98.09%
Random Forest	95.27%	94.22%	94.22%
MLP	97.26%	97.91%	97.91%

These results compare favorably to those presented in the literature [138], [139], [140], [141], [142], [143] with the added caveat that our experiment used a bigger number of samples than most previous works. See results in Table 14. It is verified that the LPC and MFCC characteristics can generate better performance in the classification of normal and abnormal heart sounds. However, the number of training samples is not sufficient to guarantee a robust classification model. This work was published in the article [144].

Table 14. Comparison of heart sound classification results

Reference	Accuracy
[138]	86%
[139]	92%
[140]	92%
[141]	96%
[142]	80%
[143]	92%
Proposed method	97%

4.2.2. Proposed method using EWT and power features

At this point, it is decided to use the advantages that EWT offers to eliminate unwanted noise, and the low computational cost that is required to calculate power values. Considering that heart murmurs are manifested in systolic and diastolic intervals, it is proposed to extract power values in these intervals to train Machine Learning models and improve classification performance. Additionally, the performance results obtained with the method proposed in the previous section, and others proposed in the state of the art, are compared.

4.2.2.1. Feature extraction

After the stages of noise reduction (pre-processing) and segmentation using EWT described in section 4.2.1, we proceed to the feature extraction for each cardiac cycle. At this stage we try to improve the limitations of high computational cost by reducing the feature vector size, requiring in turn simpler classification models. Health professionals use different attributes of heart murmurs to achieve their discrimination, the most common of which are their location, duration, pitch and shape [146]. These attributes are related to the different types of murmurs shown in Figure 13 [147].

Taking as a reference the analysis carried out on the representation of the different types of murmurs shown in Figure 13, we propose the division of the systolic and diastolic interval into three segments and calculate the signal power in each segment, obtaining a total of six characteristics (three in the systole and three in the diastole; see Figure 14). In this way, we try to emulate the attributes (location, duration, pitch and shape) that doctors use to aurally detect some type of heart murmur.

The power of a signal is defined as the amount of energy consumed in a time interval. This calculation is widely used to characterize a signal [148]. In the discrete domain, the power of a signal is given by

$$P = \frac{1}{2N+1} \sum_{n=-N}^{n=+N} |x(n)|^2 \tag{21}$$

The signal is called the power signal when $0 < P < \infty$. Figure 14 shows the distribution of the power features.

With this method, it is possible to establish criteria to advance the classification of types of murmurs. For example, if the value of the power P1 of the systole is greater than the P2 and P3 powers, it can be assumed that the type of abnormality is protosystolic.

At this stage, this study is not carried out since not all the samples used in the experiment are labeled with the corresponding type of abnormality. Therefore, it is decided to work only on the classification of normal and abnormal sounds and leave the identification of specific anomalies for future work when a suitable training database is available.

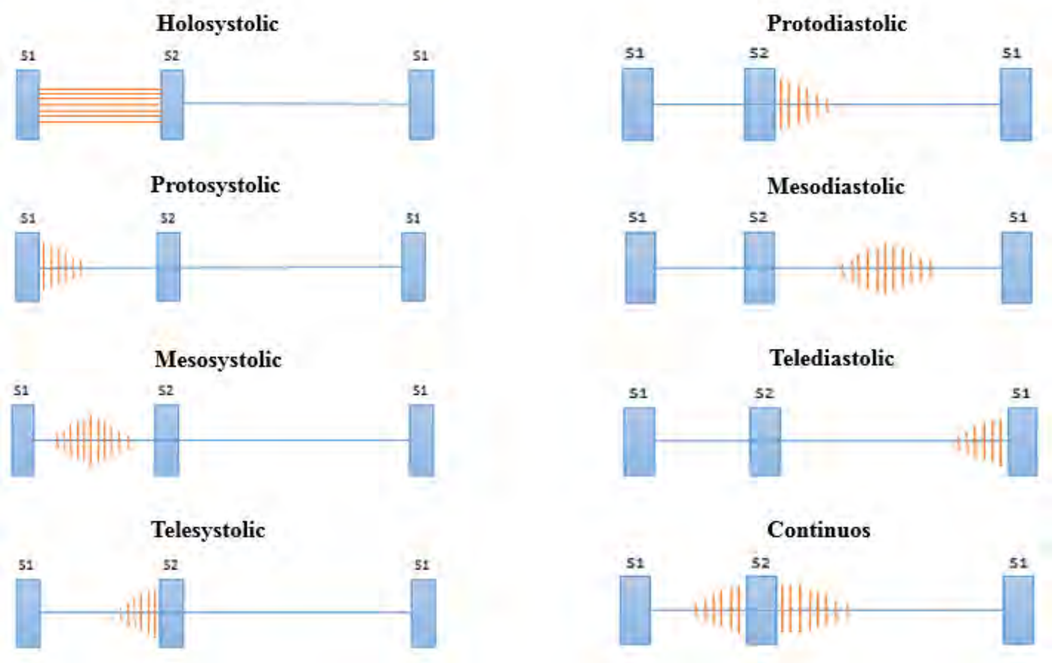


Figure 13. Types of murmurs according to their location.

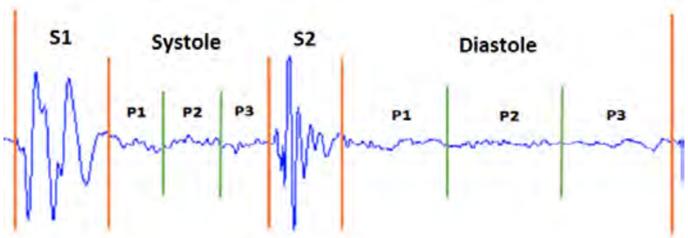


Figure 14. Distribution of power features for a normal heart sound.

4.2.2.2. Experiments and results

This work makes the comparison of models [49], [50], [51], [52] and [144] with the proposed method, since they used characteristics in the domains of time, time–frequency and perceptual domains and techniques used for audio recognition—specifically, the voice. Therefore, different techniques were implemented to extract the characteristics that the authors used in their works. Tables 15 and 16 describe in detail the characteristics used in [49] and [50], respectively. Regarding the classification stage, in [144], the SVM, KNN, MLP and random forest algorithms were used; in [49], deep neural networks and SVM were used, with latter showing the best performance; and in [50], the best performance was obtained using the XgBoost algorithm.

Table 15. Features used in [49].

Domain	Features
Perceptual	19 Mel-frequency cepstral coefficient (MFCC)
Time-frequency	24 DWT features
Perceptual + Time-frequency	19 MFCC and 24 DWT features

Table 16. Features used in [50].

Domain	Features
Statistical	Mean value, median value, standard deviation, mean absolute deviation, quartile 25, quartile 75, iqr, skewness, kurtosis, coefficient of variation
Frequency	Entropy, dominant frequency value, dominant frequency magnitude, dominant frequency ratio
Perceptual	13 MFCC

In the case of the model in [51], 1D and 2D convolutional neural networks (CNN) were used. In the 1D-CNN model, the authors performed the normalization of 1000 samples for each cardiac signal in order to use it as the input to the model. For the 2D-CNN model, they extracted 12 MFCC features in each 30 ms frame, obtaining a 96 × 12 feature matrix. Table 17 shows the configurations of the 1D-CNN and 2D-CNN model.

Finally, in the model in [52], the authors used 12.5 s of each recording of heart sounds, making a total of 50,000 samples. Then, they applied a decimation with a factor of eight twice until they obtained a total of 782 samples in each recording. This model used recurrent neural networks, specifically long short-term memory (LSTM). Table 18 shows the model configuration.

Table 17. Summary of 1D-CNN and 2D-CNN model configurations in [51].

1D-CNN model		2D-CNN model	
Layer	**Output shape**	**Layer**	**Output shape**
Input	1000 × 1	Input	96 × 12
Conv (kernel = 6; strides = 1)	1000 × 8	Conv (kernel = 4; strides = 1)	96 × 12 × 16
Batch-Norm	1000 × 8	Batch-Norm	96 × 12 × 16
Activation Function (ReLu)	1000 × 8	Activation Function (ReLu)	96 × 12 × 16
MaxPool (kernel = 2; strides = 2)	500 × 8	MaxPool (kernel = 2; stride = 2)	48 × 6 × 16
Conv (kernel = 6; strides = 1)	500 × 8	Conv (kernel = 4; strides = 1)	48 × 6 × 16
Dropout (0.4)	500 × 8	Dropout (0.5)	48 × 6 × 16
Activation Func. (ReLu)	500 × 8	Activation Func. (ReLu)	48 × 6 × 16
MaxPool (kernel = 2; strides = 2)	250 × 8	MaxPool (kernel = 2; strides = 2)	24 × 3 × 16
Conv (kernel = 6; strides = 1)	250 × 8	Conv (kernel = 4; strides = 1)	24 × 3 × 16
Dropout (0.4)	250 × 8	Dropout (0.5)	24 × 3 × 16
Activation Func. (ReLu)	250 × 8	Activation Func. (ReLu)	24 × 3 × 16
MaxPool (kernel = 2; strides = 2)	125 × 8	MaxPool (kernel = 2; strides = 2)	12 × 1 × 16
Flatten	1000	Flatten	192
Dense	512	Dense	256
Dropout (0.4)	512	Dropout (0.4)	256
SoftMax	2	SoftMax	2

Table 18. Summary of LSTM model configurations in [52].

Layer	Output shape
Input	782 × 1
LSTM	782 × 64
Dropout (0.35)	782 × 64
LSTM	1 × 32
Dropout (0.35)	1 × 32
Dense	2
SoftMax	2

In the models in [49], [50] and [144], segmentation was performed manually for the identification of the cardiac cycle, S1 and S2 sounds and systolic and diastolic intervals, unlike the proposed method that uses the automatic segmentation method described in Section 4.1. In the case of the models proposed in [51], normalization was applied to 1000 samples in each cardiac cycle and the MFCC features were calculated. Furthermore, for the model in [52], the decimation process was carried out in each recording, until a total of 782 samples were obtained in each signal.

The same dataset described in section 4.2.1.1 was used in this experiment. The well-known tool for data mining and machine learning Weka (Knowledge Environment of Waikato University) was used to construct classification models [150]. The power characteristics were used as inputs for the four classifiers (SVM, KNN, RF and MLP). A cross validation of 10 folds was used to evaluate the performance of each classifier.

Table 19 shows the comparison of the accuracy results between the methods in [49], [50], [144] and the proposed method using the different ML models. Table 20, 21 and 22 show the results of specificity, sensitivity, and the area under the receiver operating characteristic curve (AUC), respectively. Table 23 presents the comparison of the results obtained with the proposed method using the KNN classifier and the models proposed in [51] and [52].

Table 19. Results of accuracy between the methods in [49], [50], [144] and the proposed method.

Feature Extraction	Classifier			
	SVM	**KNN**	**RF**	**MLP**
[49]: MFCC	74.65%	85.96%	87.20%	85.83%
[49]: DWT	86,95%	88.19%	92.17%	90.31%
[49]: MFCC + DWT	90.68%	91.18%	91.42%	91.55%
[50]: Statistical, frequency and perceptual	84.47%	93.66%	93.66%	92.54%
[144]: LPC	88.19%	92.91%	92.91%	94.16%
[144]: MFCC	94.90%	94.65%	96.52%	96.52%
[144]: LPC + MFCC	96.27%	95.52%	95.27%	97.26%
Proposed Method: EWT + Power	92.42%	**99.25%**	99.00%	98.63%

Table 20. Results of specificity between the methods in [49], [50], [144] and the proposed.

Feature Extraction	Classifier			
	SVM	**KNN**	**RF**	**MLP**
[49]: MFCC	76.92%	86.92%	87.94%	82.82%
[49]: DWT	85.38%	83.84%	91.79%	88.46%
[49]: MFCC + DWT	90.76%	90.00%	90.00%	91.53%
[50]: Statistical, frequency and perceptual	86.06%	95.90%	95.18%	94.69%
[144]: LPC	91.54%	96.63%	91.31%	94.31%
[144]: MFCC	94.85%	96.76%	95.47%	97.13%
[144]: LPC + MFCC	97.61%	98.09%	94.22%	97.91%
Proposed Method: EWT + Power	**100.00%**	**100.00%**	99.22%	**100.00%**

Table 21. Results of sensibility between the methods [49], [50], [144] and the proposed.

Feature Extraction	Classifier			
	SVM	**KNN**	**RF**	**MLP**
[49]: MFCC	72.53%	85.06%	86.50%	88.67%
[49]: DWT	88.43%	92.28%	92.53%	92.04%
[49]: MFCC + DWT	90.60%	92.28%	92.77%	91.56%
[50]: Statistical, frequency and perceptual	83.84%	91.28%	92.05%	90.25%
[144]: LPC	85.55%	89.95%	94.52%	94.01%
[144]: MFCC	94.85%	96.76%	95.47%	97.13%
[144]: LPC + MFCC	97.61%	98.09%	94.22%	97.91%
Proposed Method: EWT + Power	87.18%	98.57%	**98.80%**	97.41%

Table 22. Results of AUC between the methods [49], [50], [144] and the proposed method.

Feature Extraction	Classifier			
	SVM	**KNN**	**RF**	**MLP**
[49]: MFCC	74.73%	86.55%	95.26%	93.05%
[49]: DWT	86.91%	87.58%	97.94%	95.09%
[49]: MFCC + DWT	90.69%	91.91%	98.20%	96.22%
[50]: Statistical, frequency and perceptual	84.45%	94.10%	98.13%	97.66%
[144]: LPC	88.05%	92.19%	97.95%	97.81%
[144]: MFCC	94.90%	94.63%	99.53%	99.37%
[144]: LPC + MFCC	96.22%	95.17%	99.46%	99.60%
Proposed Method: EWT + Power	92.10%	91.81%	99.62%	98.75%

Our method obtained an accuracy of 99.25%, a specificity of 100%, a sensitivity of 98.57% and an AUC of 91.81% using the KNN classifier, which was the best result obtained overall. Furthermore, with all other classifiers tested, our method ranks at or close to the top, suggesting that the proposed segmentation and feature extraction algorithms are indeed useful, irrespective of the classification model applied on their output.

Table 23. Comparison of results between the methods in [51, 52] and the proposed method.

Method	**Accuracy**	**Specificity**	**Sensibility**	**AUC**
[51]: 1D-CNN	91.80%	85.43%	97.34%	91.39%
[51]: 2D-CNN and MFCC	86.83%	80.52%	93.19%	86.86%
[52]: LSTM	54.42%	100%	0%	50%
Proposed Method: EWT + Power + KNN	99.25%	100%	98.57%	91.81%

Finally, Table 24 presents the confusion matrix for each classification model using as inputs the characteristics extracted in the proposed method. It can be seen that, in all cases, a good performance was obtained in the detection of normal heart sounds, taking into account the fact that a 10-fold cross validation was applied.

Table 24. Confusion matrix of proposed method.

Machine Learning		**Confusion Matrix**	
		Normal	**Abnormal**
SVM	Normal	415	0
	Abnormal	61	329
KNN	Normal	415	0
	Abnormal	6	384
Random Forest	Normal	412	3
	Abnormal	5	385
ANN	Normal	415	0
	Abnormal	11	379

The main limitation that exists in the proposed method is when the recording of the cardiac signal has ambient noises with a high amplitude, since these noises can be contained in different frequency bands. Therefore, automatic segmentation can identify false positives in the systolic or diastolic interval. Similarly, the power features can vary when the signal has these types of noise and the classifier could in turn be confused regarding whether a heart murmur is present, with the detected murmur actually being an ambient noise. These results were published in the article [151].

5. Synthesis of Heart sounds

In this chapter we present the advances obtained during research in the synthesis of heart sounds. After exploring the mathematical models proposed in the state of the art and evaluating the possibility of implementing GAN models for the generation of biomedical signals, in this section we propose different GAN-based architectures and incorporate mathematical models that help define a normal heart sound pattern and types of murmurs as realistic as possible, considering the main limitations described in section 2.3.1.

The first part of this chapter describes the implementation of a GAN model for the generation of normal heart sounds. Different validation tests using Mel Cepstral Distortion (MCD) and classification models are presented. These results compare favorably with the mathematical model [23]. Although good results were obtained with normal cardiac signals, a great limitation is evident in generating abnormal types of sounds, since there is a low availability of this type of signals. For this reason, a model based on GAN is proposed that consists of refining the characteristics of an ideal signal obtained with a mathematical model from real signals. We have called this model FeaturesGAN and it is considered one of the main contributions of this research. Validation results are presented using techniques such as: MCD, SSIM, PCA and t-SNE, the performance of classification models is also evaluated and MOS tests are carried out with doctors obtaining good scoring results.

5.1. Synthesis of Normal Heart Sounds Using GAN and EWT

In this section, we propose a model based on Generative Adversary Networks (GANs) to generate normal synthetic heart sounds. Additionally, a denoising algorithm is implemented using the empirical wavelet transform (EWT), allowing a decrease in the number of epochs and the computational cost that the GAN model requires. A distortion metric (mel–cepstral distortion) was used to objectively assess the quality of synthetic heart sounds. The proposed method was favorably compared with a mathematical model that is based on the morphology of the phonocardiography (PCG) signal published as the state-of-the-art. Additionally, different heart sound classification models proposed as state-of-the-art were also used to test the performance of such models when the GAN-generated synthetic signals were used as test dataset. In this experiment, good accuracy results were obtained with most of the implemented models, suggesting that the GAN-generated sounds correctly capture the characteristics of natural heart sounds.

5.1.1. Proposed method

The proposed method is made up of two main stages as shown in Figure 15. The first stage consists of the implementation of a GAN architecture to generate a synthetic heart sound, and the second stage is in charge of reducing the noise of the synthetic signal using the Empirical Wavelet Transform (EWT). This last stage consists of a post-processing applied to the signal generated by the GAN, in order to attenuate the noise level. Therefore, it makes possible to reduce the number of epochs (and consequently the computational cost) required to train the GAN until obtaining a low noise output signal. Figure 16 shows the diagram of the implemented GAN architecture, and each of its components is described below.

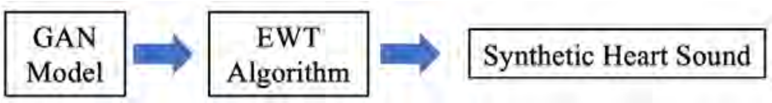


Figure 15. General diagram of proposed method.

Noise: A Gaussian noise with a size of 2000 samples is used as input to the generator. The mean and standard deviation of the noise’s distribution are 0 and 1 respectively.

Generator Model: Figure 17 shows a diagram of the generating network, it begins with a dense layer with ReLu activation function; followed by three convolutional layers with filters of size 128, 64 and 1 respectively, each of these layers have ReLu activation function, kernel size of 3 and stride of 1; finally, there is a dense layer with tanh activation function. The Padding parameter is set to 'same' to maintain the same data dimension in the input and output of the convolutional layer.

Discriminator Model: Figure 18 shows a diagram of the discriminator network. It begins with a dense layer with ReLu activation function; followed by 4 convolutional layers with filters of size 256, 128, 64 and 32 respectively, each of these layers uses Leaky ReLu activation function, kernel size of 3 and stride of 1, additionally, between each convolution layer there is a Dropout of 0.25; finally, there is a dense layer with tanh activation function. The Padding parameter is set to 'same' to maintain the same data dimension in the input and output of the convolutional layer.

The architecture of the DCGAN model was taken as the basis for the implementation of the generator and discriminator, taking into account the use of 1D convolutional layers. This type of architecture has been implemented in other works published in the state of the art [89, 94]. For our work, some hyperparameters were modified in order to reduce the computational cost.

Dataset of Heart Sounds: 100 normal heart sounds obtained from the Physionet database [30] are used, with a sampling frequency of 2 KHz and 1 second of duration. For this dataset, those signals with a similar heart rate were selected, that is, all signals have a similar systolic and diastolic interval duration.

Optimization**:** The Adam optimizer is used, since it is one of the best performers in this type of architecture. A Learning Rate of 0.0002 and a beta of 0.3 are set.

Loss function: Binary Cross-entropy function is used in this work. This function computes the cross-entropy loss between true labels and predicted labels.

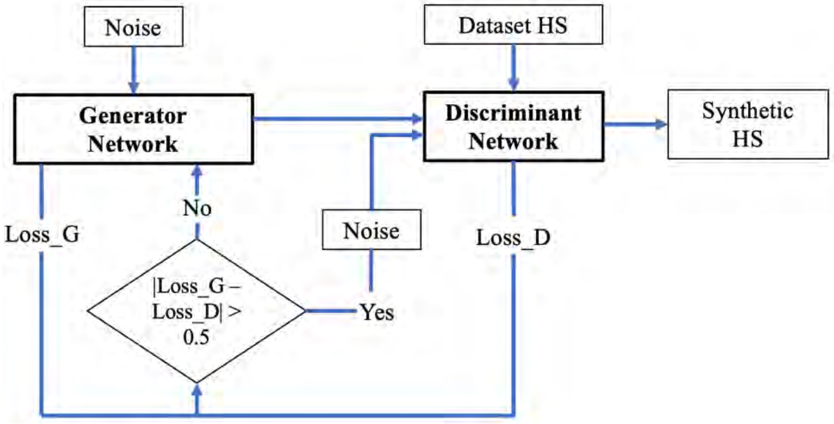


Figure 16. Proposed GAN diagram.

Subsequently, the difference between generator and discriminator losses is analyzed. If this difference is greater than 0.5, the input data to the discriminator is switched to a Gaussian noise with a mean of 0 and standard deviation of 1, not the generator output as it is otherwise. With this method, a convergence in the loss functions of the generator and discriminator can be achieved.

As mentioned before, the second stage of the proposed method aims to reduce the noise level of the synthetic signal generated by the GAN model. It is understood that as the number of epochs in the training of the generator and discriminator models increases, the noise of the synthetic signal is attenuated, however, it requires many epochs, and in turn, a long computation time [**152**]. Therefore, in order to reduce the number of GAN training epochs required to generate synthetic signals with acceptable noise levels, it was decided to introduce a post-processing stage using the algorithm proposed in [**118**], called Empirical Wavelet Transform (EWT). The EWT allows the extraction of different components of a signal by designing an appropriate filter bank. The theory of the EWT algorithm is described in detail in [**118**]. This algorithm has been used in different signal processing applications [119], [120] and [121]. In [**151**] a modified version of this algorithm was used as a pre-processing stage in the analysis of heart sounds. Its implementation is described in more detail in [**151**]. In this work it was decided to use as a reference the method proposed in [**151**] to reduce the noise of the synthetic signal.

Taking into account that the frequency range of the S1 and S2 sounds is between 20 **-** 200 Hz [**122**], it was decided to modify the edge selection method that determines the number of components for the EWT algorithm. The signal is then broken down into two frequency bands, the first component corresponds to the frequency range between 0 - 200 Hz, while the second component corresponds to frequencies over 200 Hz. Therefore, in this work the signal corresponding to the first component is used.

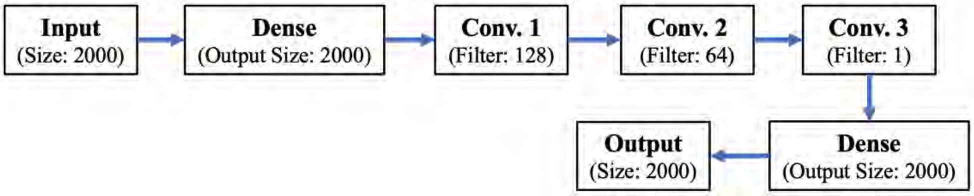


Figure 17. Architecture of the generator model.

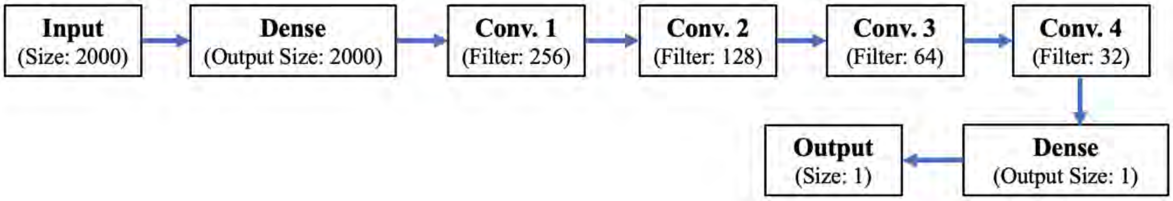


Figure 18. Architecture of the discriminator model.

Taking into account that the input of the GAN model is a Gaussian noise, and in turn, the output of the model during the first epochs of training is expected to be a signal mixed with Gaussian noise, it is decided to do a test using a real heart signal mixed with Gaussian noise to evaluate the performance of the proposed EWT filter on signals with the same expected characteristics of the Generator's output. Figures 19a and 19b show, respectively, a real heart sound and the same heart sound mixed with low amplitude Gaussian noise, while Figures 19c and 19d show their respective Fourier Transforms (FFT). In this last figure, it can be seen in blue the frequency components between 0 and 200 Hz, and in green the frequency components above 200 Hz caused by the Gaussian noise. The signal shown in Figure 19b was then used as an input example to the proposed EWT algorithm to illustrate its de-noising action. Figure 19d shows the two frequency bands extracted with EWT using the FFT, while Figures 19e and 19f show the two components extracted from the noisy signal in the time domain. As can be seen, the signal obtained in Figure 19e presents a lower noise level and is comparable with the original cardiac signal shown in Figure 19a.

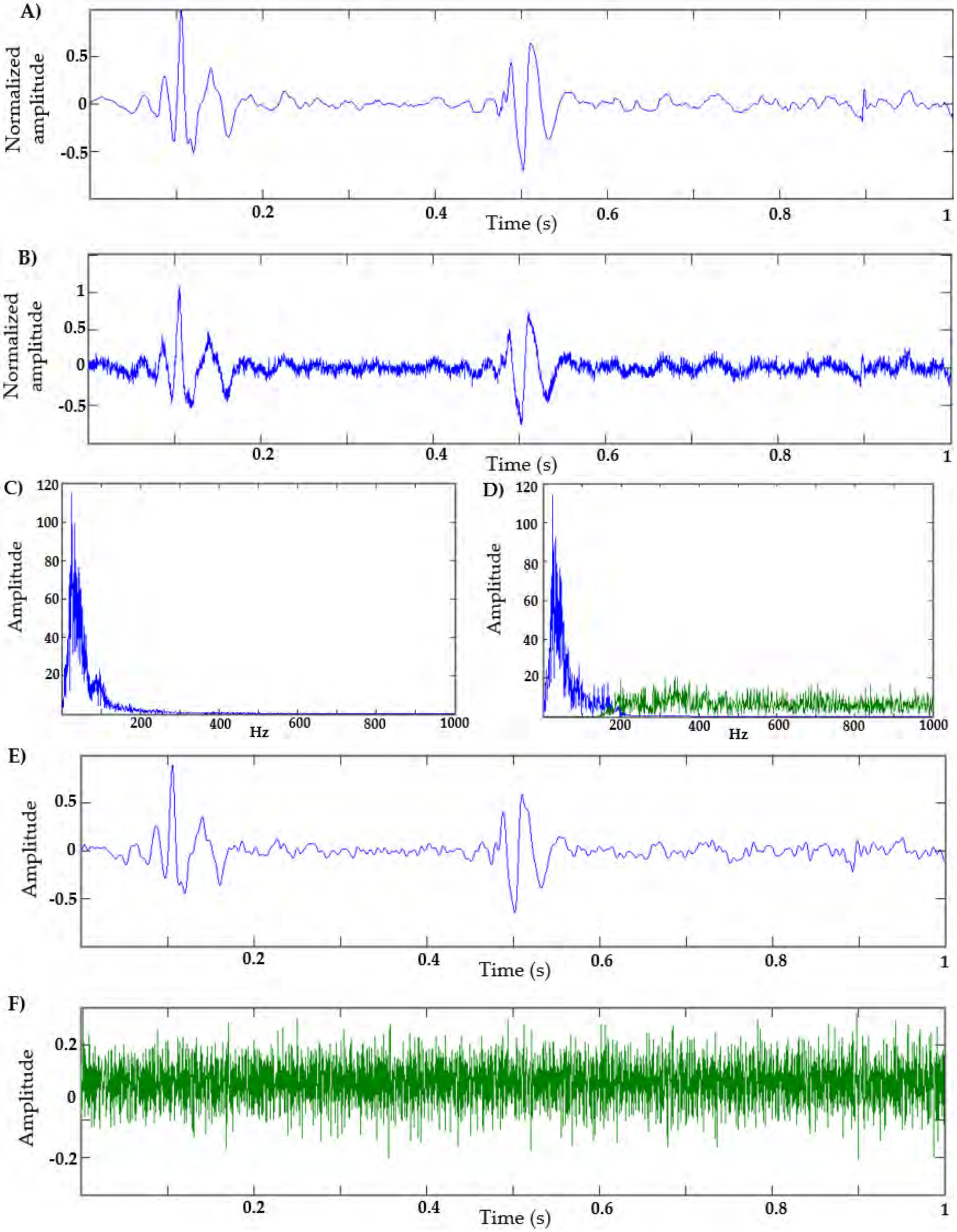
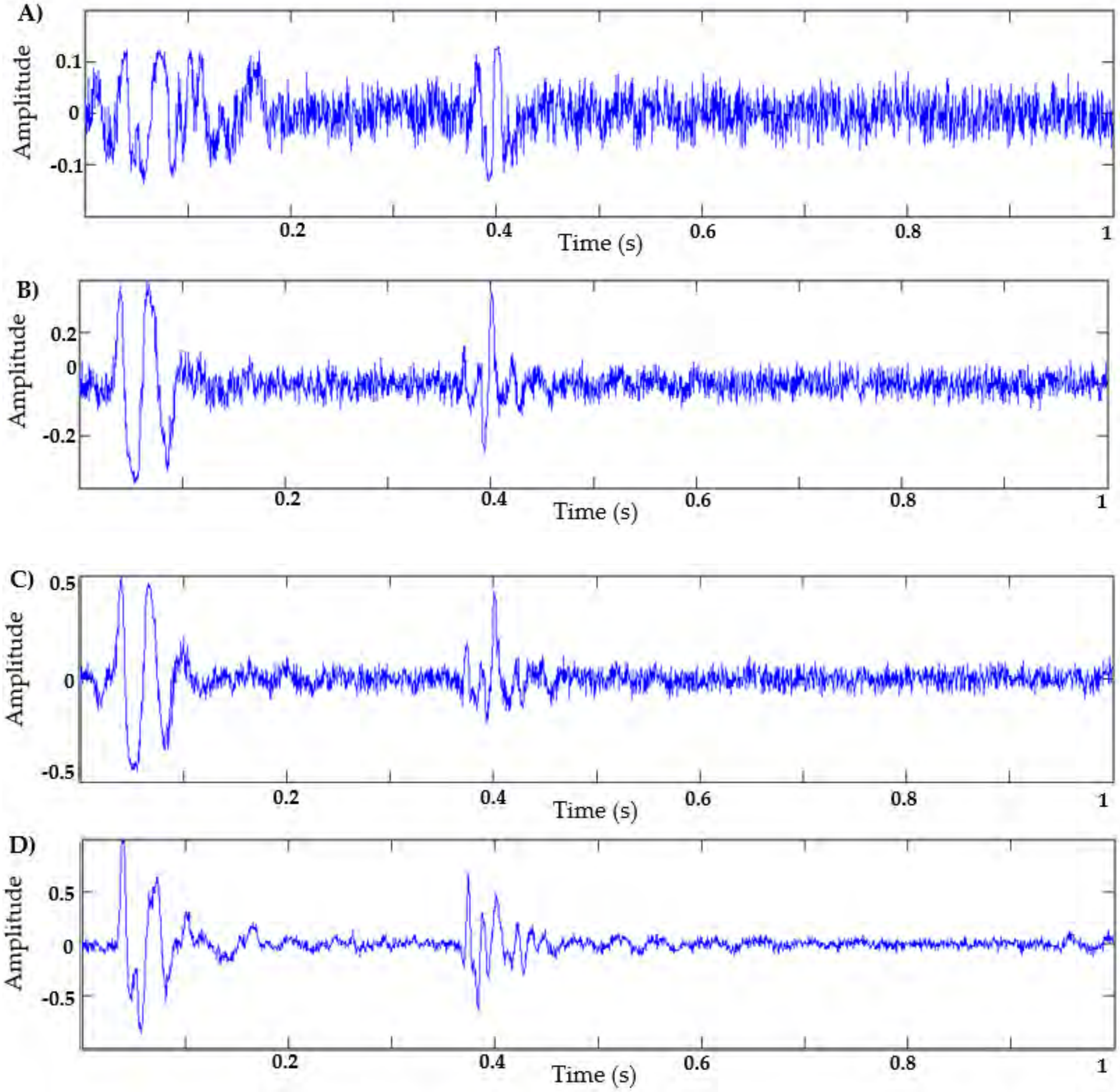


Figure 19. A) Real heart sound; B) Real heart sound with Gaussian noise; C) FFT of real heart sound; D) FFT of heart sound with Gaussian noise; E) EWT component of the noisy signal in the frequency range of 0 – 200 Hz; F) EWT component of the noisy signal in the frequency range greater than 200 Hz.

5.1.2. Experiments and Results

The proposed GAN model was trained for a total of 2000 epochs. Figure 20 shows sample output signals generated at different epochs. As can be observed, as the epochs increase, the signals present a more realistic form. In this work, the EWT filter is a post-processing stage that is applied to the synthetic signal generated with 2000 training epochs, in order to reduce the noise level of the generator output. Therefore, the EWT filter is not part of the training loop for the GAN. This number of epochs was determined after observing the synthetic signals obtained at different training points (from 100 epochs to 12000 epochs). It was observed that from 2000 epochs, the synthetic signal has a shape very similar to a natural signal, but with a relatively high noise level, as shown in Figure 20e. Therefore, it was decided to generate the synthetic signals up to 2000 epochs, and subsequently apply the proposed EWT algorithm. Figure 20f and 20g shows the results of a synthetic signal generated with 12000 training epochs without applying an EWT algorithm or a natural signal, respectively.



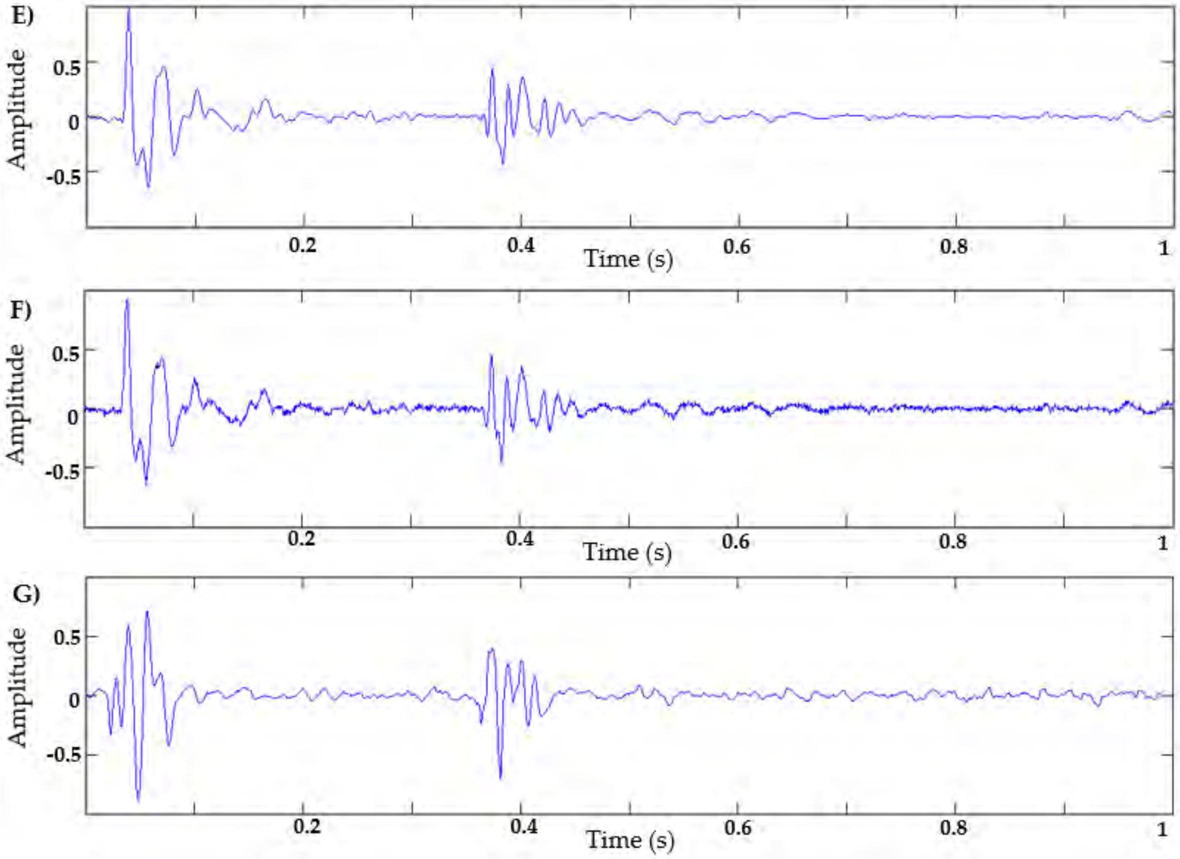


Figure 20. A) Synthetic signal with 100 epochs; B) synthetic signal with 500 epochs; C) synthetic signal with 1000 epochs; D) synthetic signal with 2000 epochs; E) synthetic signal with 2000 epochs + EWT; F) synthetic signal with 12000 epochs; G) natural signal of heart sound.

In this work, a comparison is made between the proposed method and a mathematical model proposed in [**23**], in order to determine which method generates a cardiac signal realistic enough to be used by a classification model. The mathematical method [**23**] is inspired by a dynamic model that generates a synthetic electrocardiographic signal, as described [**154**]. Therefore, this model is based on the morphology of the phonocardiographic signal and has been used as a reference in other proposed methods [**153**]. The equation for the reference model [**23**] is described below (equation 22):

$$\dot{z} = -\sum_{i\,\in\{S1^{-}S1^{+}S2^{-}S2^{+}\}} \left(\frac{a_i}{\sigma_i}(\theta - \mu_i)e^{\left(-\frac{(\theta-\mu_i)^2}{2\sigma_i^2}\right)} cos(2\pi f_i\theta - \varphi_i) + 2\pi\alpha_i f_i\, e^{\left(-\frac{(\theta-\mu_i)^2}{2\sigma_i^2}\right)} sin(2\pi f_i\theta - \varphi_i) \right)$$

Where α_i, μ_i and σ_i are the parameters of amplitude, center and width of the Gaussian terms, respectively; f_i and φ_i are the frequency and the phase shift of the sinusoidal terms, respectively; and θ is an independent parameter in radians that varies in the range -π, π for each beat. The parameters used by the authors in [23] are summarized in Table 25.

The ordinary differential equation $\dot{z}$ was solved using the numerical Runge-Kutta method of fourth order, using the Matlab software. Using the values in Table 27, we obtain the graph in Figure 21A, which represents the S1 and S2 sounds of a cardiac cycle. Figure 21B shows a natural signal of heart sound.

Table 25. Parameters used in [**23**] to generate normal heart sounds.

Index (i)	S1 (-)	S1 (+)	S2 (-)	S2 (+)
α_i	0.4250	0.6875	0.5575	0.4775
μ_i (radians)	$\pi/12$	$3\pi/19$	$3\pi/4$	$7\pi/9$
σ_i	0.1090	0.0816	0.0723	0.1060
f_i (Hz)	10.484	11.874	11.316	10.882
φ_i (radians)	$3\pi/4$	$9\pi/11$	$7\pi/8$	$3\pi/4$

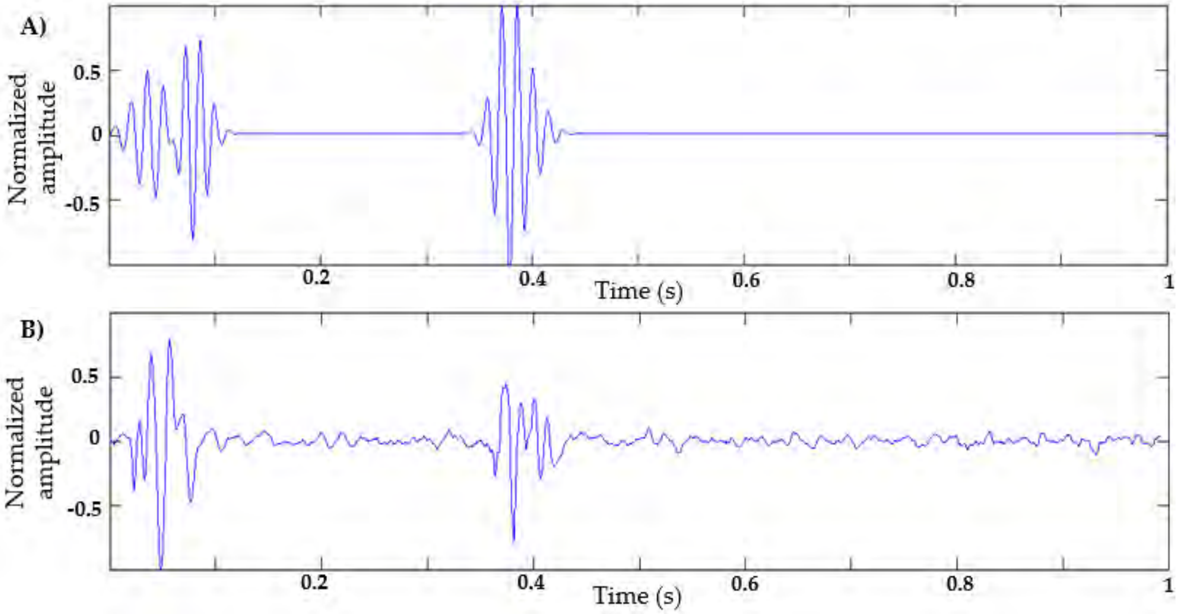


Figure 21. A) Synthetic heart sound generated by the model [23]; B) natural signal of the heart sound.

5.1.2.1. Results using Mel Cepstral Distortion (MCD)

Mel-Cepstral Distortion (MCD) is a metric widely used to objectively evaluate audio quality [**155**], and its calculation is based on Mel-frequency cepstral coefficients (MFCC). This method has been widely used in the evaluation of voice signals, since many automatic voice recognition models use feature vectors based on MFCC coefficients [**155**]. The parameters used for MFCC extraction are the following: the length of the analysis window (frame) in seconds is 0.03 s, the step between successive windows in second is 0.015 s, the number of cepstral to return in each windows (frame) is 14, the number of filters in the filterbank is 22, the FFT size is 4000, the lowest band edge of mel filters is 0, the highest band edge of mel filters is 0.5, and no window function is applied to the analysis window of each frame.

Basically, MCD is a measure of the difference between two MFCC sequences. In Vasilijevic and Petrinovic [**156**], different ways of calculating this distortion are presented. In equation (3), the formula used in [**155**] is defined, where C_{MFCC} and $C_{M\hat{F}CC}$ are the MFCC vectors of a frame of the original and study signal, respectively, and L represents the number of coefficients in that frame. MCD_{FRAME} represents the MCD result obtained in a frame.

$$MCD_{FRAME} = \sum_{l=1}^{L} \left(C_{MFCC}[l] - C_{M\hat{F}CC}[l] \right)^2 \qquad \textbf{(23)}$$

In this work, it was decided to use this objective measurement method to evaluate the similarity between natural and synthetic heart signals, taking into account that heart sounds are audio signals that are typically evaluated using human hearing, and the MFCC coefficients have already been used in the analysis of heart sound signals [49], [50], [144] and [151].

A set of 400 natural normal heart sounds taken from the Physionet [**16**] and Pascal [**17**] databases were used. Each signal was cut to a single cardiac cycle, with normalized duration of 1 s, applying a resampling on the signal. Signals were also normalized in amplitude, and those signals with a similar heart rate were chosen. Those natural signals are compared to a total of 50 synthetic heart sounds generated using the proposed method, and 50 synthetic signals were generated using the model [**23**]. In the case of model [**23**], the α_i parameters were obtained with random variables in the range of 0.3 to 0.7, in order to generate different wave signals. The other parameters were established as shown in Table 3. Additionally, the synthetic signal was mixed with a white Gaussian noise, as indicated in the article [**23**]. These synthetic signals have a sampling rate of 2 KHz, a duration of 1 s, and are amplitude-normalized. All signals (natural and synthetic) have a similar heart rate-- that is, the size of the systolic and diastolic interval is similar in all the signals.

The first evaluation step was to calculate the MCD between the natural signals-- that is, the MCD between each natural signal and the remaining natural samples. A total of 399 MCD values were computed and then averaged. This same procedure was applied with the synthetic signals, i.e., computing the MCD between each synthetic signal and each natural signal, obtaining 400 MCD values that were then averaged. Figure 22 shows a schematic of the procedure to compute the MCD. Figure 23 shows the results of the average MCD between the natural signals (blue color), the average distortions of the synthetic signals generated with the proposed method (red color), the average distortions of the synthetic signals generated with the proposed method without applying an EWT algorithm (dark blue color), and the average distortion of the synthetic signals generated with the model in [**23**] (green color) using the MCD method.

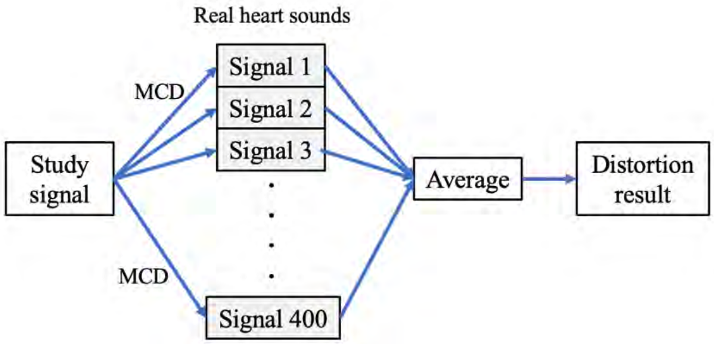


Figure 22. General diagram of the procedure to calculate the signal distortion.

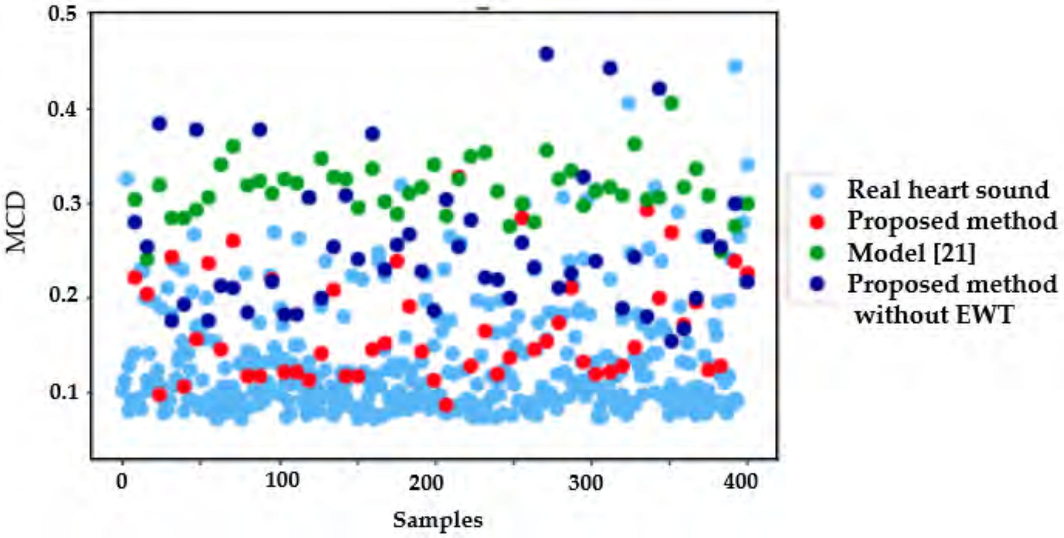


Figure 23. Result of MCD distortions.

To verify that the generator in the GAN model was not just copying some of the training examples, we computed the MCD distortion of a synthetic signal against each of the signals used in the training dataset. Figure 24 shows the resulting MCD values, with and without the EWT postprocessing. It can be seen that in none of the cases is there a MCD value equal to approaching zero; therefore, the generated signal is not a copy of any of the training inputs.

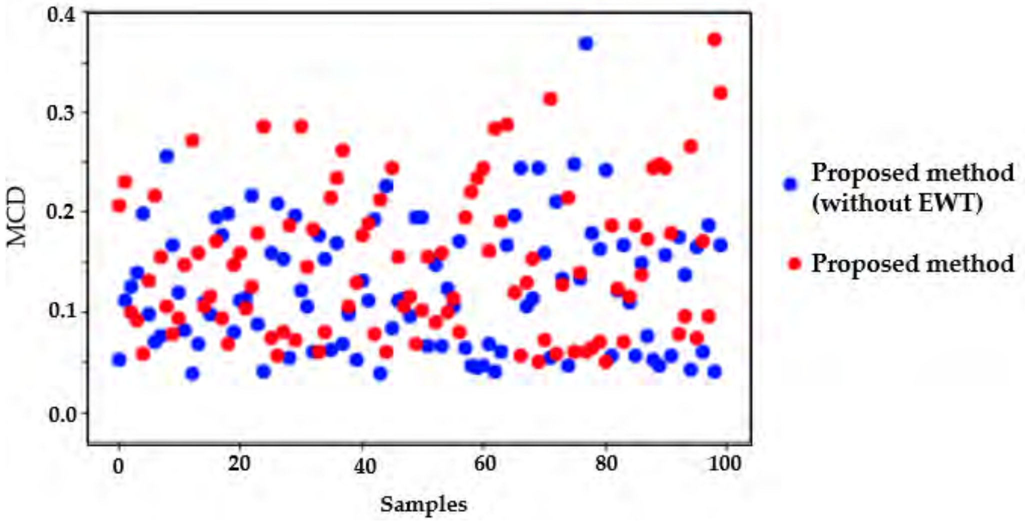


Figure 24. Result of MCD distortion using one synthetic signal with training dataset

It can be seen in Figure 11 that the distortions of the natural signals and the signals generated using the proposed method are in the same range, unlike the distortion obtained with the signals generated using the model [**23**].

5.1.2.2. Results using classification models

In this section, different heart sound classification models published in the state-of-the-art are tested [49], [50], [144] and [151]. These models focus on discrimination between normal and abnormal heart sounds. They were trained with a total of 805 heart sounds (415 normal and 390 abnormal), obtained from the following databases: the PhysioNet/Computing in Cardiology Challenge 2016 [**16**], Pascal challenge database [**17**], Database of the University of Michigan [**13**], Database of the University of Washington [**14**], Thinklabs database (digital stethoscope) [**15**] and 3M database (digital stethoscope) [**145**]. Table 26 presents the different characteristics extracted in the proposed classification methods [49], [50], [144] and [151]. These characteristics belong to the domains of time, frequency, time-frequency and perceptual.

Table 26. Features extracted in models [49], [50], [144] and [151].

Reference	Features
[**151**]	Six power values (three in systole and three in diastole)
[**49**]	Nineteen mel-frequency cepstral coefficients (MFCC) and 24 Discrete Wavelet Transform features
[**50**]	**Statistical domain**: Mean value, median value, standard deviation, mean absolute deviation, quartile 25, quartile 75, IQR, skewness, kurtosis, coefficient of variation. **Frequency domain**: Entropy, dominant frequency value, dominant frequency magnitude, dominant frequency ratio. **Perceptual domain**: 13 MFCC
[**144**]	Six linear prediction coefficients (LPC) + 14 MFCC per segment (S1, S2, systole and diastole)

Each feature set was used as input to the following machine learning (ML) models: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and Multilayer Perceptron (MLP). In Table 27, the accuracy results of each one of the combinations of characteristics with the ML models are presented, applying a 10-fold cross validation. The analysis of these results is described in more detail in [**151**].

Table 27. Accuracy results of the methods proposed in [49], [50], [144] and [151], taken from article [151].

Feature Extraction	Classifier			
	SVM	KNN	RF	MLP
[151]: EWT + Power	92.42%	99.25%	99.00%	98.63%
[49]: MFCC + DWT	90.68%	91.18%	91.42%	91.55%
[50]: Statistical, frequency and perceptual	84.47%	93.66%	93.66%	92.54%
[144]: LPC + MFCC	96.27%	95.52%	95.27%	97.26%

In this work, these classification models were used to test the synthetic signals generated with the proposed method (GAN). Therefore, 50 synthetic signals were used as the test dataset, and the accuracy results are presented in Table 28. The same procedure was done with the synthetic signals without applying the EWT algorithm, with the accuracy results presented in Table 29.

Table 28. Accuracy results of synthetic signals, using the trained models proposed in articles [49], [50], [144] and [151].

Feature Extraction	Classifier			
	SVM	KNN	RF	MLP
[151]: EWT + Power	100%	100%	100%	100%
[49]: MFCC + DWT	80%	90%	78%	82%
[50]: Statistical, frequency and perceptual	98%	78%	78%	60%
[144]: LPC + MFCC	98%	96%	96%	82%

Table 29. Accuracy results of synthetic signals without applying EWT algorithm, using the trained models proposed in articles [49], [50], [144] and [151].

Feature Extraction	Classifier			
	SVM	KNN	RF	MLP
[151]: EWT + Power	100%	100%	100%	100%
[49]: MFCC + DWT	85%	88%	80%	90%
[50]: Statistical, frequency and perceptual	90%	78%	75%	60%
[144]: LPC + MFCC	95%	90%	90%	78%

The best results were obtained with the power characteristics proposed in [151]. However, in several combinations of characteristics and ML models results of precision greater than 90% were obtained, as was the case of the combination of LPC and MFCC proposed in [144]. From these results it can be argued that the synthetic signals generated with the proposed method have similar characteristics to the natural signals, since the classification results on both type of signals are similar.

5.2. FeaturesGAN: adversarial model for the synthesis of heart sound and murmurs

In the previous section, a GAN-based method was presented for the generation of normal cardiac sounds. Different validation experiments were performed using Mel Cepstral Distortion metric and the performance of different Machine Learning models, so it is possible to evidence a strong indication that synthetic signals can be used to improve the performance of cardiac sound classification models by increasing the number of available training samples. However, one of the main drawbacks or limitations in the proposed model was the presence of the mode collapse in the training stage of the GAN network (this phenomenon is described in section 2.5).

Although a diversity was obtained in the synthetic samples, it was necessary to retraining the GAN model several times to reach a good number of different samples. Undoubtedly one of the main factors that caused a mode collapse in the model was the small number of actual samples used to train the discriminator model. Given this limitation, a method capable of generating cardiac sounds with different types of murmurs is proposed assuming that the number of real signals is even smaller, this method has been named FeaturesGAN.

5.2.1. General description

FeaturesGAN main objective is to take advantage of the phenomenon of mode collapse that occurs in the training of the GAN model to combine an ideal signal (obtained from a mathematical model) with characteristics in the domain of time, frequency or perceptual of real signals, as described in Figure 25. In this way, the model produces many synthetic signals, preserving the main characteristics of the available real signals. In other words, FeaturesGAN makes a refinement in the characteristics of the signals obtained by the mathematical model to achieve a more realistic result.

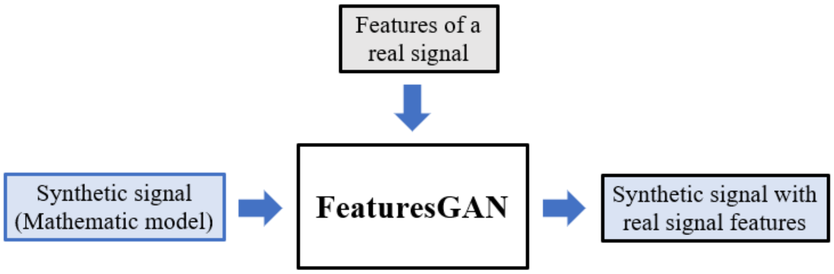

Features of a real signal
Synthetic signal (Mathematic model)
FeaturesGAN
Synthetic signal with real signal features

Figure 25. General diagram of FeaturesGAN

The main novelty of this approach is the structure of each sample of the dataset that is used to train the discriminant model, that is, each sample consists of a concatenation of an ideal signal in time (obtained by a mathematical model) and characteristic vectors o transformations of a real signal, as shown in the diagram of Figure 26.

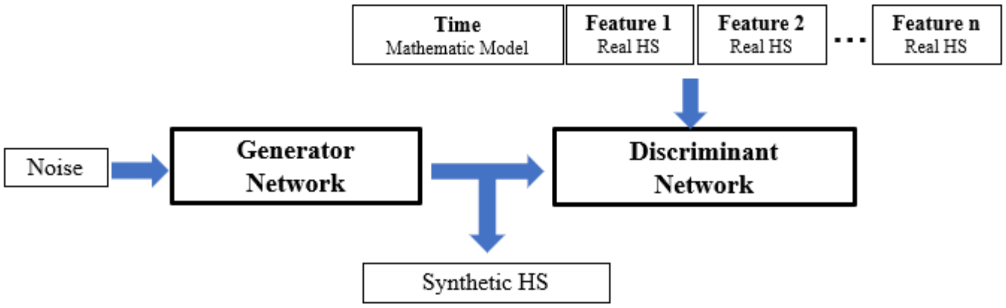


Figure 26. General architecture of FeaturesGAN

The generator model must extract the different vectors of features in each iteration, in order to produce a signal in the time domain very similar to one obtained with the mathematical model, and in turn with features very similar to the available real signals.

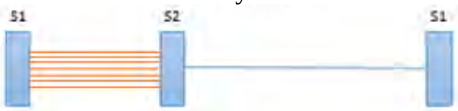
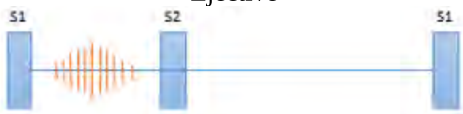
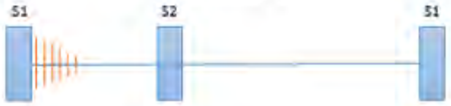
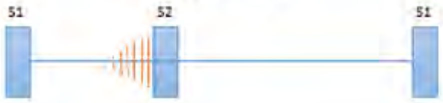
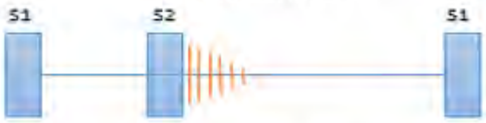
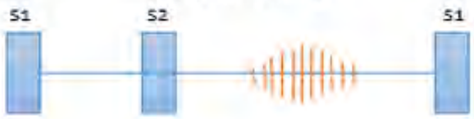
5.2.2. Definition of mathematical models for the generation of normal heart sounds and types of murmurs

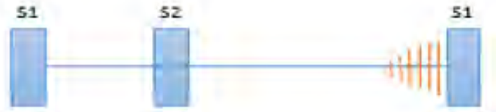
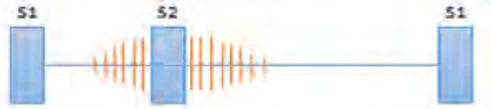
Taking into account that this method requires a signal that represents a cardiac sound in the time domain, the mathematical model proposed in [23] is used for the generation of S1 and S2 sounds. Random variables were added in the amplitude, centered and wide parameters to obtain different waveforms for S1 and S2. For the case of cardiac murmurs, a mathematical model is defined that is capable of producing a signal with a morphology similar to the different types of murmurs described in Table 10. The different types of murmurs are described below and subsequently the respective mathematical models.

5.2.2.1. Cardiac murmurs

Cardiac murmurs are noise caused by a turbulent blood flow through heart valves or near the heart [146]. These noises can be characterized according to several criteria, such as: i) localization in the cardiac cycle: systolic puffs (located in the systole, between S1 and S2), diastolic (located in the diastole, between S2 and the S1 sound of the next cardiac cycle), and continuous (begin in the systole and surpass S2 to end in diastole); ii) Duration: According to its extension in the systole or diastole, there is talk of short (protosystolic for example) and long (pansystolic for example); iii) Morphology: It refers to the dynamic aspect of the breath, can be presented as a homogeneous or rhomboid intensity [147]. Table 30 illustrates the different types of murmurs taking into account the location, duration and morphology criteria.

Table 30. Description of the different types of cardiac murmurs [147]

Type of cardiac murmur	Description
Pansystolic 	They occupy all the systole without varying their morphology (rectangular). They usually appear in the insufficiency of the atrioventricular valves, and in most interventricular communications.
Ejective 	They are rhomboidal murmurs and are auscultated when there is stenosis in ventricular output tracts and in pulmonary or aortic valves.
Protosystolic 	They start up close to S1 to decrease in intensity and end before S2. They are characteristic of small muscle interventricular communications.
Telesystolic 	They are short murmurs, located in the middle or at the end of the systole. They are rare in pediatrics. They usually associate with mild pathology of the mitral valve.
Protodiastolic 	They are short murmurs, of decreasing intensity. They are produced by the insufficiency of sigmoid, or pulmonary or aortic valves.
Mesodiastolic 	They are rhomboidal murmurs. They occupy the center of Diastole. They are produced by increasing flow through the auricular valves or in the stenosis of them.

Telediastolic  	They occupy the end of the diastole, they are usually increasing intensity, and are characteristic of mitral or tricuspid stenosis, coinciding with the contraction of the corresponding atrium.
Continuous  	They originate in the systole and surpass the S2 ending in diastole. There is a communication between an arterial vessel and another venous.

5.2.2.2. Mathematical model of cardiac murmurs

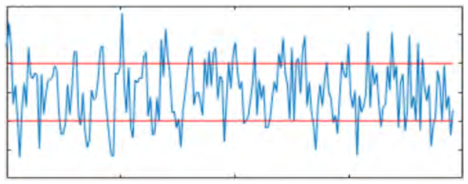
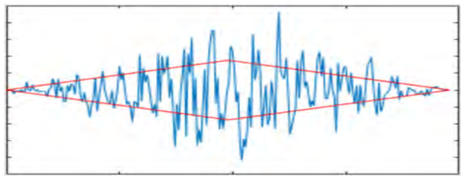
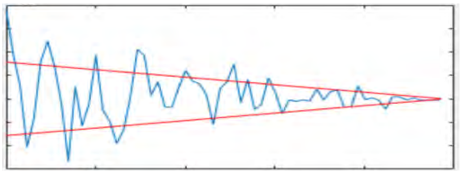
Taking into account the location and morphology of the different types of cardiac murmurs described in Table 30, and the model [23] that represents the generation of sounds S1 and S2, this section describes a mathematical model that we designed in order to generate a type of murmur. It is worth mentioning that this is a own model, which is not in the state of the art. The mathematical model is described below:

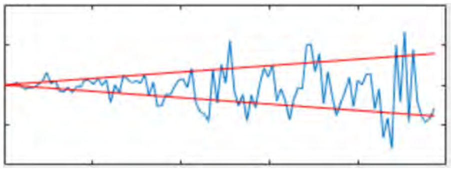
$$\delta(t) = \alpha_s \Delta(t)[cos(2\pi f t) + \beta] \qquad parat_i > t > t_f \qquad \mathbf{(24)}$$

Where α_s is a random variable that represents the amplitude of the murmur; β is a noise that is added with the sinusoidal signal; t_i and t_f are the initial and final points of the murmur respectively. $\Delta(t)$ is a triangular function, which varies depends on the type of murmur.

Table 31 describes the term $\Delta(t)$ for each type of murmur and an example of the generated signal. Where σ_s is a random variable that represents the width of the murmur, comprised between the systolic or diastolic interval, according to the type of abnormality (ejective or mesodiastolic); μ_s represents the centering of the murmur.

Table 31. Definition of the term $\Delta(t)$ in different types of cardiac murmurs

Type of cardiac murmur	$\Delta(t)$	**Signal**
Pansystolic	$\Delta(t)$ = 1	
Ejective	$\Delta(t) = \begin{cases} t - \frac{\sigma_s}{2} parat_i > t > \mu_s \\ -t + \sigma_s para\mu_s > t > t_f \end{cases}$	
Protosystolic or Protodiastolic	$\Delta(t) = t - \frac{\sigma_s}{2} parat_i > t > t_f$	

Telesystolic or Telediastolic	$\Delta(t) = -t + \sigma_s parat_i > t > t_f$	

Therefore, the new model to generate abnormal cardiac sounds is the sum of the model [23] and the model that represents the type of murmur, as denoted in the following equation:

$$HS\ Model = \dot{z} + \gamma. \qquad \mathbf{(25)}$$

Where $\dot{z}$ represents the model [23] and γ is described as follows:

$$\gamma = \begin{cases} 0 & \text{para } 0 < t < \sigma_{S1^+} \\ \delta(t) & \text{para } \sigma_{S1^+} < t < \sigma_{S2^-} \\ 0 & \text{para } \sigma_{S2^-} < t < 2\pi \end{cases} \qquad \mathbf{(26)}$$

Tables 32 and 33 present the parameters used to generate the different types of systolic and diastolic puffs respectively. A random variable was set in the amplitude parameter in order to provide variability in the waveform. Regarding the parameters of width and centered, the values were established taking into account the location of the sounds S1 and S2 (see table 27), the location of the different types of murmurs (see table 33), and that the duration of the signal is 2π. The frequency parameter is fixed with a value of 10, being an approximate value to the one used in the model [23]. Figures 27 and 28 shows examples of real and synthetic signals, in these figures it can be observed that the morphology and location of murmurs are similar.

Table 32. Parameters of systolic murmurs

Parameter	Pansystolic	Ejective	Protosystolic	Telesystolic
α_s (Amplitude)	Random variable with a range between 0.001 – 0.01			
σ_s (Width)	1.16 (Systolic interval)	Random variable with a range between $37\,\pi/100$ y $3\,\pi/50$		
μ_s (Centered)	$37\,\pi/100$	$37\,\pi/100$	$27\,\pi/100$	$\pi/100$
F (Frequency)	10			

Table 33. Parameters of diastolic murmurs

Parameter	Protodiastolic	Mesodiastolic	Telediastolic
α_s (Amplitude)	Random variable with a range between 0.001 – 0.01		
σ_s (Width)	Random variable with a range between $11\,\pi/10$ y $3\,\pi/50$		
μ_s (Centered)	$\pi/5$	$11\,\pi/10$	$9\,\pi/10$
F (Frequency)	10		

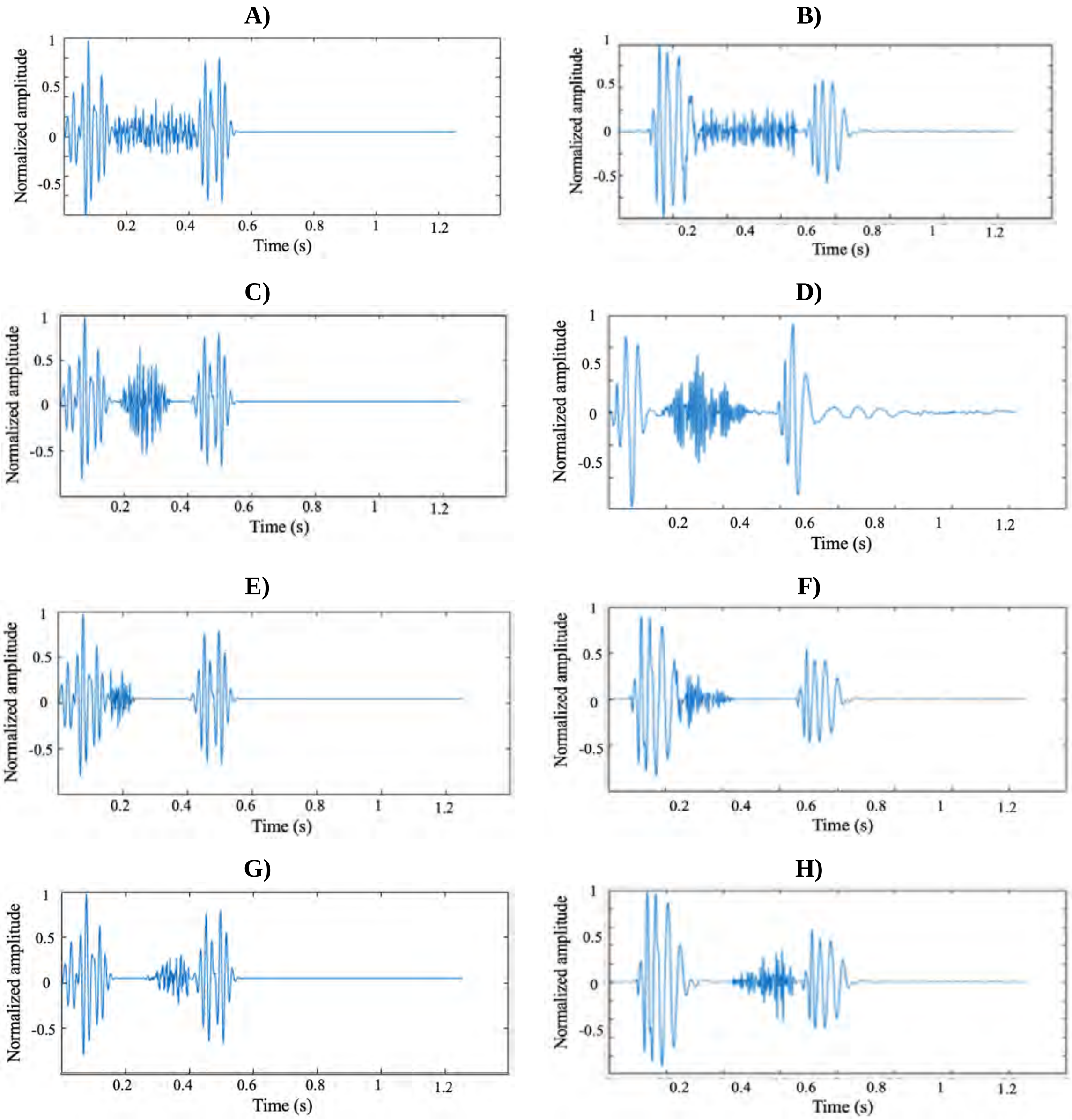
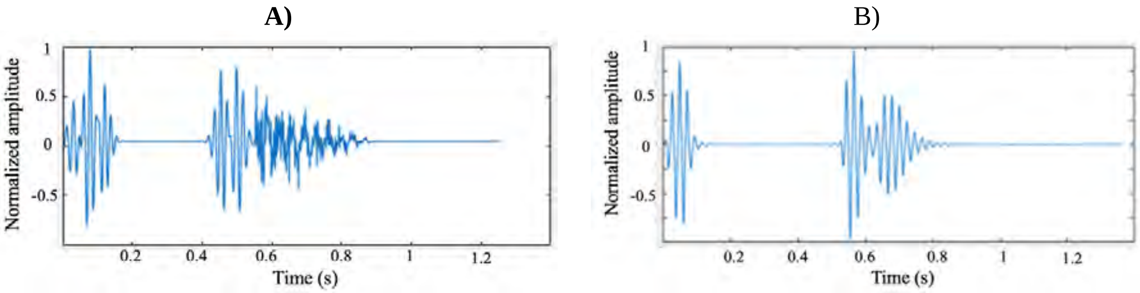


Figure 27. Examples of heart sounds with systolic murmurs. A) Pansystolic - Synthetic, B) Pansystolic - Real. C) Ejective - Synthetic; D) Ejective - Real; E) Protosystolic - Synthetic; F) Protosystolic - Real; G) Telesystolic - Synthetic; H) Telesystolic - Real.



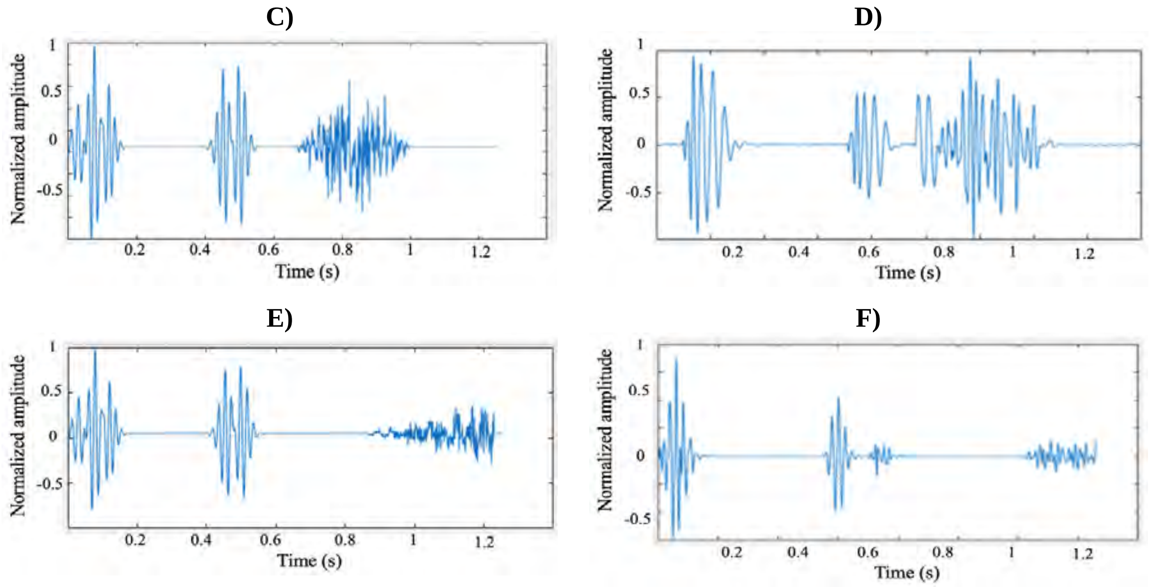


Figure 28. Examples of heart sounds with diastolic murmurs. A) Protodiastolic - Synthetic, B) Protodiastolic - Real. C) Mesodiastolic - Synthetic; D) Mesodiastolic - Real; E) Telediastolic - Synthetic; F) Telediastolic – Real.

5.2.3. FeaturesGAN using MFCC features

This section describes the proposed method of the FeaturesGAN model using MFCC characteristics of cardiac sounds. Figure 29 shows a general diagram of the architecture, in which the discriminant model is trained with signals representing the concatenation of an ideal cardiac sounds (obtained with the mathematical model) and MFCC features vector extracted from a dataset of real signals. The training methodology of this model is very similar to that presented in section 5.1.1. (See Figure 16), including hyperparameter configurations such as: optimization and loss functions.

The main objective of this architecture is to ensure that the generating model is capable of producing signals that are very similar in time domain to signals generated by the mathematical model, and in turn have MFCC features very similar to real signals.

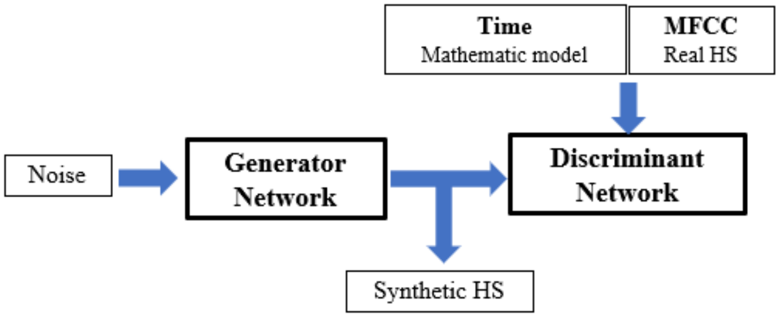


Figure 29. General architecture of FeaturesGAN using MFCC features

Figure 30 shows the architecture of the generator for this experiment. The hyperparmeters used in the dense and convolutional layers are the same proposed in section 5.1.1. (Figure 17). However, MFCC coefficients at the signal obtained in the last dense layer are calculated in this experiment, and after both signals are concatenated, as shown in Figure 11. For the extraction of the MFCC coefficients, the signal was segmented using windows of 64 ms with 75% overlap, 14 MFCC coefficients were calculated in each window. Considering that the size of the signal is 2000 samples, the MFCC vector was 182 samples.

Considering that the goal of the generator is to deceive the discriminator by producing signals that have MFCC features very similar to the real signals, it is decided to replicate and concatenate the vector of features 6 times so that there is a balance between the signal size in the domain of the Time (2000 samples) and the MFCC feature vector (1092 samples). In this way, as the training epochs increase, the generator enters mode collapse and begins to produce signals very similar to those used in the discriminator training dataset.

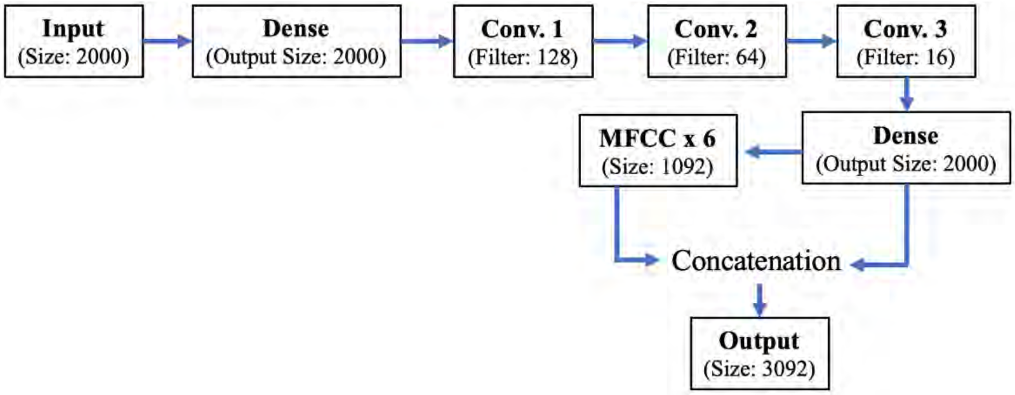


Figure 30. Architecture of the generator model for FeaturesGAN using MFCC features.

Figure 31 describes the architecture of the discriminant model; the configurations of each layer are similar to the proposals in section 5.1.1. (Figure 18), the only difference is that the input signals have a size of 3098 samples.

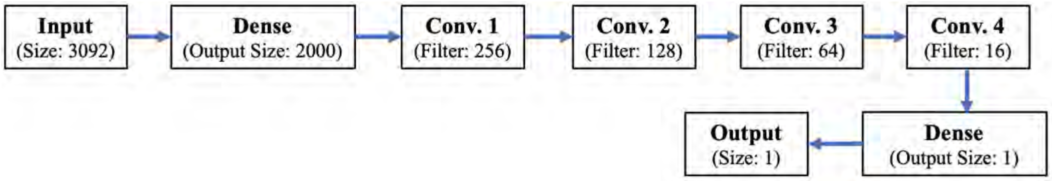


Figure 31. Architecture of the discriminant model for FeaturesGAN using MFCC features.

Experiments were carried out using normal and abnormal heart sounds. The set of normal heart sound data was obtained from Physionet [16], 50 cardiac cycles were collected with similar cardiac frequencies, that is, the systolic and diastolic segments have similar durations in each of the samples. Additionally, 50 normal heart sounds were generated using the mathematical model [23], the parameters were also configured so that the heart rate is similar to the real signals.

In Figure 32A there is a signal that represents the concatenation between a normal cardiac sound obtained with the mathematical model (left) and the MFCC features vector extracted from the same signal (right), in Figure 32B the same type of concatenation is shown but with a real heart sound. It can be observed that the MFCC features vector differ a lot between both signals, one of the main reasons is the absence of noise in the systolic and diastolic interval of the signal generated with the mathematical model. Figure 32C shows the concatenation of the signal obtained by the mathematical model and the MFCC features vector of the real signal, this signal is an example of the data set used for discriminator training. Finally, Figure 32D shows the result of the generator output after 5000 epochs of training, it can be seen that the signal in the time domain (left) was refined with real MFCC features.

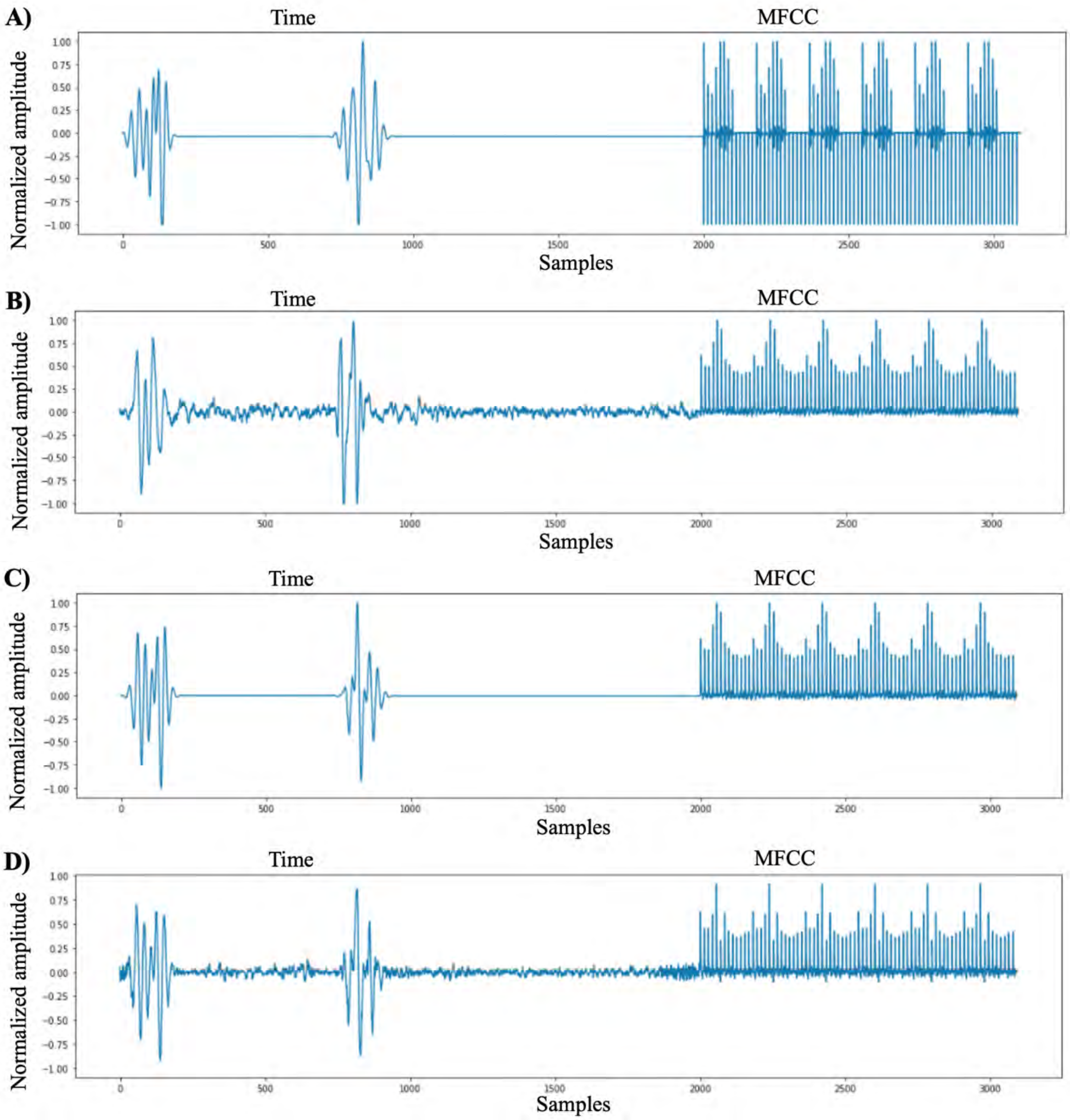


Figure 32. Examples of cardiac signals concatenation in the time domain and their respective MFCC feature vector. A) Signal obtained from the mathematical model, B) Real signal obtained from a database, C) Signal used for discriminator training, D) Signal obtained after 5000 training epochs.

Figures 33A, 33B and 33C shows the result in the domain of the frequencies of the signal obtained by the mathematical model, of the actual cardiac signal and the signal generated by the FeaturesGAN model respectively. Similarly, refinement can be observed in the high frequency range of the synthetic signal obtained by FeaturesGAN.

The same experiment was repeated for abnormal heart sounds. For this case, 50 cardiac sounds of the Physionet database [16] that have an ejective systolic murmur and with a similar heart rate were also selected. Additionally, abnormal cardiac sounds were generated from the mathematical model proposed in section 5.2.2.1., taking into account the parameters required for a signal with the same type of murmur and heart rate.

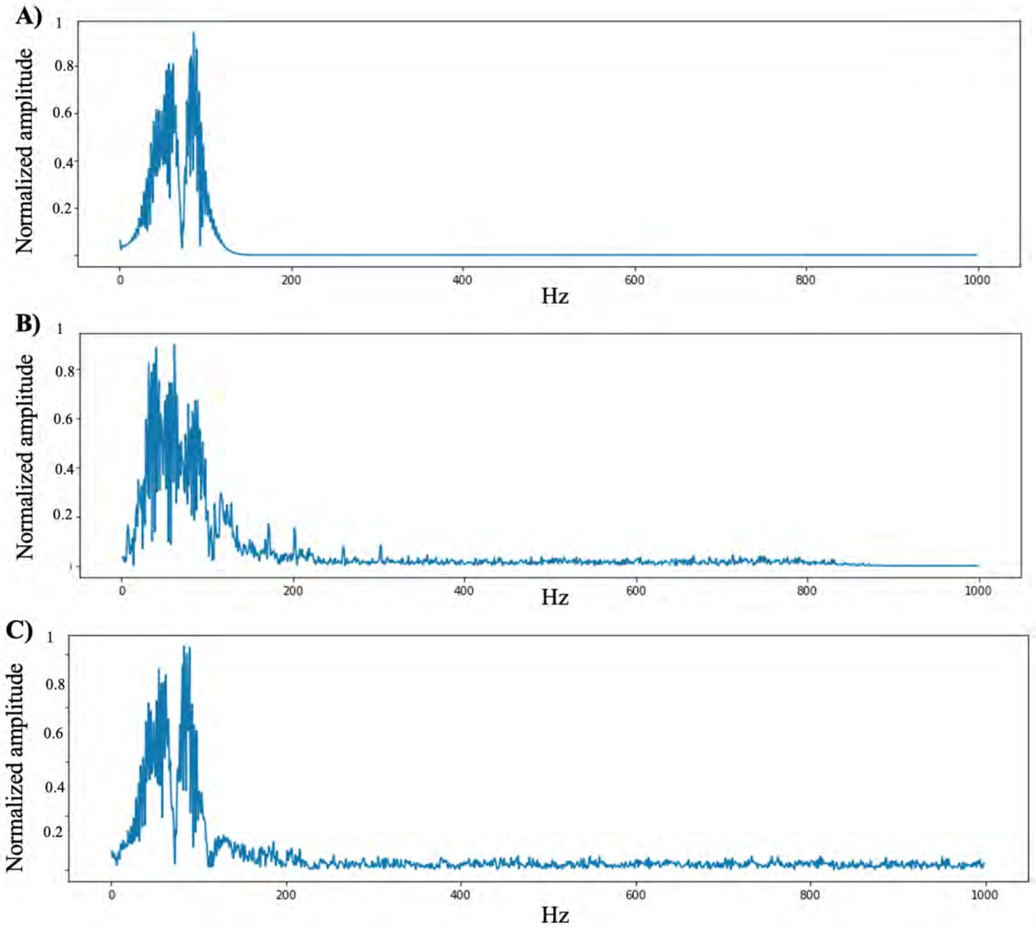


Figure 33. Result of cardiac signals in the frequency domain. A) Normal heart sound obtained from the mathematical model, B) Real signal, C) Synthetic signal obtained from the proposed FeaturesGAN model.

Figure 34 shows representations of the concatenation between signals with systolic murmur in the time domain and its respective MFCC features vector, very similar to what is illustrated in Figure 33 with normal cardiac sounds. Figure 34D shows the result of the generator output after 5000 epochs of training. Therefore, it can be noted that the resulting signal has MFCC features very similar to the real signal (Figure 34B) compared to those signals obtained by the mathematical model.

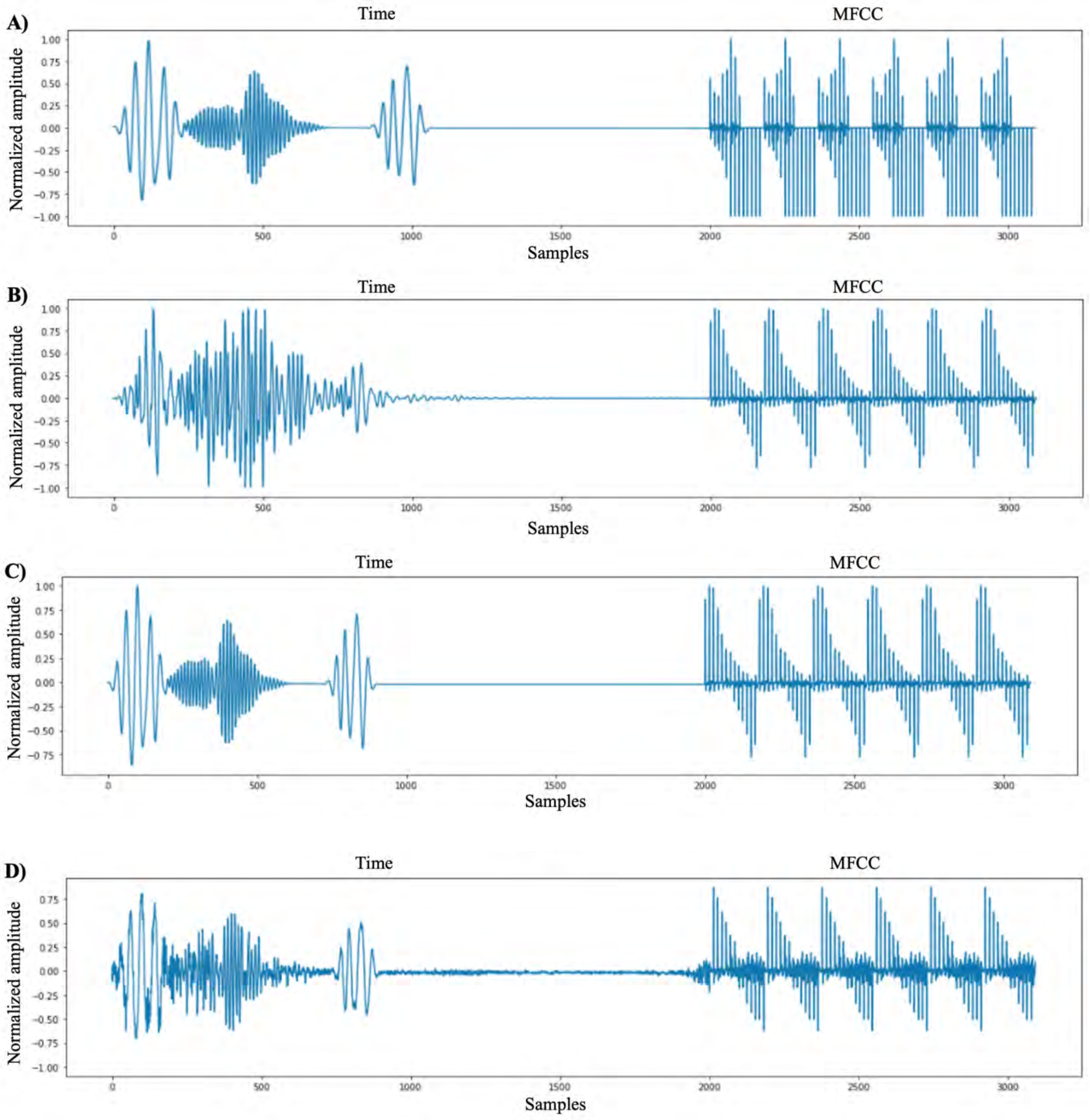


Figure 34. Examples of abnormal cardiac signals concatenation in the time domain and their respective MFCC feature vector. A) Signal obtained from the mathematical model, B) Real signal obtained from a database, C) Signal used for discriminator training, D) Signal obtained after 5000 training epochs.

Likewise, Figure 35 shows the results in the frequency domain for the signal obtained by the mathematical model (Figure 35A), the real signal (Figure 35B) and the signal obtained by the FeaturesGAN model (Figure 35C). Although a change is observed in the resulting signal from the generator with respect to the signal obtained by the mathematical model, the FFT signal does not have a behavior similar to real signals. Therefore, it is very important to consider the characteristics in the mastery of the frequencies for the refinement of these synthetic signals.

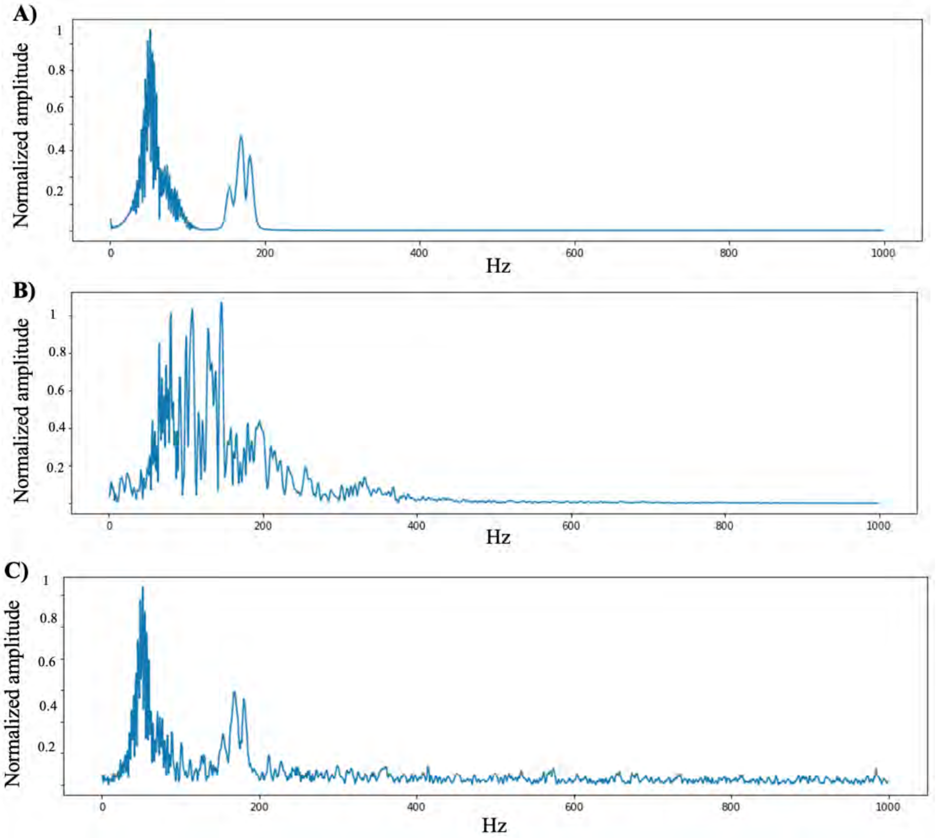


Figure 35. Result of abnormal cardiac signals in the frequency domain. A) Abnormal heart sound obtained from the mathematical model, B) Real signal, C) Synthetic signal obtained from the proposed FeaturesGAN model.

5.2.4. FeaturesGAN using MFCC and FFT features

Considering the limitations that were presented in the results of FeaturesGAN using only MFCC features for the synthesis of heart sounds with murmurs, in this section we present the same experiment using MFCC and FFT features in the FeaturesGAN model. Figure 36 shows a general diagram of the new FeaturesGAN architecture, unlike the previous experiment, in this case an FFT feature vector is added to the concatenation of signals used for discriminator training. Similarly, all the configurations and hyperparameters used in the previous experiments are preserved.

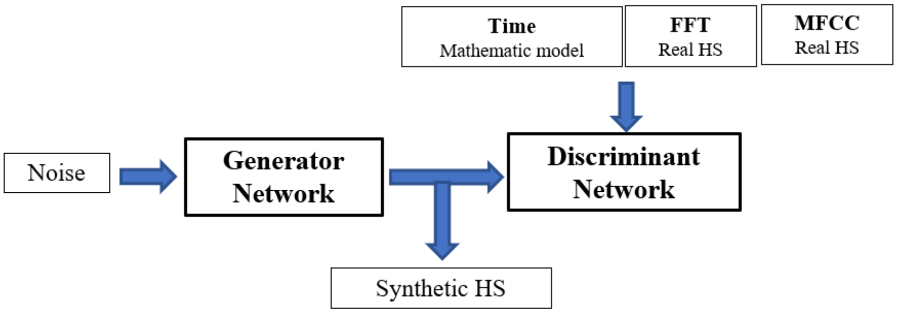


Figure 36. General architecture of FeaturesGAN using MFCC and FFT features

Figure 37 shows the architecture of the generator proposed in this experiment. In this case, the FFT and MFCC features are calculated at the output of the last dense layer. Subsequently, the 3 signals are concatenated, that is, the output of the last dense layer, the FFT signal and the MFCC feature vector. Similarly, the MFCC feature vector is replicated 6 times to a size of 1098 samples and the FFT signal is doubled to a size of 2000 samples. Therefore, the size of the resulting signal is 5098 samples. The dense and convolutional layers have the settings and hyperparameters used in the previous experiments.

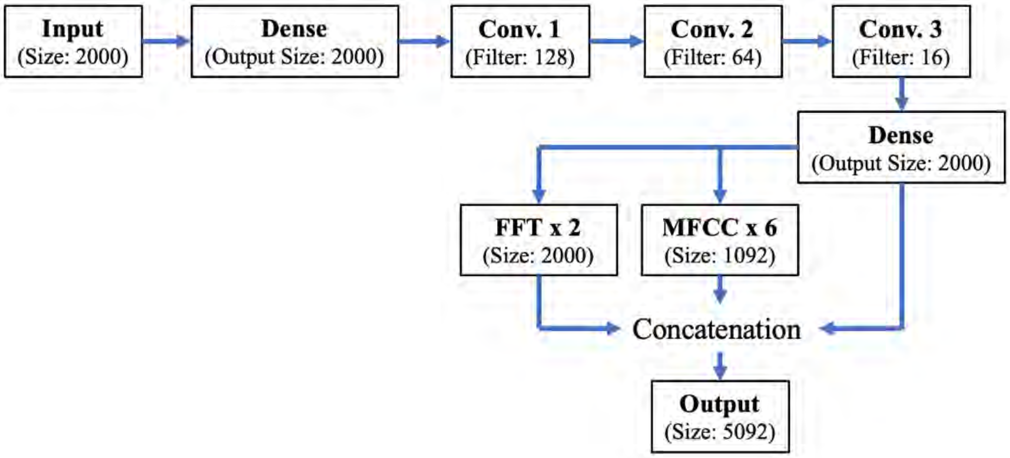


Figure 37. Architecture of the generator model for FeaturesGAN using MFCC and FFT features.

For the case of the Discriminator model, Figure 38 shows the diagram of its architecture. The settings and hyperparameters are similar to those used in section 5.1.1. (Figure 17) and section 5.2.3. (Figure 31), the only difference is that the size of the input signals is 5098 samples.

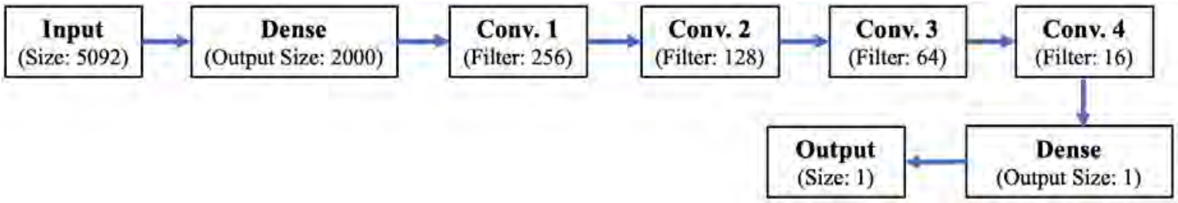


Figure 38. Architecture of the discriminant model for FeaturesGAN using MFCC and FFT features.

Based on the fact that the sounds S1, S2 and murmurs present different frequency components and, in this experiment, the FFT signal is used in the training of the model, it is considered to perform a murmur synthesis separate of the sounds S1 and S2, as shown in Figure 39. Therefore, a data set of 50 signals is constructed, in which each signal is composed of a murmur obtained from the mathematical model, an FFT signal (duplicated) and an MFCC feature vector (replicated 6 times) obtained from the murmur of a real signal, that is, in each real signal the sounds S1 and S2 were suppressed, in order to analyze only the murmur. After training for 10,000 epochs, the Generator produces a signal like the one shown in Figure 39B. Therefore, the synthetic murmur presents FFT and MFCC features very similar to those obtained with the murmurs of the real signals.

For the generation of sounds S1 and S2, the same method proposed in the previous section was used, considering that the parameters of the mathematical model [23] are good indicators for a representation of the signal in the frequency domain.

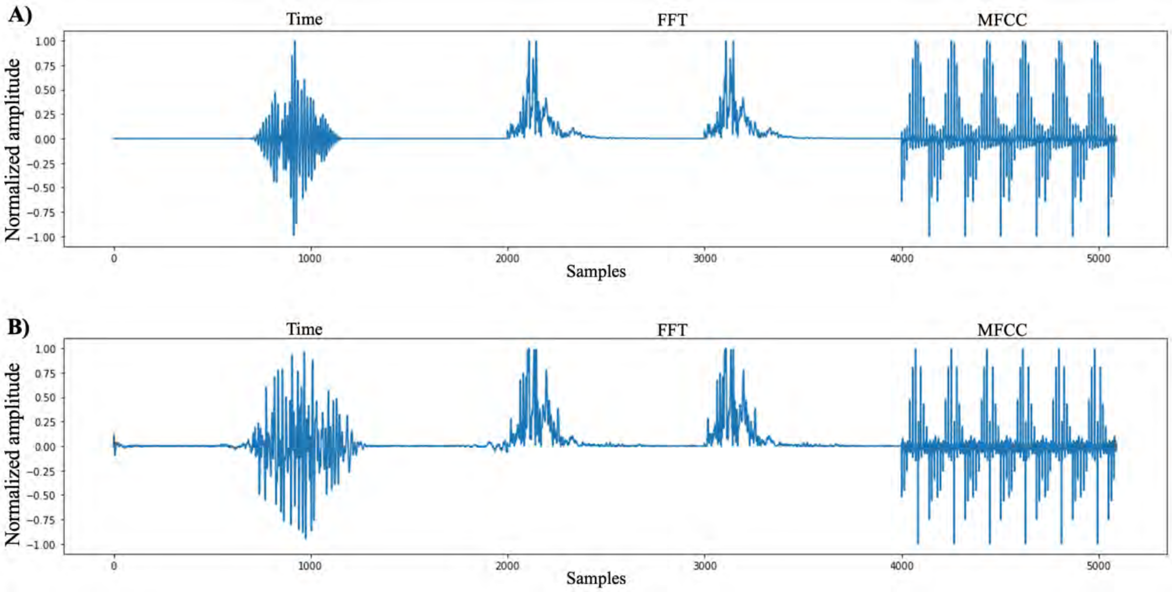


Figure 39. Examples of abnormal cardiac signals concatenation in the time domain and their respective FFT and MFCC feature vector. A) Signal used for discriminator training, B) Signal obtained after 10000 training epochs.

Figure 40 shows results of heart sounds with ejective murmur in systole in the time and frequency domain. The result obtained by the FeaturesGAN model generated a significant change in the FFT signal, showing a behavior very similar to a real signal.

A more rigorous analysis of the synthetic signals obtained from the FeaturesGAN model will be presented in the next sections.

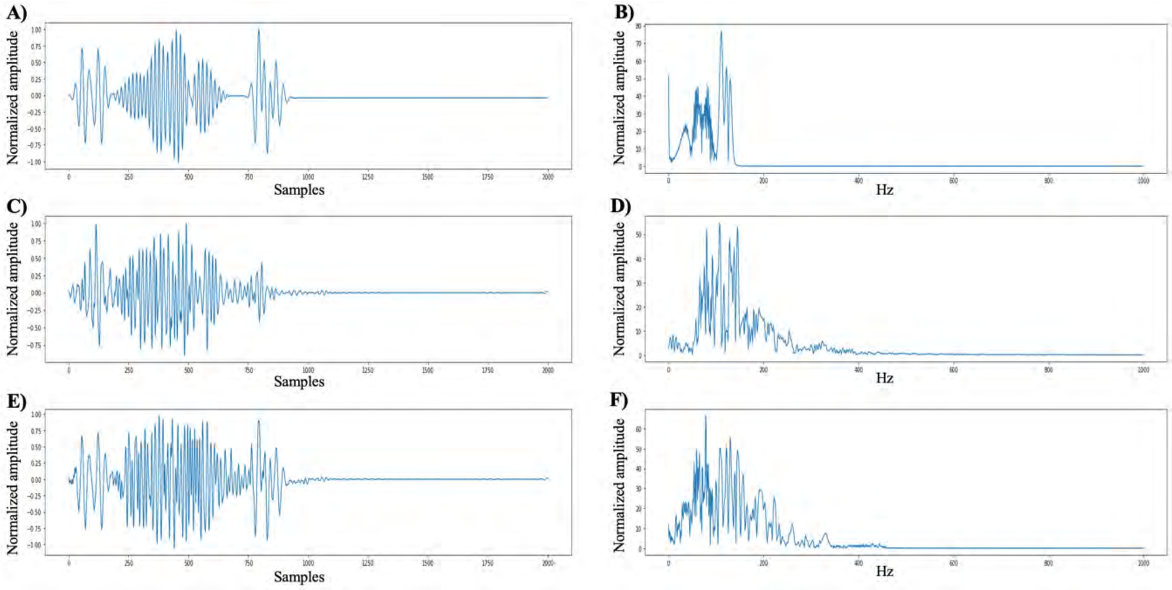


Figure 40. Result of abnormal signals in the time and frequency domain. A) Abnormal heart sound obtained from the mathematical model, B) Real signal, C) Synthetic signal obtained from the proposed FeaturesGAN model.

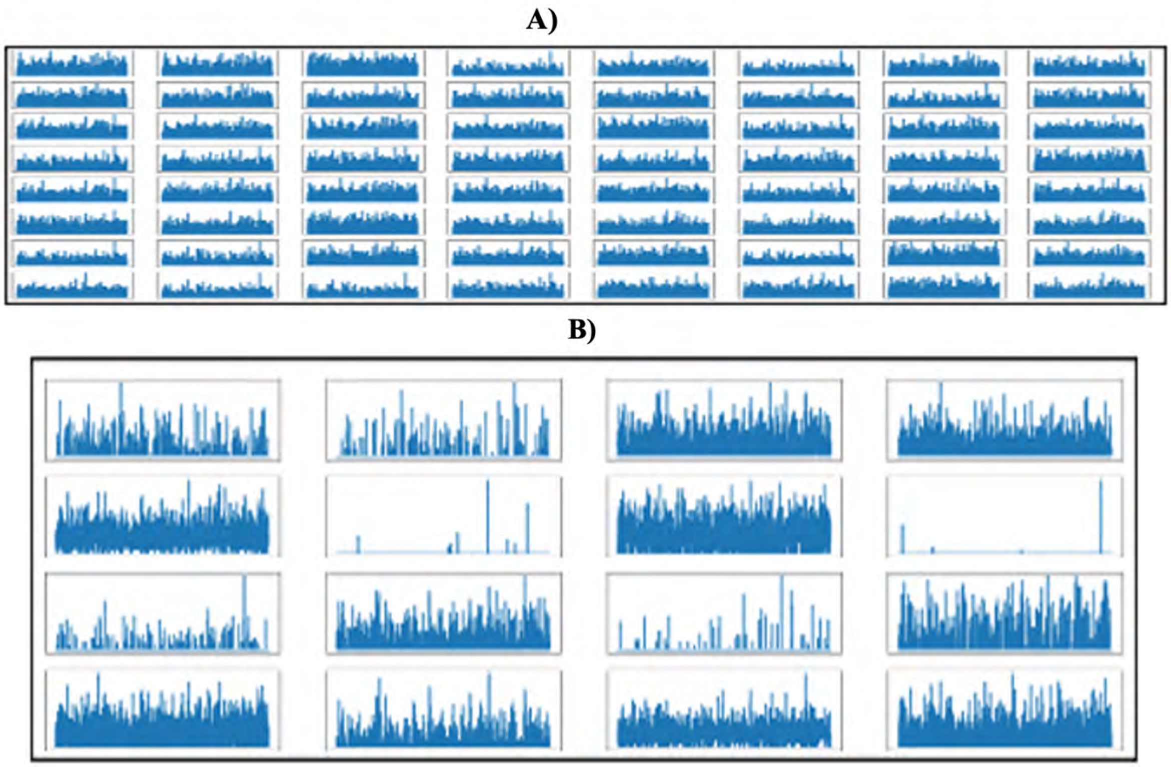
5.2.5. Analysis of results

This section presents different analyses of synthetic signal results obtained with the proposed FeaturesGAN model. It begins with a review of the Generator and Discriminator model in the training stage. Subsequently, different ways are proposed to evaluate the quality of synthetic signals in an objective way, for which dimension reduction techniques such as PCA and t-SNE are used in order to visualize the real and synthetic signals in a 2D plane, and in turn analyze the clustering between them. On the other hand, widely used methods for the objective evaluation of audio and image quality such as MCD and SSIM, respectively, are implemented; the performance of different heart sound classification models using synthetic signals is also evaluated, and finally, MOS tests with medical experts in murmur identification to obtain a subjective evaluation of the quality of the sounds.

5.2.5.1. Analysis of Generator and Discriminant models

This section presents an analysis of the Generator and Discriminator models in the training process. The objective is to visualize what happens in each of the model layers to understand the mode collapse phenomenon in greater detail. On this occasion, the analysis is performed from the generation of a normal heart sound based on the GAN architecture proposed in section 5.1.1.

Initially, a display is made of the output of each of the Generator layers (convolutional and dense), the discriminator output, and the Generator and discriminator losses at certain training epochs. Figure 41 shows the results of feature maps obtained in various layers of the Generator model after 2000 training epochs. In the fourth convolutional layer (see Figure 41-D), it can be seen that the signal samples are close to zero. Therefore, the last dense layer is receiving a vector of zeros after a number of epochs. However, as the iterations increase, the output of this dense layer is a signal very similar to that of a heart sound. For this reason, it is decided to visualize the weights and biases of the dense layer at certain instants of epochs, in order to analyze the behavior of this layer.



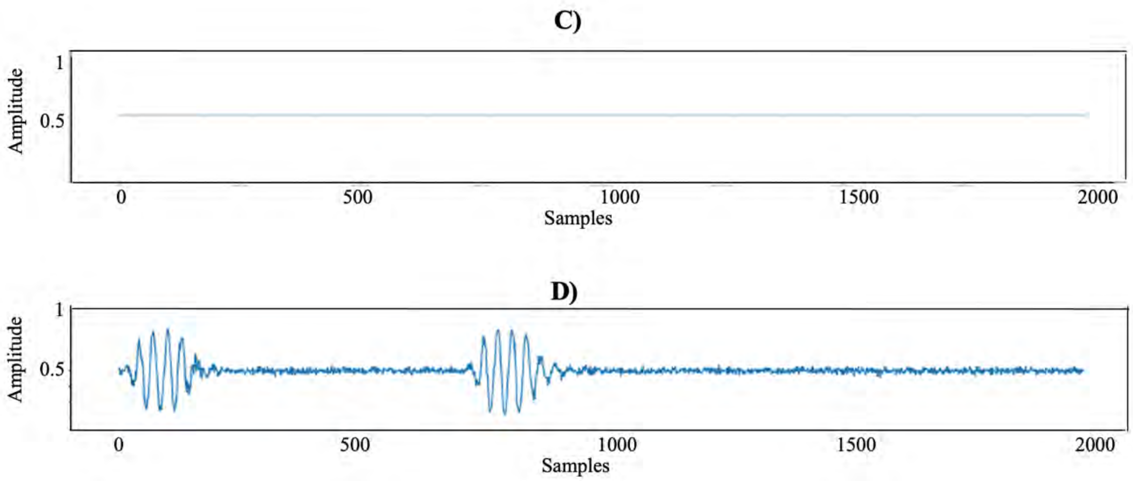


Figure 41. Features map of the generating model with 2000 training epochs. A) First convolutional layer with 128 filters (only the first 64 maps are shown); B) Second convolutional layer with 64 filters (Only the first 16 maps are displayed); C) Third convolutional layer with 1 filter; D) Dense layer with output of 2000 samples.

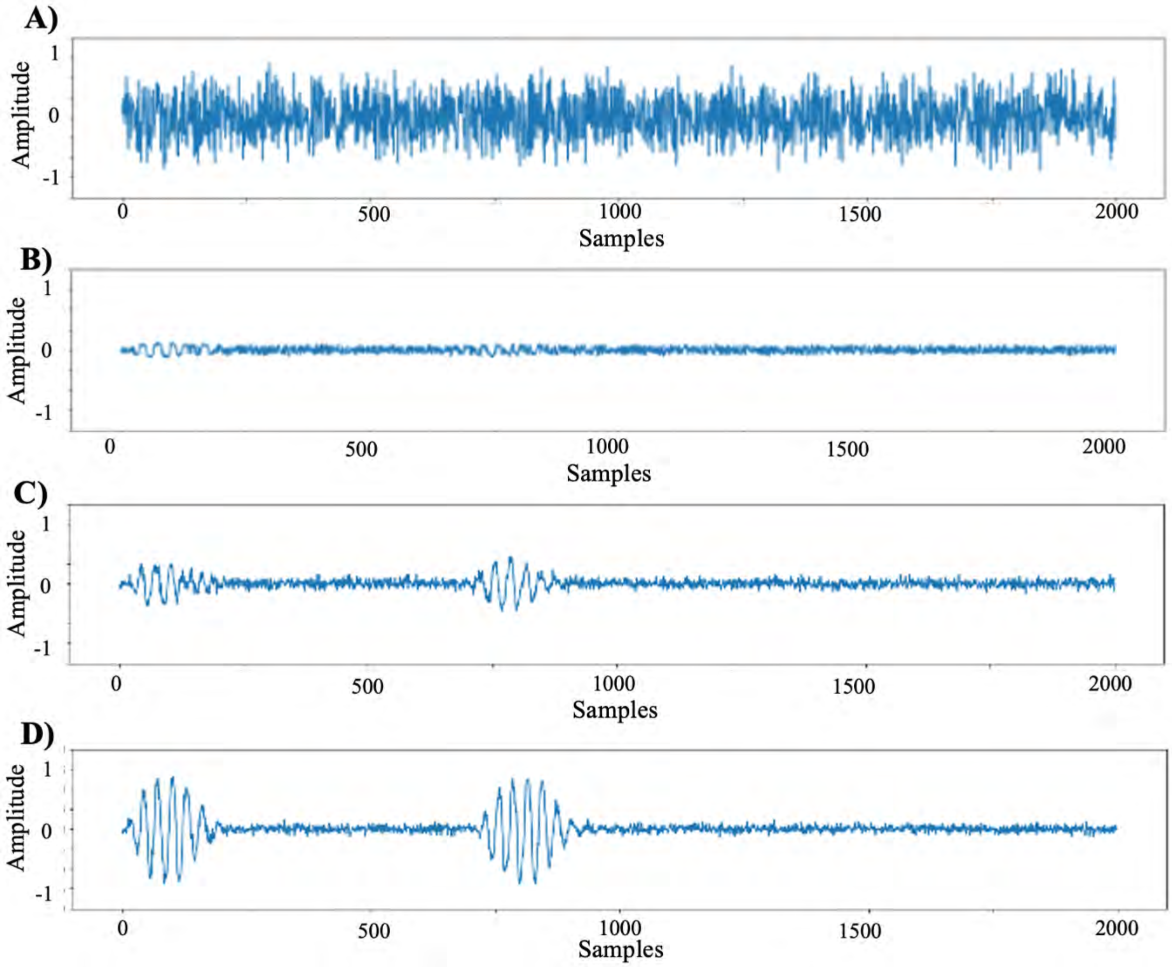


Figure 42. Bias visualization in the last dense layer in different training epochs. A) 10 epochs; B) 100 epochs; C) 500 epochs; D) 2000 epochs.

Taking into account that after a number of iterations the weights of the dense layer begin to multiply with values close to zero, it is concluded that the biases are the values that are adjusted to obtain a signal that allows the discriminator to be fooled. Figure 42 shows the bias result at different epochs, as can be seen, the bias values after 1000 epochs represent the shape of a real heart sound. Therefore, the convolutional layers would not be contributing relevant information to the model, unlike the dense layer, mainly the biases, which are continually adjusting their weights to generate a sufficiently real signal.

A possible reason why the model presents this behavior is the use of few real samples to train the discriminator. In this way, the model goes into mode collapse quickly and always tries to find the same path so that the generator can fool the discriminator. Figure 43 shows the values of discriminator loss (blue line), generator losses (red line), and discriminator output (green line) at different epochs. An oscillatory behavior can be observed in the signals, this being a very common case when the mode collapse is present in the training stage.

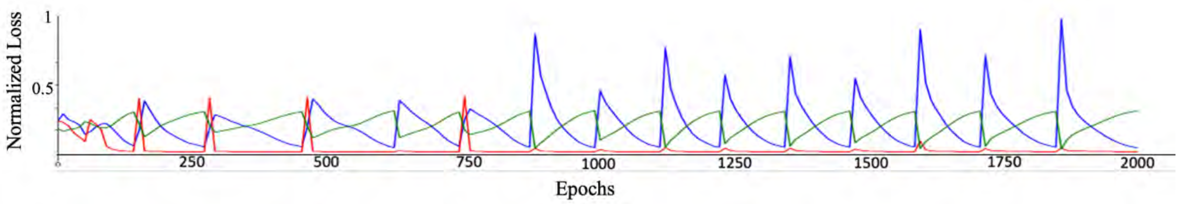


Figure 43. Discriminator output and losses. Green signal: Discriminator output; Red signal: Discriminator loss; Blue signal: Generator loss

Taking into account that the signals are generated from the biases of the last dense layer, and failing that, the convolutional layers are not providing the expected characteristics, it is decided to modify the Generator model using a partially connected neural network. In this case, the input of the model receives a noise signal that is multiplied with the weights of the network and then added to the signal of the mathematical model. In this way, the execution time is reduced to 80%, since it does not use convolutional layers and the neural network is not fully connected. By incorporating the deterministic signal (mathematical model) in the new generator model, the network approach becomes a signal refiner, that is, FeaturesGAN is a refiner of synthetic signals that were obtained by mathematical models. The architecture is described in figure 44.

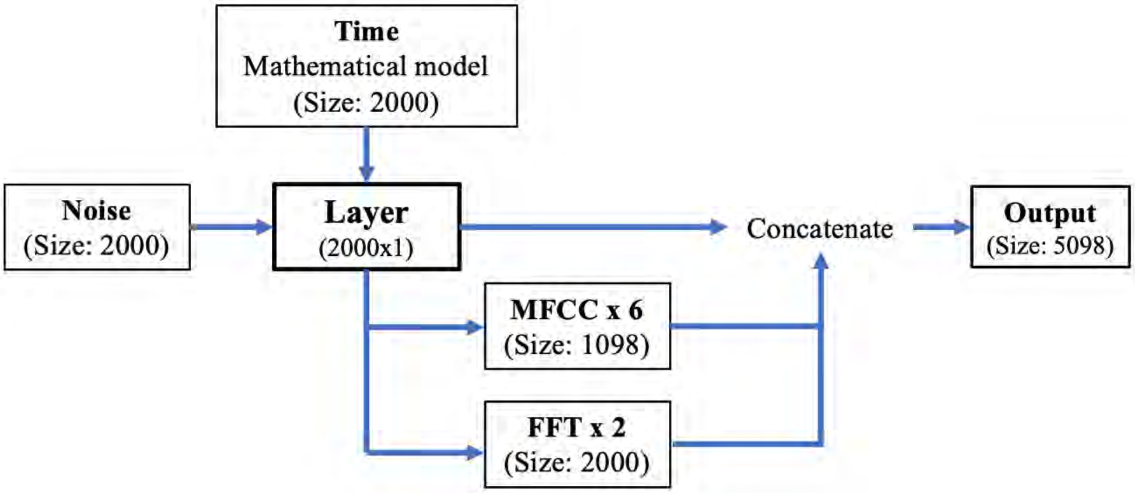


Figure 44. Diagram of the modified generator model

Similar to the analysis of the Generator model, the feature maps of each convolutional layer were analyzed at different instants of epochs for the Discriminator model. Figure 45 shows examples of various layers of the network after 2000 training epochs, in which no significant findings or patterns were found. The outputs of the discriminator were also visualized at different times, in order to analyze how the performance of the discriminator behaves as the generating model adjusts its weights. It is worth mentioning that the real signals were labeled with the value of 1. In Figure 46 presents the results of the discriminator for 50 signals obtained from the generator at different epochs, in which it is observed that the output of the discriminator approaches 1 as the epochs increase.

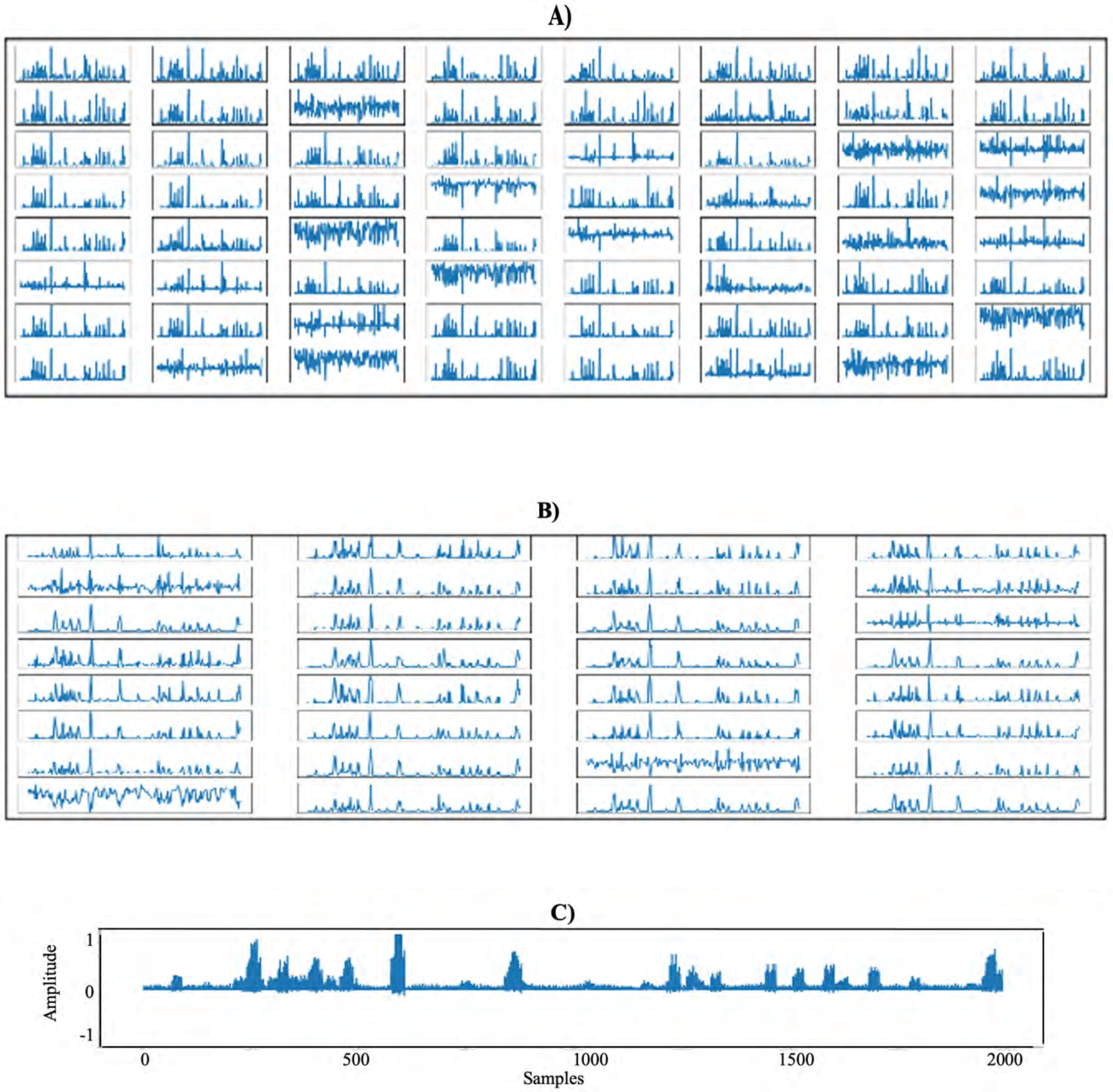


Figure 45. Features map in the discriminator layers. A) Third convolutional layer with 64 filters; B) Fourth convolutional layer with 32 filters; C) Convolutional layer with 1 filter

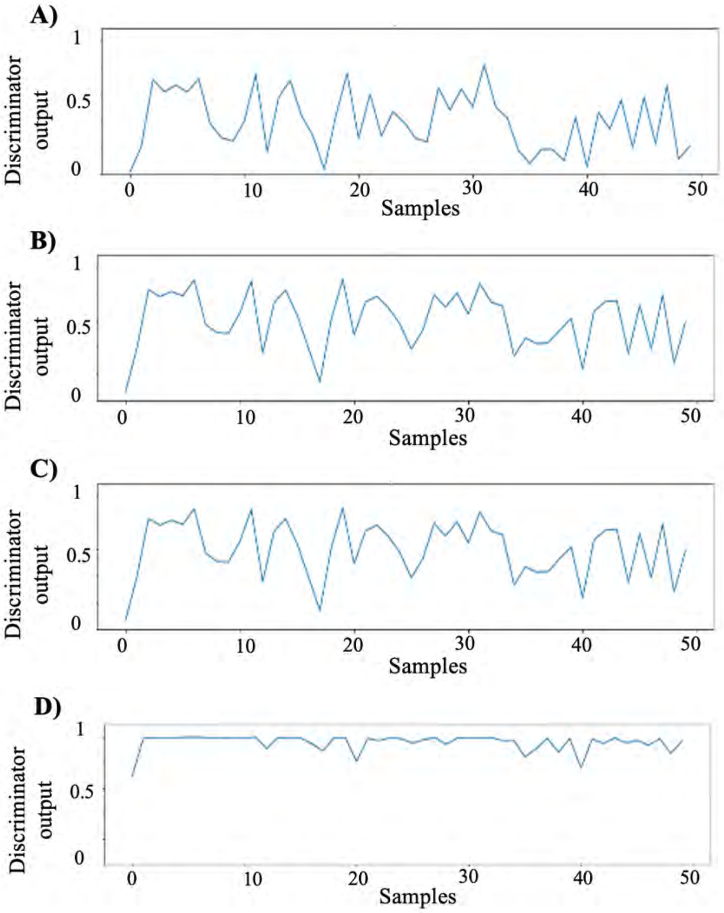


Figure 46. Discriminator outputs using 50 generated samples as the test data set in different training epochs. A) 10 epochs. B) 100 epochs. C) 500 epochs. D) 2000 epochs.

5.2.5.2. PCA and t-SNE

An analysis of the results was made from the visualization of features extracted using Principal Component Analysis (PCA) and T-distributed stochastic neighbor embedding (t-SNE), both methods are widely used for dimension reduction. PCA consists of expressing a set of data or characteristics in a set of linear combinations of uncorrelated factors [157]. On the other hand, t-SNE creates a probability distribution that represents the similarities between neighboring data in a high-dimensional space and in a lower-dimensional space [158]. Different experiments have been carried out in the literature to compare which of the two methods has a better performance, in many cases t-SNE shows a better clustering result compared to PCA [159]. Both methods will be used for the analysis of real and synthetic heart sounds.

The objective is to use PCA and t-SNE to reduce the dimension of the signals and to be able to do an analysis in a 2D plane. In this way, the clustering of the different types of signals is verified, such as: real signals obtained from different databases, synthetic signals obtained by the mathematical model and synthetic signals obtained by the FeaturesGAN model. Several analysis experiments using normal and abnormal cardiac signals are presented below.

Experiment 1: In this experiment, 900 signals representing normal heart sounds were used and are distributed as follows: 300 real signals, 300 signals obtained by the mathematical model [23], and 300 signals obtained by the FeaturesGAN model using MFCC features. PCA and t-SNE were applied to the dataset to reduce its dimensions into two components. Figure 47A and 47B show the results of PCA and t-SNE, respectively, where the green dots represent the real signals, the blue dots are the signals obtained by the FeaturesGAN model, and the red dots are the signals generated by the mathematical model. Figures 47C, 47D and 47E show an example of each of the signal types used in this analysis.

A very pronounced clustering of the real signals (green dots) and a high similarity between the synthetic signals (blue and red dots) can be observed. It is worth mentioning that the sounds S1 and S2 were obtained with the model proposed in [23], however, a clustering between the real and synthetic signals is not observed, implying that there are characteristics that differ between them. Given this situation, it is proposed to perform a refinement of the sounds S1 and S2 generated by the model [23] using the FeaturesGAN method with MFCC and FFT features. The procedure is similar to that used in murmur generation as described in section 5.2.4.

Figure 48 shows examples of sounds S1 in the time and frequency domain, figures 48A, 48B and 48C refer to the signal S1 obtained by the model [23], the segment S1 obtained from a real signal and the signal resulting from the FeaturesGAN model respectively. In Figure 48C, a change in the frequency domain of the synthetic signal can be observed since FFT features were used in the FeaturesGAN model.

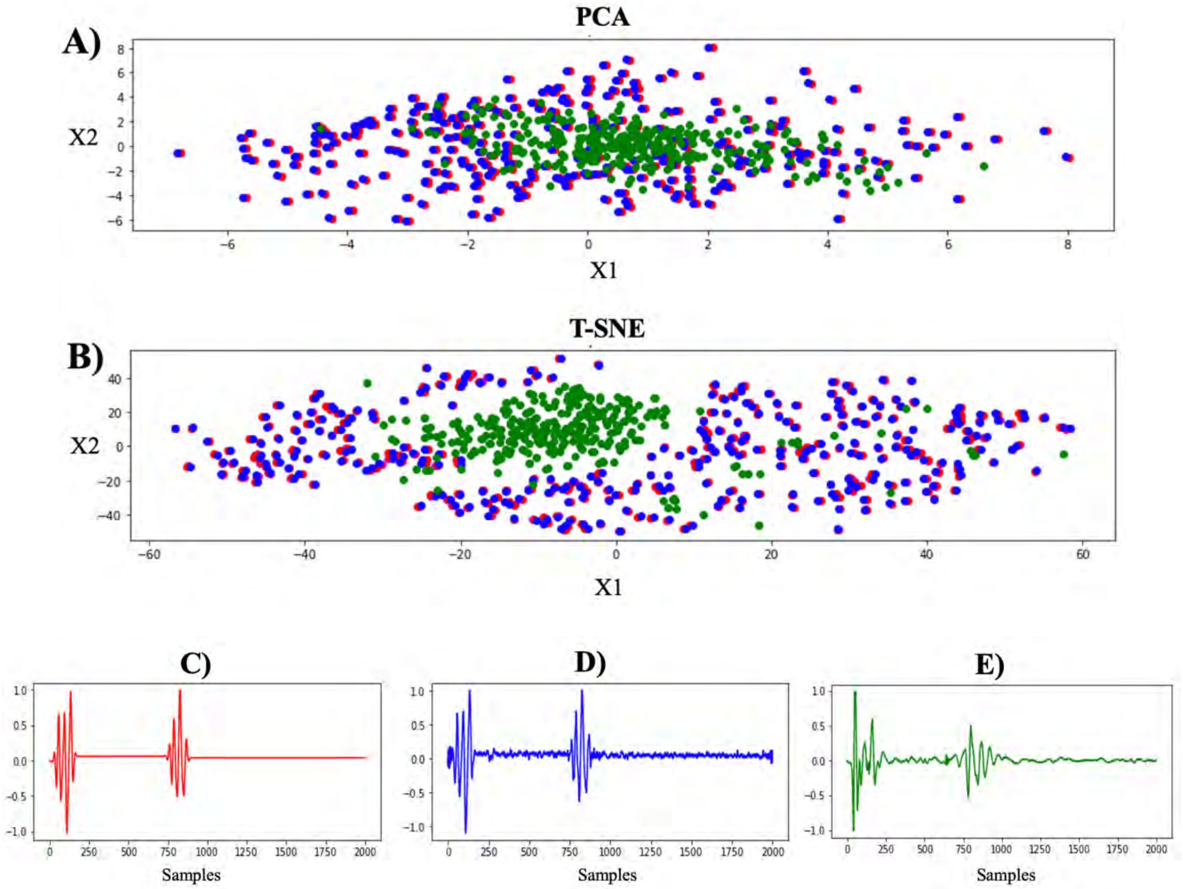


Figure 47. A) Clustering result using PCA; B) Clustering result using T-SNE; C) Signal obtained from the Mathematical Model; D) Real Cardiac Signal; E) Signal obtained from the FeaturesGAN model using MFCC. The Y-axis represents the normalized amplitude of the signal.

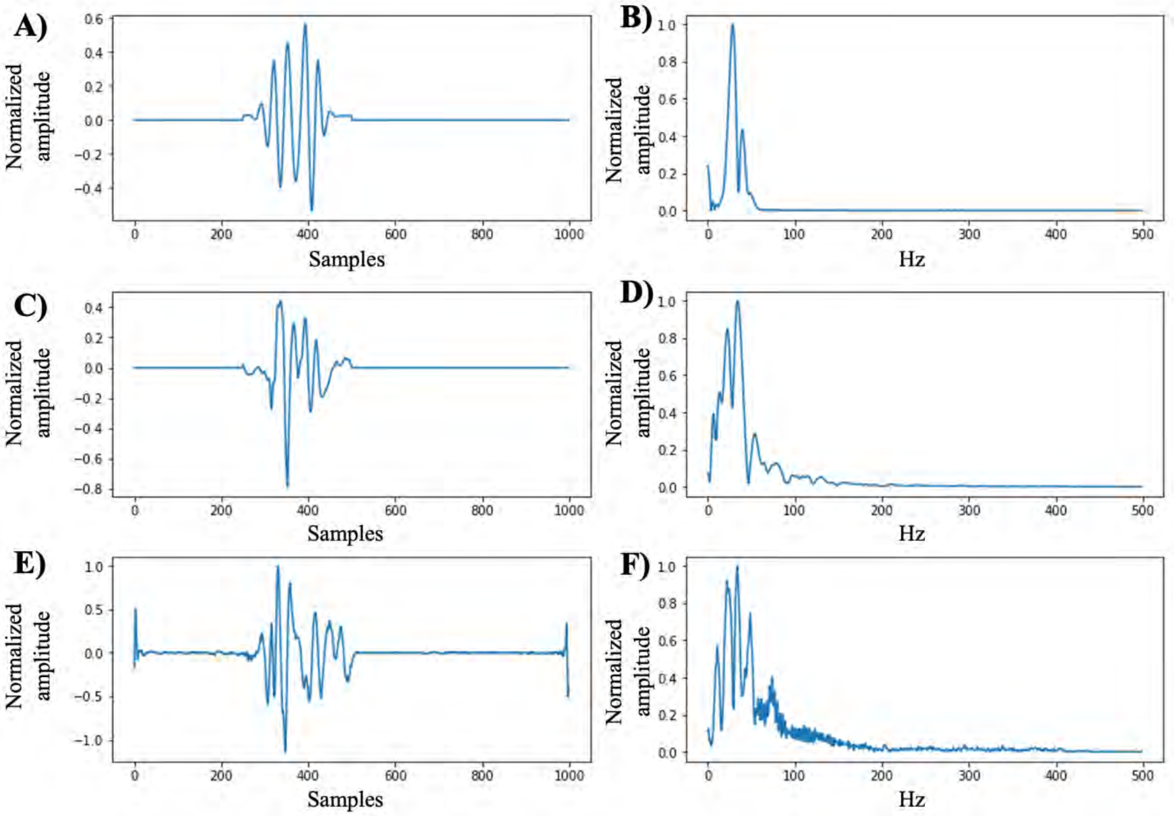
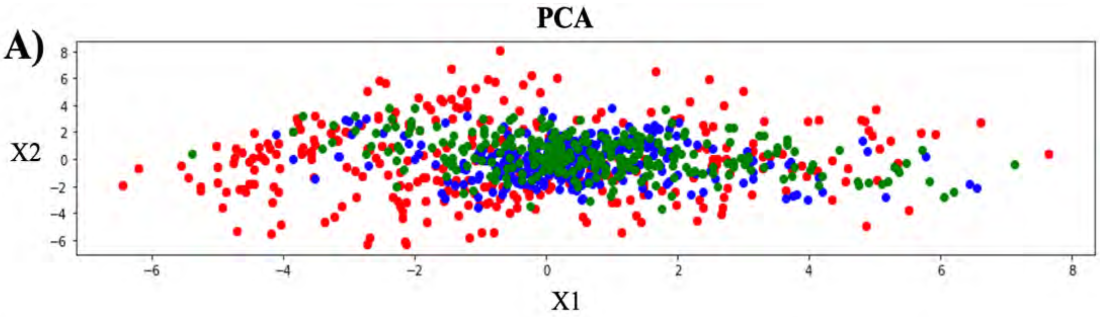


Figure 48. Sound S1 results using FeaturesGAN with MFCC and FFT features. Sound S1 obtained by the mathematical model: A) Signal in the time domain; B) Signal in the frequency domain. S1 sounds extracted from a real cardiac signal: C) Signal in the time domain; D) Signal in the frequency domain. Sound S1 generated by FeaturesGAN: E) Signal in the time domain; F) Signal in the frequency domain.

After refining the sounds S1 and S2 using FeaturesGAN, the same visualization experiment using PCA and t-SNE is performed on the same dataset. Figure 49A and 49B show the results of PCA and t-SNE respectively. A clustering can be observed between the real signals (green dots) and the synthetic signals generated by FeaturesGAN (blue dots), unlike the signals generated by the model [23]. It is also observed that the t-SNE method performs a more pronounced clustering than PCA. These results show that the adjustment made in the sounds S1 and S2 with the FeaturesGAN model allows the generation of a cardiac signal with more realistic characteristics, unlike using only the mathematical model. Figures 49C, 49D and 49E show an example of each of the signal types used in this analysis.



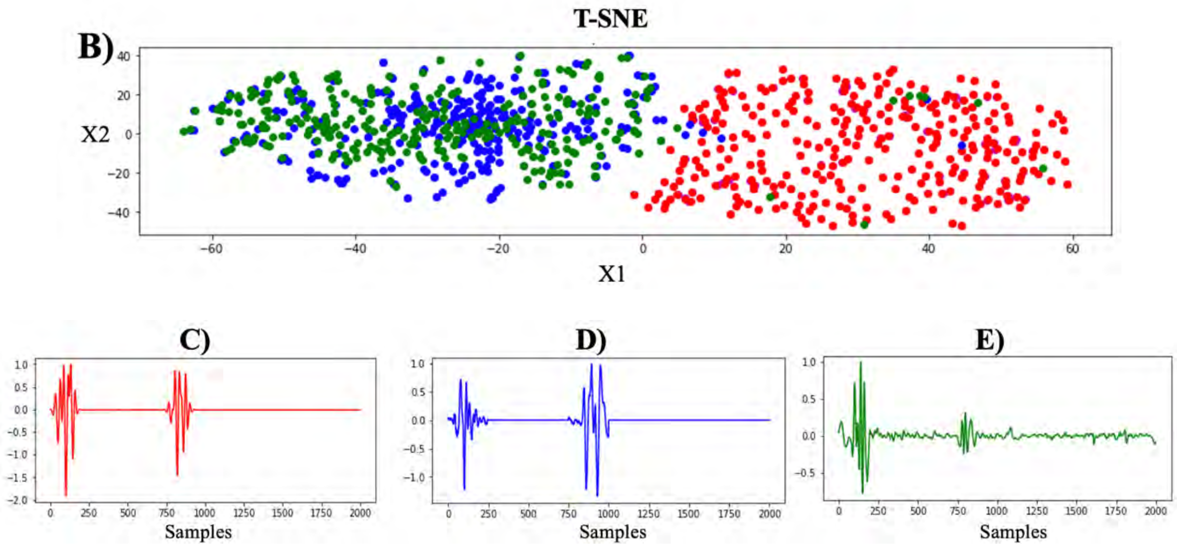


Figure 49. A) Clustering result using PCA; B) Clustering result using T-SNE; C) Signal obtained from the Mathematical Model; D) Real Cardiac Signal; E) Signal obtained from the FeaturesGAN model using MFCC and FFT. The Y-axis represents the normalized amplitude of the signal

Experiment 2: In this second experiment an analysis of the signals generated by FeaturesGAN is done using MFCC features. 200 signals representing the concatenation of a heart sound in the time domain and its respective MFCC feature vector were used as described in section 5.2.3. These signals are distributed as follows: 50 signals obtained by the model [23] (see Figure 50A), 50 real signals obtained from different databases (See Figure 50B), 50 signals representing the training data set of the Discriminant model (see Figure 50C) and 50 signals obtained by the FeaturesGAN generator model after 5000 training epochs. Unlike experiment 1, fewer samples were used in this experiment to better visualize the mix and clusters of signal types. However, the behavior is similar using a larger number of samples.

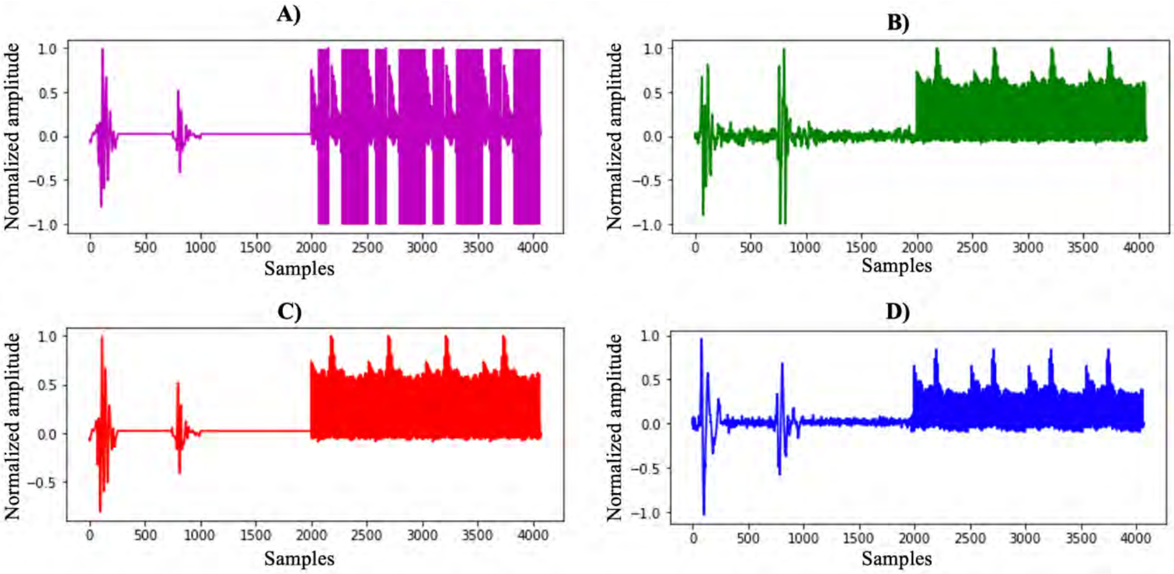


Figure 50. Concatenated signals used in the experiment. A) Mathematical model signal. B) Actual signal. C) Signal used in the discriminator. D) Signal obtained from FeaturesGAN after 5000 epochs.

Figure 51 shows the results of PCA and t-SNE. In both cases, the set of signals that correspond to the mathematical model (purple dots) are far from the rest of the signals, and in turn, the real signals (green dots), the signals used in the discriminator (red dots) and the signals generated by FeaturesGAN (blue dots) are mixed. Annex 1 and 2 shows the PCA and t-SNE results respectively at different training epochs in order to observe the behavior of the signals as the iterations increase.

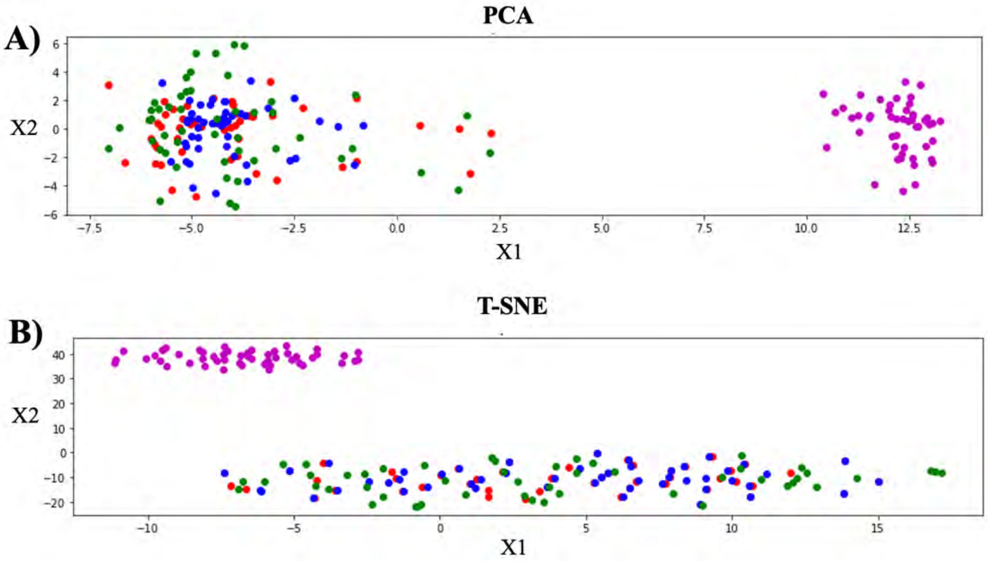


Figure 51. Clustering result using concatenated signals. A) PCA; B) T-SNE

Taking into account that FeaturesGAN tries to fit the characteristics of a signal from a mathematical model, Figure 52 shows the results of PCA and t-SNE using only the segment of MFCC characteristics of the signals. Annex 2 and 3 shows the results at different training epochs in order to analyze the evolution of these characteristics as the iterations progress. The results shown in figure 52 are very similar to those presented in figure 51. It is worth mentioning that the MFCC features that correspond to the discriminator are the same as those of the real signals, therefore the red dots are not visible in the graphs.

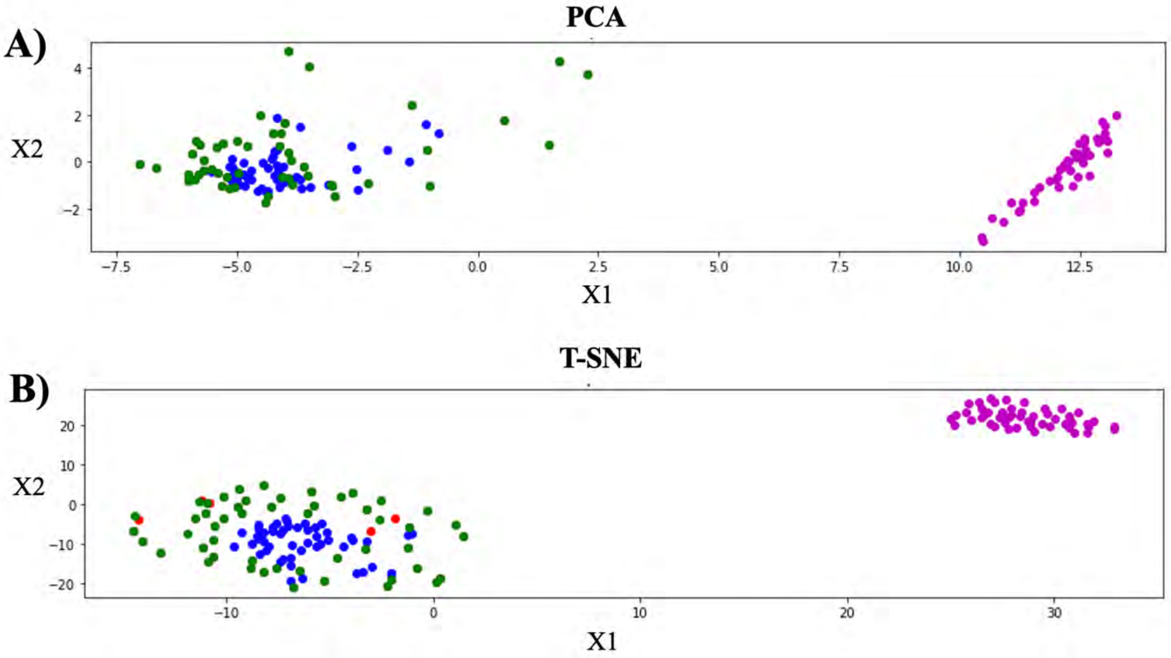


Figure 52. Clustering result using MFCC features. A) PCA; B) T-SNE

<u>Experiment 3:</u> In this third experiment, the same exercise as the previous experiment is done, but this time the FeaturesGAN model is analyzed using MFCC and FFT features in a type of murmur. 200 signals representing the concatenation of a time-domain murmur and its respective MFCC and FFT feature vector were used as described in section 5.2.4. These signals are distributed as follows: 50 signals obtained from the mathematical model proposed in section 5.2.2. (see figure 53A), 50 real signals obtained from different databases that correspond to the same type of murmur (see figure 53B), 50 signals that represent the training data set of the Discriminator model (see figure 53C) and 50 signals obtained by the FeaturesGAN generator model after 10000 training epochs (see figure 53D). Annex 3 and 4 shows the PCA and t-SNE results respectively at different training epochs in order to observe the behavior of the signals as the iterations increase.

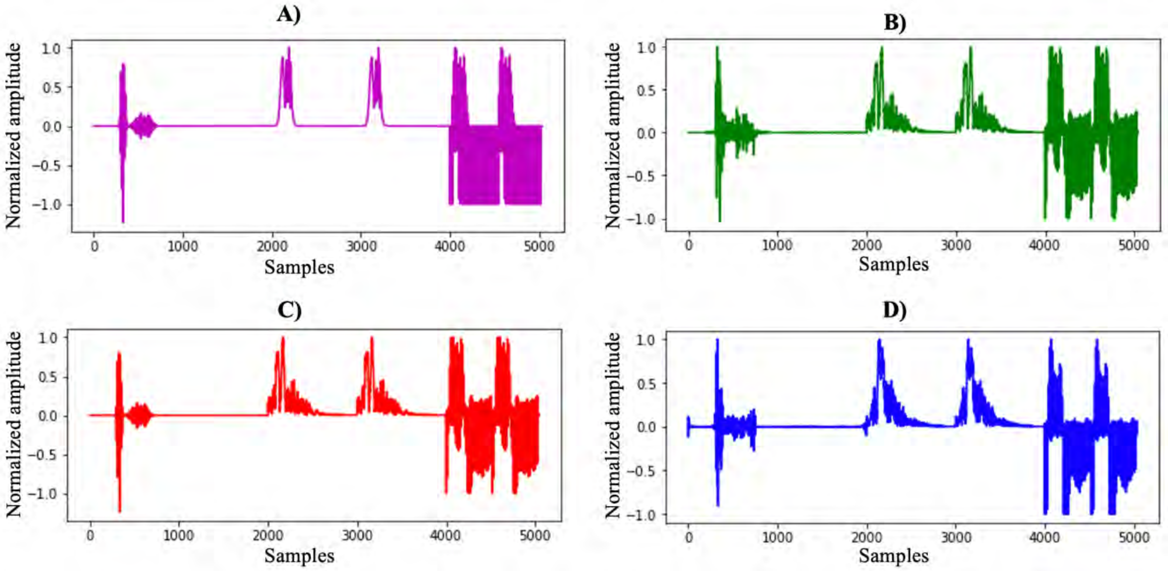


Figure 53. Concatenated signals used in the experiment. A) Mathematical model signal. B) Real signal. C) Signal used in the discriminator. D) Signal obtained from FeaturesGAN after 10000 epochs.

Figure 54 shows the results of PCA and t-SNE. The results were very similar to those obtained in the previous experiment, where the real signals (green dots) are mixed with the signals generated by the FeaturesGAN model (blue dots) and the signals used in discriminator training (red dots). Subsequently, the same procedure was performed using only the feature segment FFT and MFCC.

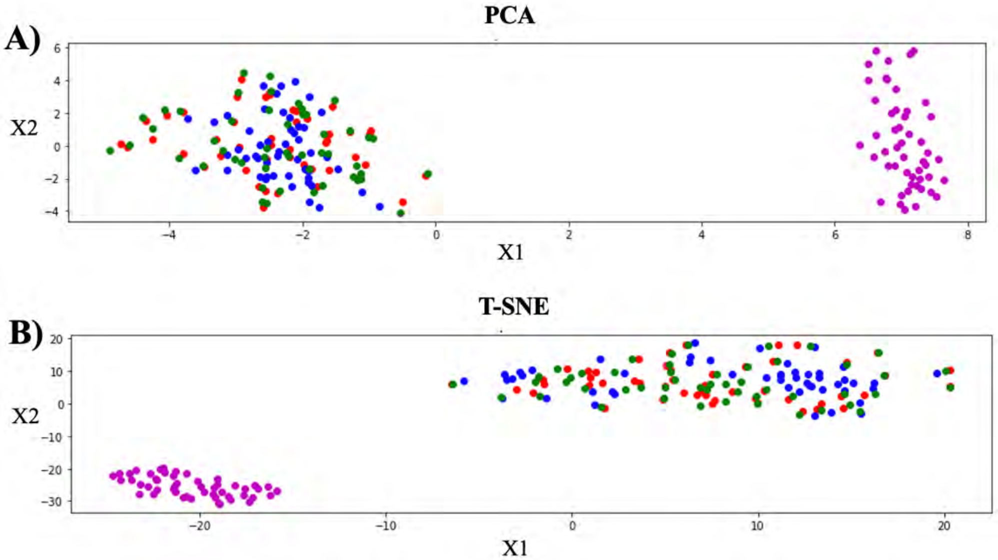


Figure 54. Clustering result using concatenated signals. A) PCA; B) T-SNE

Figures 55 and 56 show the results of PCA and t-SNE for the set of FFT and MFCC features, respectively, although the same clustering behavior is shown in both types of signals, with the MFCC features the result is more pronounced. Annexes 5 to 10 show the PCA and t-SNE results at different times of training.

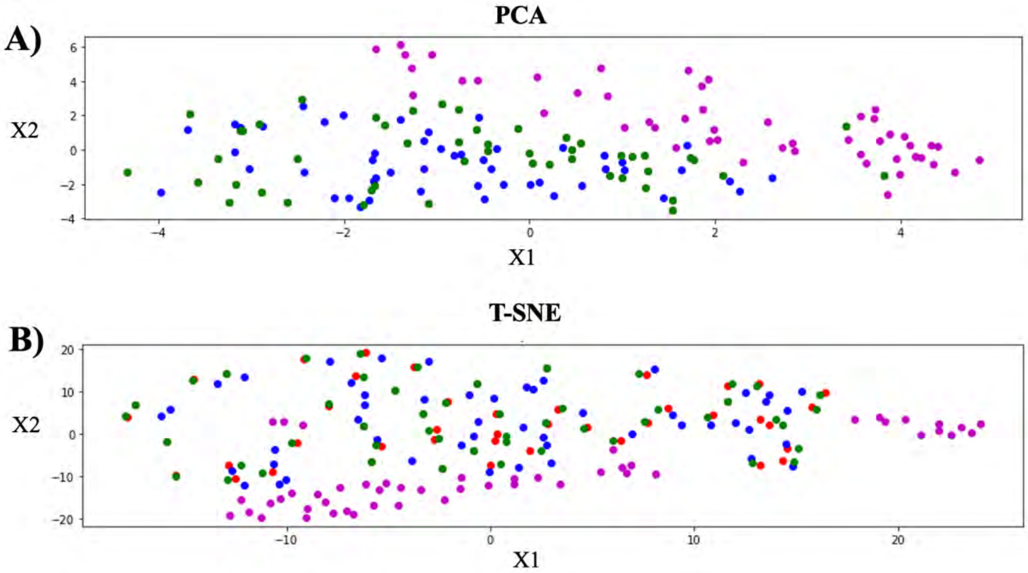


Figure 55. Clustering result using FFT features. A) PCA; B) T-SNE

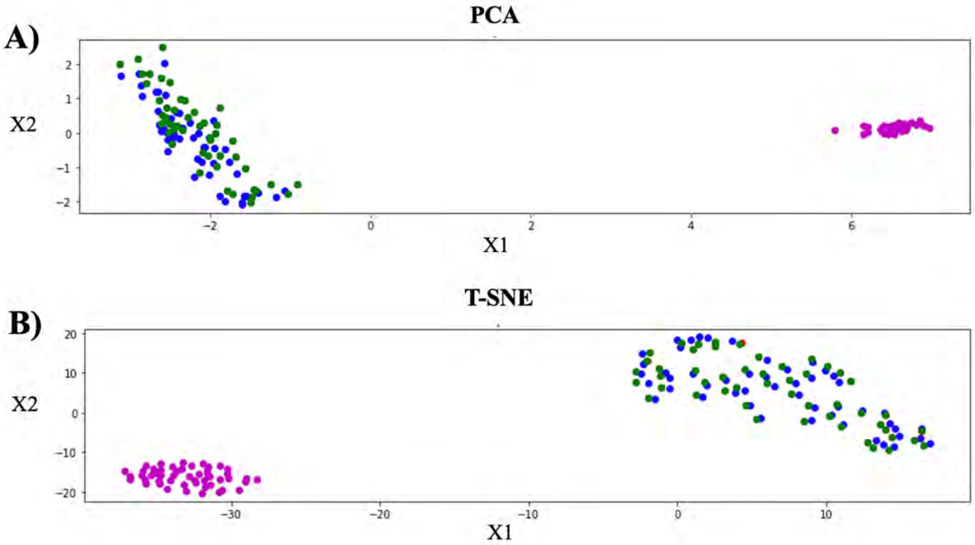


Figure 56. Clustering result using MFCC features. A) PCA; B) T-SNE

5.2.5.3. MCD and SSIM

In this section, another objective evaluation is made based on Mel Cepstral Distortion (MCD), which was used to evaluate the GAN model proposed in section 5.1. Additionally, the Structural Similarity Index Measure (SSIM) method is used, which is widely used to evaluate the similarity or quality of the images [**160**]. SSIM measures the perceptual difference between two similar images based on structure, brightness, and contrast. For the SSIM calculation between two images, the Scikit-Image library offered by Python was used [**161**]. The result is a value between 0 and 1, where 1 indicates that both images are identical. For this case, spectrogram images of each of the real and synthetic signals were used, taking as reference the signals obtained by the mathematical model.

The same methodology proposed in section 5.1.2.1 was implemented. for the calculation of the MCD and SSIM indices. Experiments were performed with normal heart sounds and different types of abnormalities, such as: Aortic Stenosis (AS), Mitral Regurgitation (MR) and Mitral Valve Prolapse (MVP). Figure 57 shows an example of each of the signals.

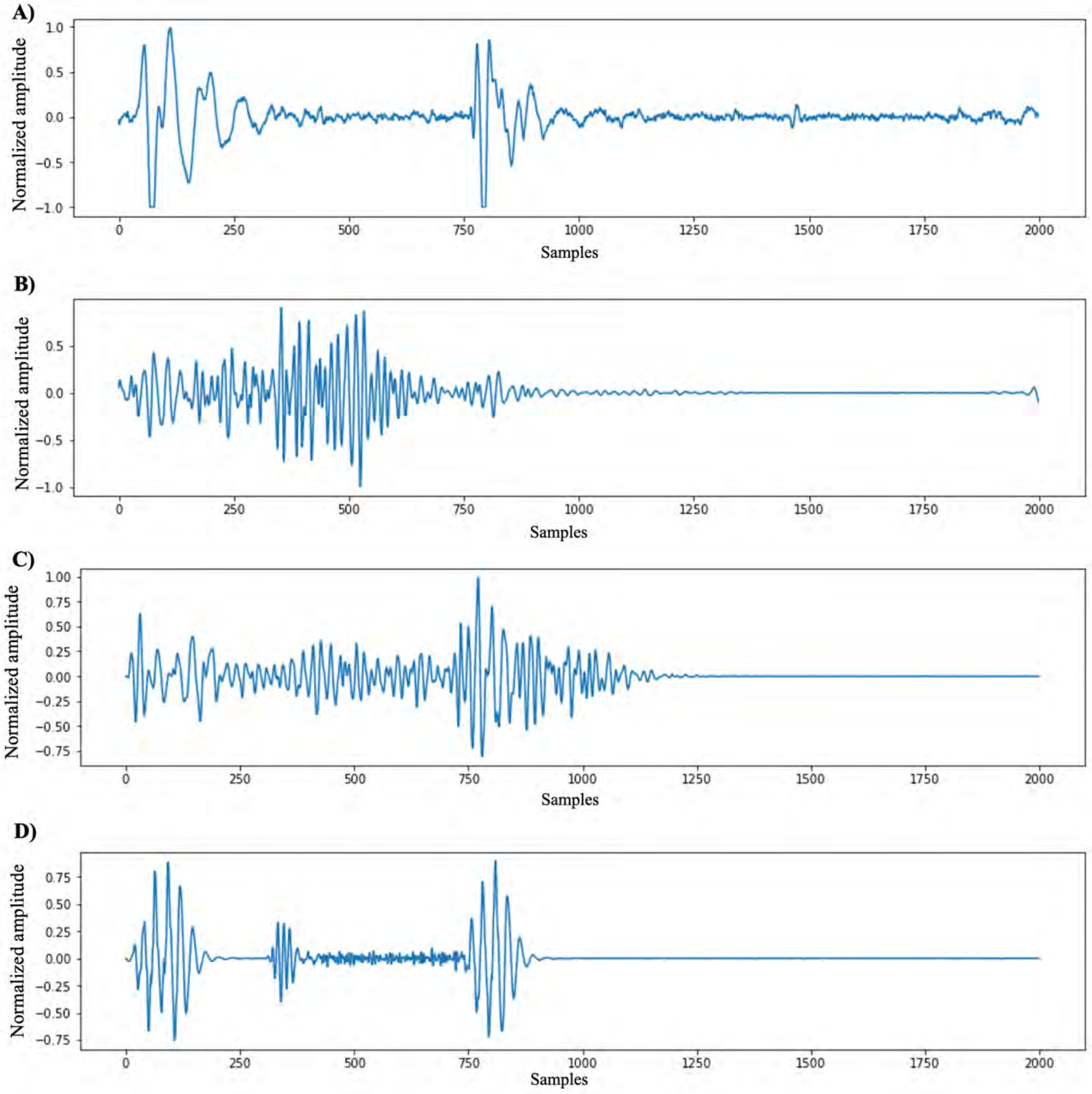


Figure 57. Examples of heart sounds. A) Normal signal, B) Aortic Stenosis, C) Mitral Regurgitation D) Mitral Valve Prolapse

In each of the experiments, a total of 150 cardiac signals were used, distributed as follows: 50 real signals obtained from different databases, 50 synthetic signals obtained by the mathematical model [23] and 50 synthetic signals obtained from the FeaturesGAN model. Figures 58A, 58B, 58C, and 58D show the MCD (Y-axis) and SSIM (X-axis) results for normal heart sounds, with aortic stenosis, with mitral regurgitation, and with Mitral Valve Prolapse, respectively. It can be seen that there is a very marked similarity between the real signals (green dots) and the synthetic signals obtained by the FeaturesGAN model (blue dots).

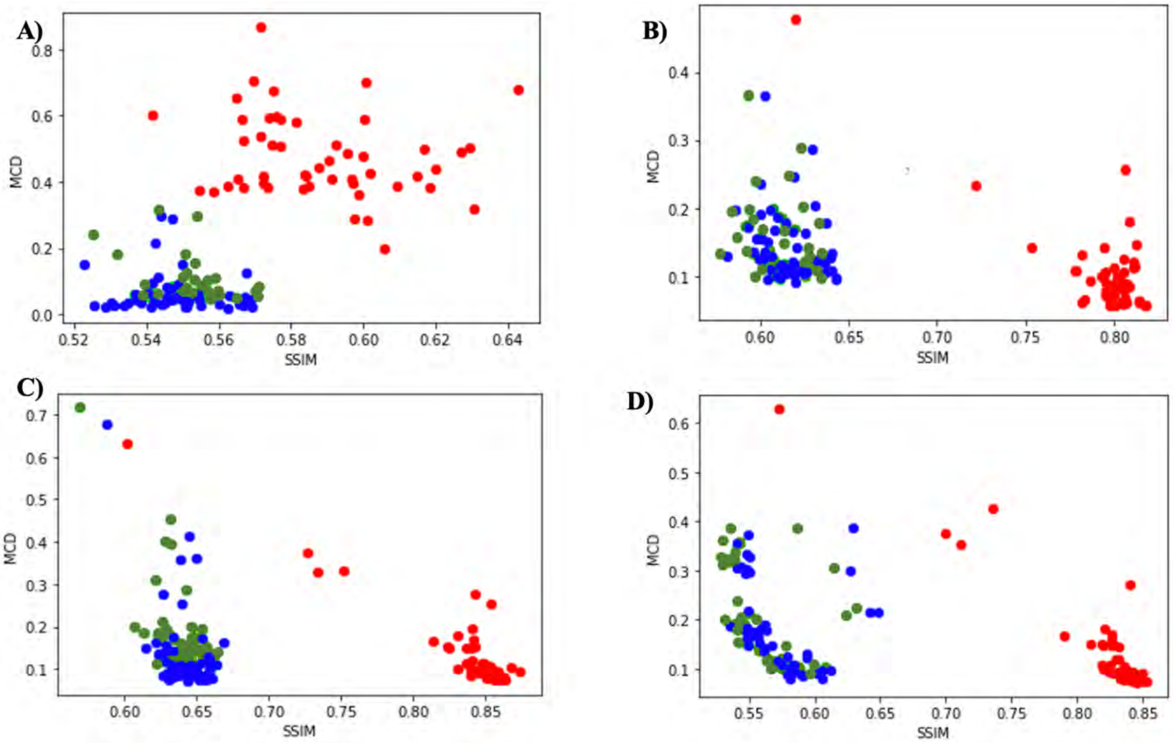


Figure 58. Mel-Cepstral Distortion (MCD) and Structural Similarity Index Measure (SSIM) results. A) Normal heart sounds; B) Abnormal heart sound with aortic stenosis; C) Abnormal heart sound with Mitral Regurgitation; D) Abnormal heart sound with Mitral Valve Prolapse. The red, blue, and green dots represent the real heart sounds, those obtained by FeaturesGAN, and those obtained by the mathematical model, respectively.

5.2.5.4. Classification models

Different proposed methods of automatic classification of normal and abnormal heart sounds are used to perform tests with synthetic signals. Similar to what was done in section 5.1.2.2. the models proposed in [49], [50], [144] and [151] were used. A total of 1000 synthetic heart sounds were used for the tests of the different pre-trained classification models, of which 500 signals correspond to normal heart sounds and the other 500 signals correspond to different types of abnormalities. Table 34 presents the accuracy results for the different feature extraction techniques and Machine Learning algorithms. The results obtained continue to be favorable using synthetic signals, in various combinations of features with ML models the accuracy results are greater than 90%.

Table 34. Accuracy results of synthetic signals using the trained models proposed in articles [49], [50], [144] and [151].

Feature Extraction	**Classifier**			
	SVM	**KNN**	**RF**	**MLP**
[151]: EWT + Power	99%	99%	98%	98%
[49]: MFCC + DWT	78%	88%	80%	90%
[50]: Statistical, frequency and perceptual	89%	86%	90%	94%
[144]: LPC + MFCC	98%	94%	90%	80%

5.2.5.5. MOS test

A dataset of real and synthetic heart sounds is established for mean opinion scoring (MOS) with clinicians to validate the quality of the signals. At this stage, normal heart sounds and types of abnormalities such as: Aortic Stenosis (AS), Mitral Regurgitation (MR) and Mitral Valve Prolapse (MVP) were selected. The MOS test is a widely used method to subjectively assess speech quality. A scale of 1 to 5 defined in the standard ITU-T P.800 is generally used to assess audio quality [162]. Table 35 specifies the meaning of each scale.

Table 35. Scales defined in the MOS test.

Scala	Description
1	Bad
2	Poor
3	Acceptable
4	Good
5	Excellent

In a MOS test, test persons listen to short speech samples (in the case of voice) and score according to perceived quality. The total MOS score is then the mean of all individual scores. For our case, the audio signals are normal and abnormal heart sounds that are evaluated by 5 doctors who are experts in identifying heart murmurs. For this, 3 types of tests were carried out:

1. **MOS test with real signals:** This test is intended to validate that the doctor agrees with the label of the respective signals (Normal, AS, MR and MVP), a total of 5 sounds of each type were used, the doctor delivers his score after listening to each type of signal, and then the average of these scores is calculated. Table 36 shows the average score results for each physician, and the overall average for each type of heart sound. According to the scoring scale, in all cases the doctors evaluated it as "Good".

Table 36. MOS test results using real signals

Doctor	Average Score			
	Normal	AS	MS	MR
1	5	4	4.2	4.7
2	4.5	4	4.3	4
3	4.2	4.7	4	4.2
4	4.6	4.6	4.5	4.8
5	4.8	4.8	4.7	4.5
Total Score	4.6	4.4	4.3	4.4

2. **MOS test with real and synthetic signals**: In this test, real and synthetic signals are used, which are previously randomized. The goal of this test is to determine if clinicians perceive a significant difference between the signals. Table 37 shows the results of this test, obtaining results on similar scales to the previous test. The synthetic signals were obtained from the FeaturesGAN model, since with this model it was possible to obtain variability in the different types of heart signals.

Table 37. MOS test results using real and synthetic signals

Doctor	Average score – Real signals				Average score – Synthetic signals			
	Normal	AS	MS	MR	Normal	AS	MS	MR
1	5.0	4.0	4.1	3.8	5.0	4.1	4.2	3.9
2	4.3	4.1	3.8	4.0	4.9	4.4	4.4	4.0
3	4.6	4.7	4.0	4.2	4.3	4.2	4.1	4.4
4	4.9	4.2	4.7	4.0	4.2	4.6	4.0	4.2
5	4.7	4.8	4.6	4.5	4.4	4.7	4.2	4.1
Total score	4.7	4.4	4.2	4.1	4.6	4.4	4.2	4.1

3. **MOS test with synthetic signals**: Finally, only synthetic signals generated from the FeaturesGAN model were used in this experiment in order to rectify the results obtained in the previous test. Table 38 shows the results of this test, showing results comparable with the real signals, obtaining a MOS scale of "good" in each of the types of signals.

Table 38. MOS test results using synthetic signals

Doctor	**Average score**			
	Normal	**AS**	**MS**	**MR**
1	4.8	4.4	4.2	3.9
2	5.0	4.5	4.1	4.1
3	4.5	4.3	4.0	4.5
4	4.7	4.5	3.9	4.6
5	4.5	4.6	4.0	4.2
Total score	4.7	4.5	4.0	4.3

6. Conclusions

In this work several algorithms are proposed for the analysis (automatic segmentation and classification) and synthesis of normal and abnormal heart sounds. In each of them, different tests, validations, and comparisons of results with the methods proposed in the state of the art were carried out.

In the stage of **ANALYSIS** of heart sounds, an algorithm for the automatic segmentation of heart sounds based on EWT and NASE was implemented, obtaining good results in the identification cardiac cycles and their segments (i.e. S1, systole, S2 and diastole) in a recording. The algorithm was tested with two datasets of heart sounds proposed in the Pascal Challenge [17]; the metric used in the challenge consists of the sum of the differences between manual segmentation labels provided in the datasets and the segmentation labels obtained with the automatic segmentation in each recording [17]. The results of the proposed methods compared favorably to methods in the current literature that used the same dataset. The error for dataset A was 843,440.8 samples and for dataset B was 17,074.1 samples, achieving a reduction of 3.45% and 41.67% respectively compared to the best result published in the state of the art [43]. Additionally, tests were performed with recordings from the Physionet database, obtaining good segmentation performance. With the help of this system, we can analyze and extract characteristics for each heart cycle segment for the development of a complete system for the automatic classification of heart sounds.

On the other hand, several combinations of features and classifiers have been presented to identify normal and abnormal sounds. The proposed method of feature extraction is based on power values in the systolic and diastolic intervals. Subsequently, four classification models were used: SVM, KNN, random forest and MLP. In addition, the characteristics proposed in [49] and [50] were extracted and the results compared with the proposed method. Similarly, models based on deep learning, such as those used in [51] and [52], were also implemented and compared with the proposed method.

A classification experiment was carried out using samples obtained from different databases downloaded from the Internet. The best results were obtained with the signal power values calculated in the systole and diastole; the classifier with the best results in accuracy and specificity was KNN with values of 99.25% and 100%, respectively, while the random forest classifier obtained the best result of sensitivity and AUC, with values of 98.8% and 99.62%, respectively.

These results compare favorably with those presented in the state-of-the-art approaches [49], [50], [51], [52] and [144] (see Tables 21, 22, 23, 24 and 25); additionally, our experiment used a greater number of testing samples compared to previous works, therefore giving more statistical weight to the results, plus a low computational cost for feature extraction and a small number of characteristics for the classification stage, and the tests were performed together with the proposed automatic segmentation algorithm. In general, the resulting metrics for the classification test are at or close to the top marks in Tables 21, 22, 23, 24 and 25, for every tested classification model. This can be seen as a strong indication that the proposed segmentation and feature extraction methods are indeed useful irrespective of the classification model that is then applied. This method could potentially be implemented in real-time and guarantees a rapid response with low computational cost.

The main limitation that exists in the proposed method is when the recording of the cardiac signal has ambient noises with a high amplitude, since these noises can be contained in different frequency bands. Therefore, automatic segmentation can identify false positives in the systolic or diastolic interval. Similarly, the power features can vary when the signal has these types of noise and the classifier could in turn be confused regarding whether a heart murmur is present, with the detected murmur actually being an ambient noise.

In the **SYNTHESIS** stage of heart sounds, two GAN models were proposed for the generation of normal signals and types of abnormalities. A GAN-based architecture was initially implemented to generate synthetic heart sounds, which can be used to train/test classification models. The proposed GAN model is accompanied by a denoising stage using the Empirical Wavelet Transform, which allows to decrease the number of training epochs and, therefore the total computational cost, obtaining a synthetic cardiac signal with a low noise level.

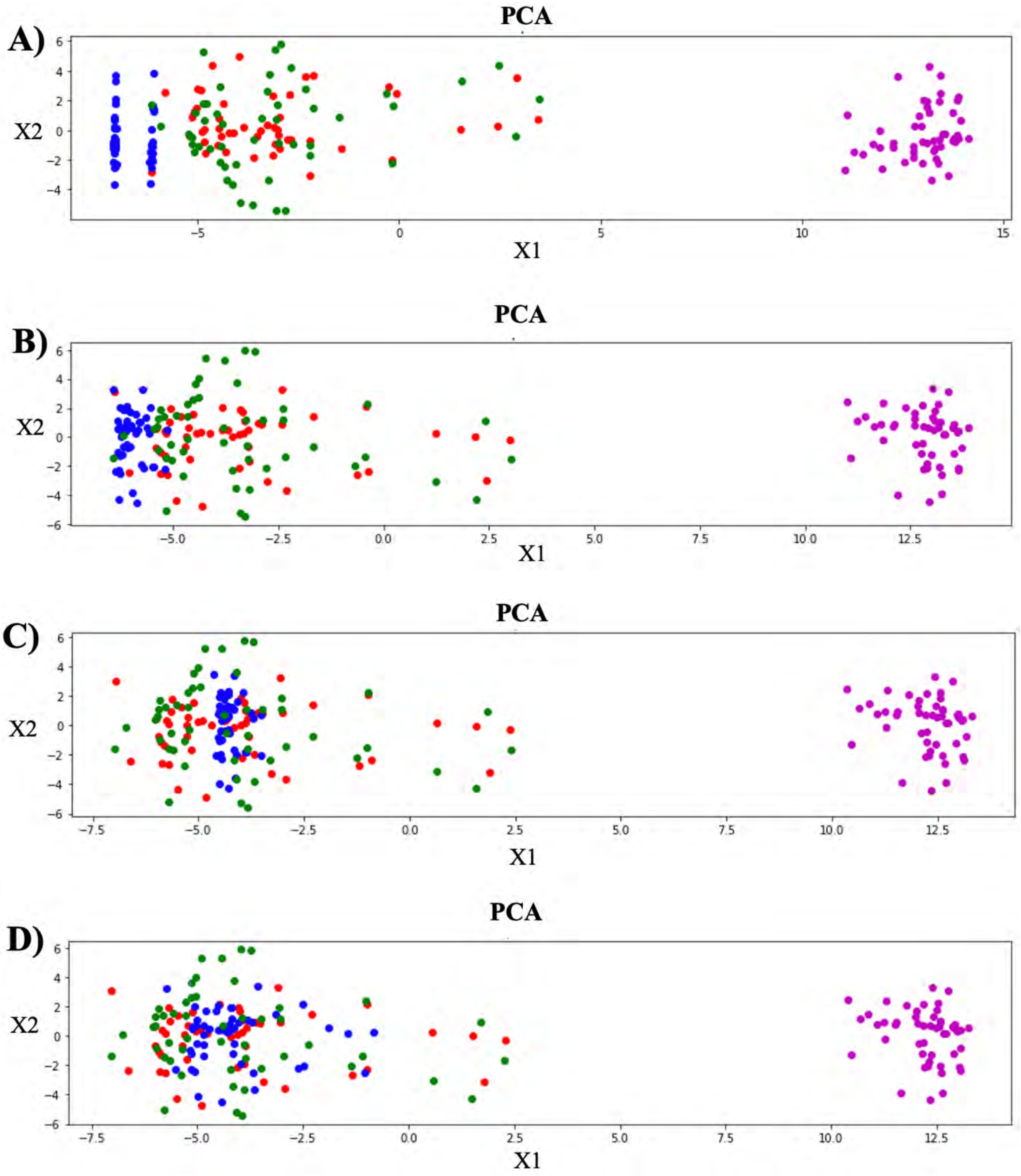
The proposed method was compared with a mathematical model proposed in the state-of-the-art [**23**]. Two evaluation tests were carried out, the first is to measure the distortion between the natural and synthetic cardiac signals, in order to objectively evaluate the similarity between them. In this case, the Mel Cepstral Distortion (MCD) method was used, this method being widely used in the evaluation of audio quality. In this test, the synthetic signal generated with the proposed method obtained a better similarity result with the natural signals, compared to the mathematical model proposed in [**23**]. The second method consisted of using different pre-trained classification Machine Learning models with good precision performance, in order to use the synthetic signals as test dataset and verify if the different ML models perform well. In this test, the power characteristics proposed in [151] with the different Machine Learning models registered the best results. Generally speaking, most of the combinations of features with classification models performed well in discriminating synthetic heart sounds as normal, as shown in Table 28.

The second proposed GAN model allows the generation of cardiac signals with murmur using few real samples for training. This model combines a synthetic signal obtained by a mathematical model with features in the frequency and perceptual domain of real signals. An analysis of the Generator and Discriminator models was carried out, achieving a decrease in the computational cost. Comparisons between the mathematical model and real signals were made by applying clustering on PCA and t-SNE features. Tests were also carried out with MCD and SSIM distortion metrics, in this last metric Spectrogram images of the synthetic signal were used to evaluate the similarity with spectrogram images of real signals. The clustering results in each of the experiments show that there is a similarity between the real signals and the signals generated by the FeaturesGAN model.

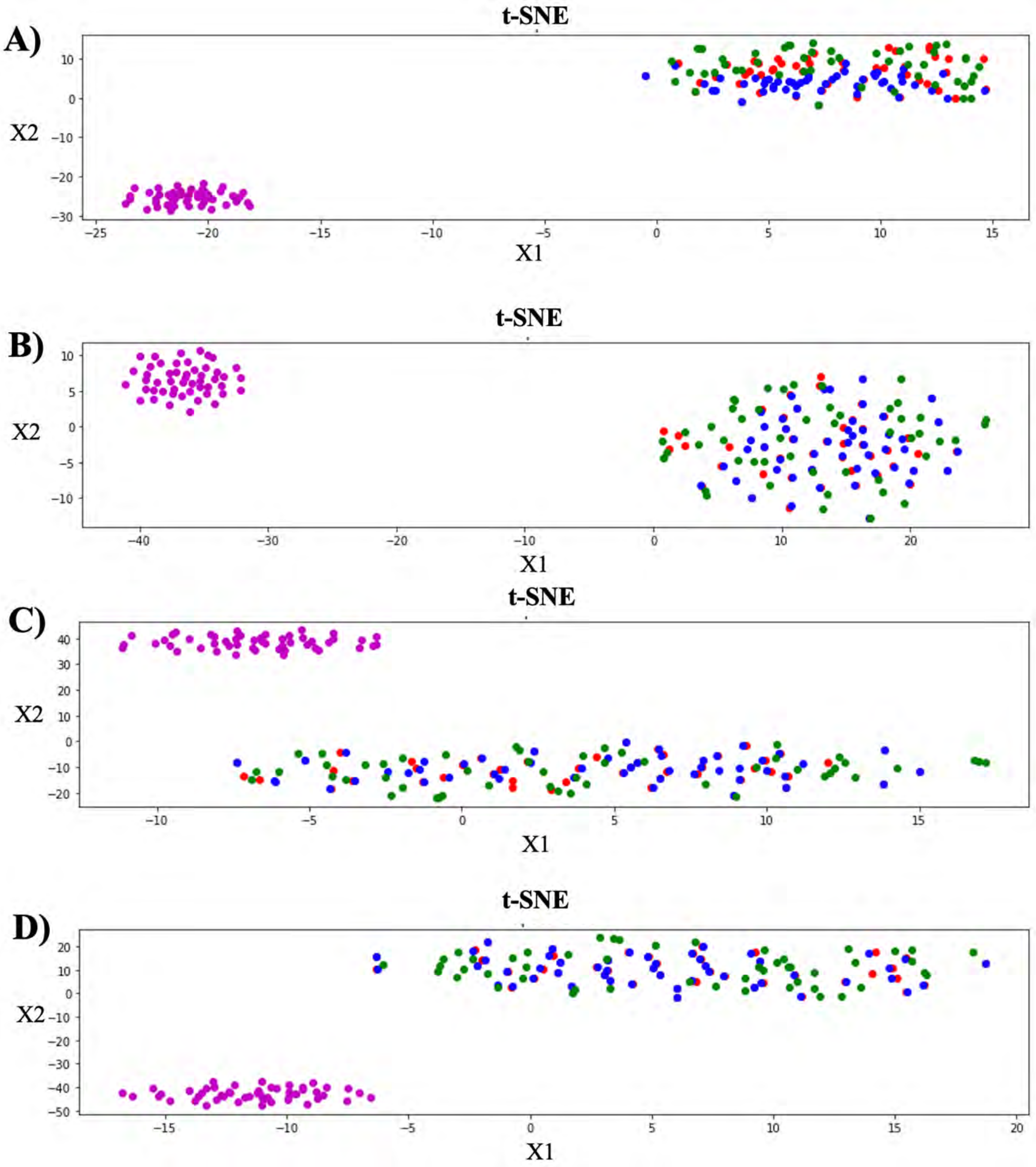
Additionally, tests were performed with the different classification models pre-trained to assess the accuracy performance using synthetic normal and abnormal signals on the test dataset. The results are favorable in the different proposed classification models. Finally, MOS tests were carried out with expert doctors to evaluate the quality of the sounds in a subjective way, indicating good opinion results. In general terms, according to the results obtained in all the validation tests, a strong indication can be seen that the synthetic signals obtained with the proposed model present very similar characteristics to the real signals, for which they can be used to improve the performance of heart sound classification models, since the number of samples in training could be increased.

Annexes

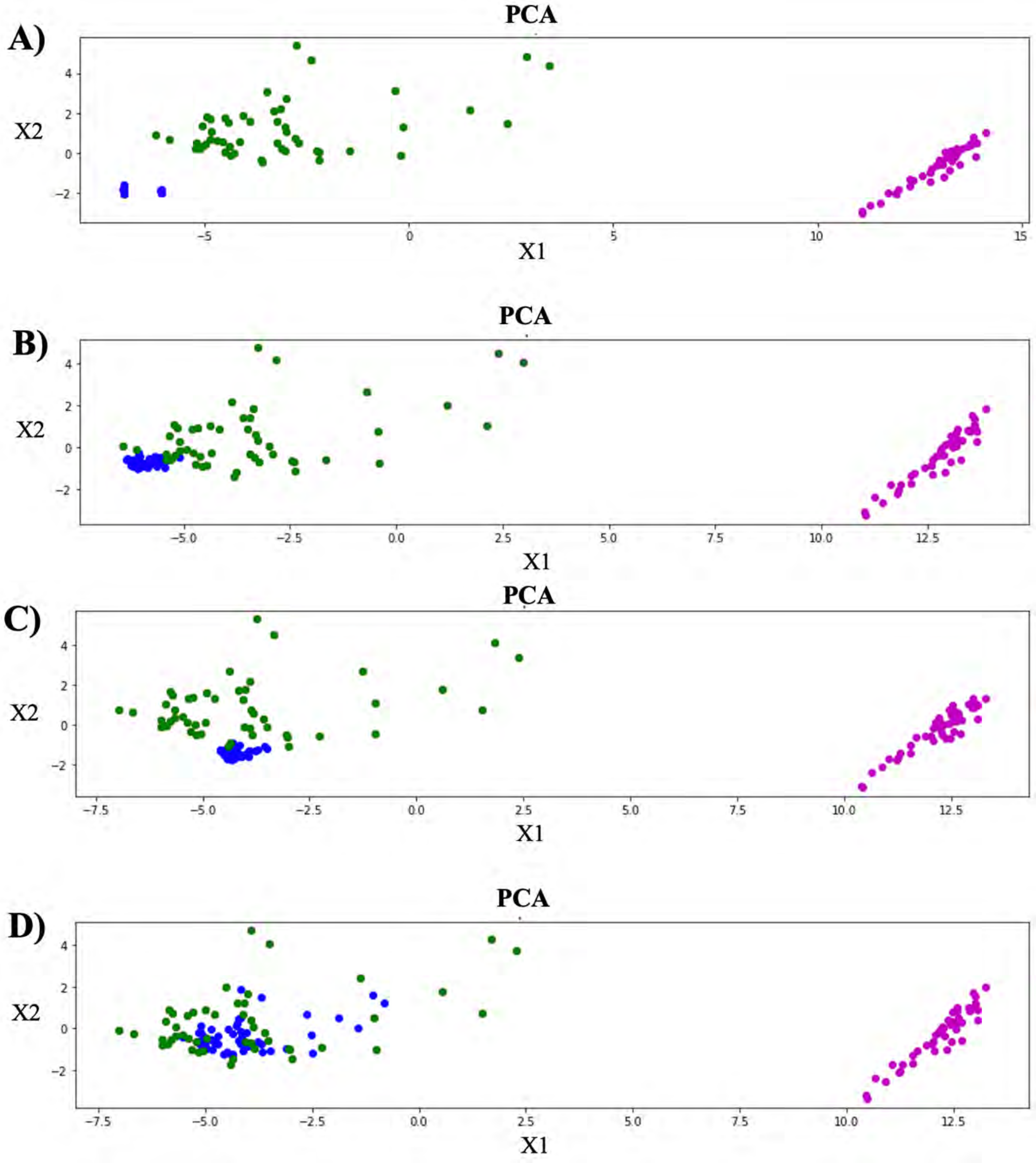
<u>Annex 1</u>: Result of clustering with PCA using concatenated signals for the generation of normal heart sounds at different training epochs. A) Epoch 100; B) Epoch 1000; C) Epoch 2500; B) Epoch 5000.



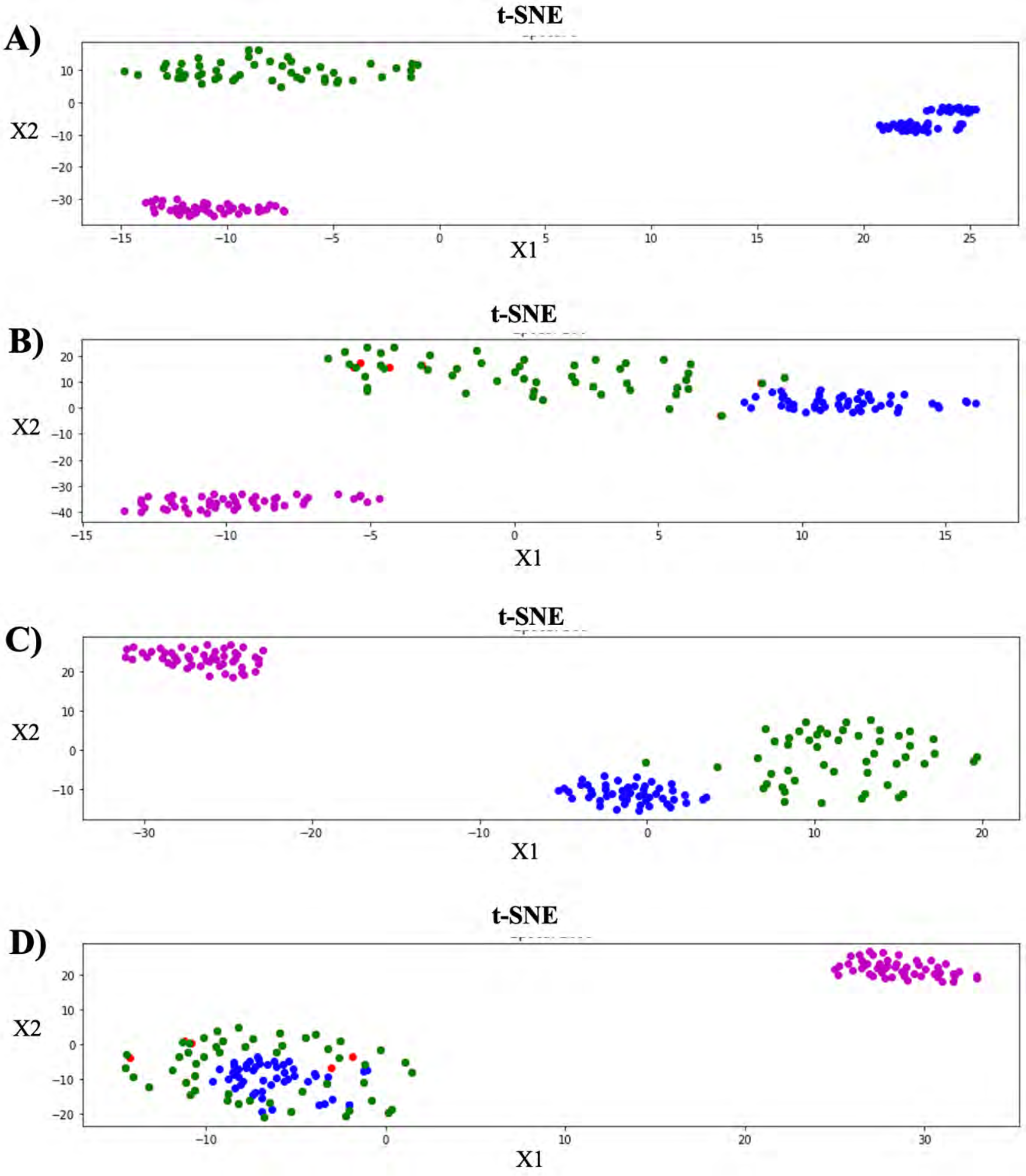
Annex 2: Result of clustering with t-SNE using concatenated signals for the generation of normal heart sounds at different training epochs. A) Epoch 100; B) Epoch 1000; C) Epoch 2500; B) Epoch 5000.



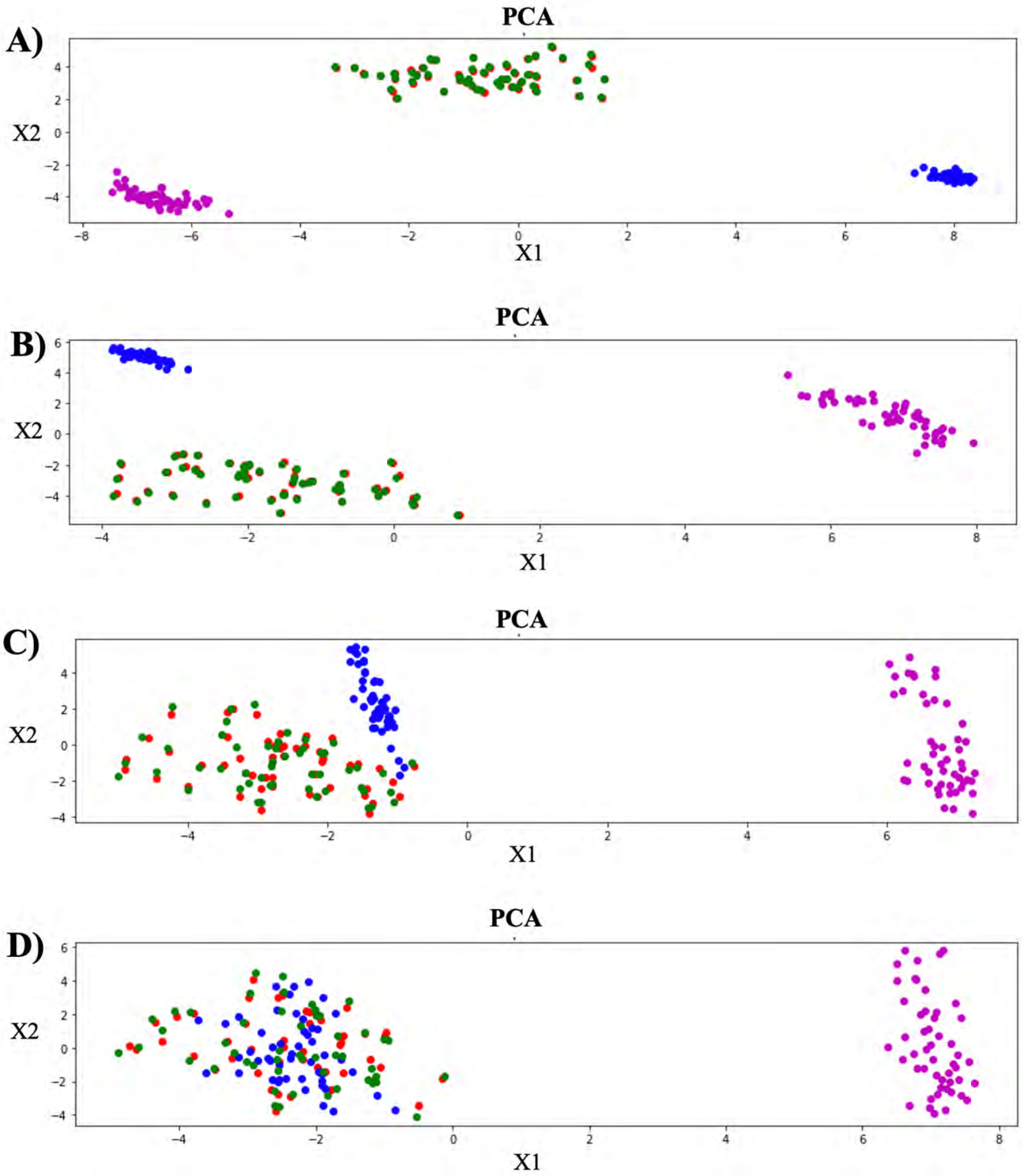
Annex 3: Result of clustering with PCA using MFCC feature segments for the generation of normal heart sounds at different training epochs. A) Epoch 100; B) Epoch 1000; C) Epoch 2500; B) Epoch 5000.



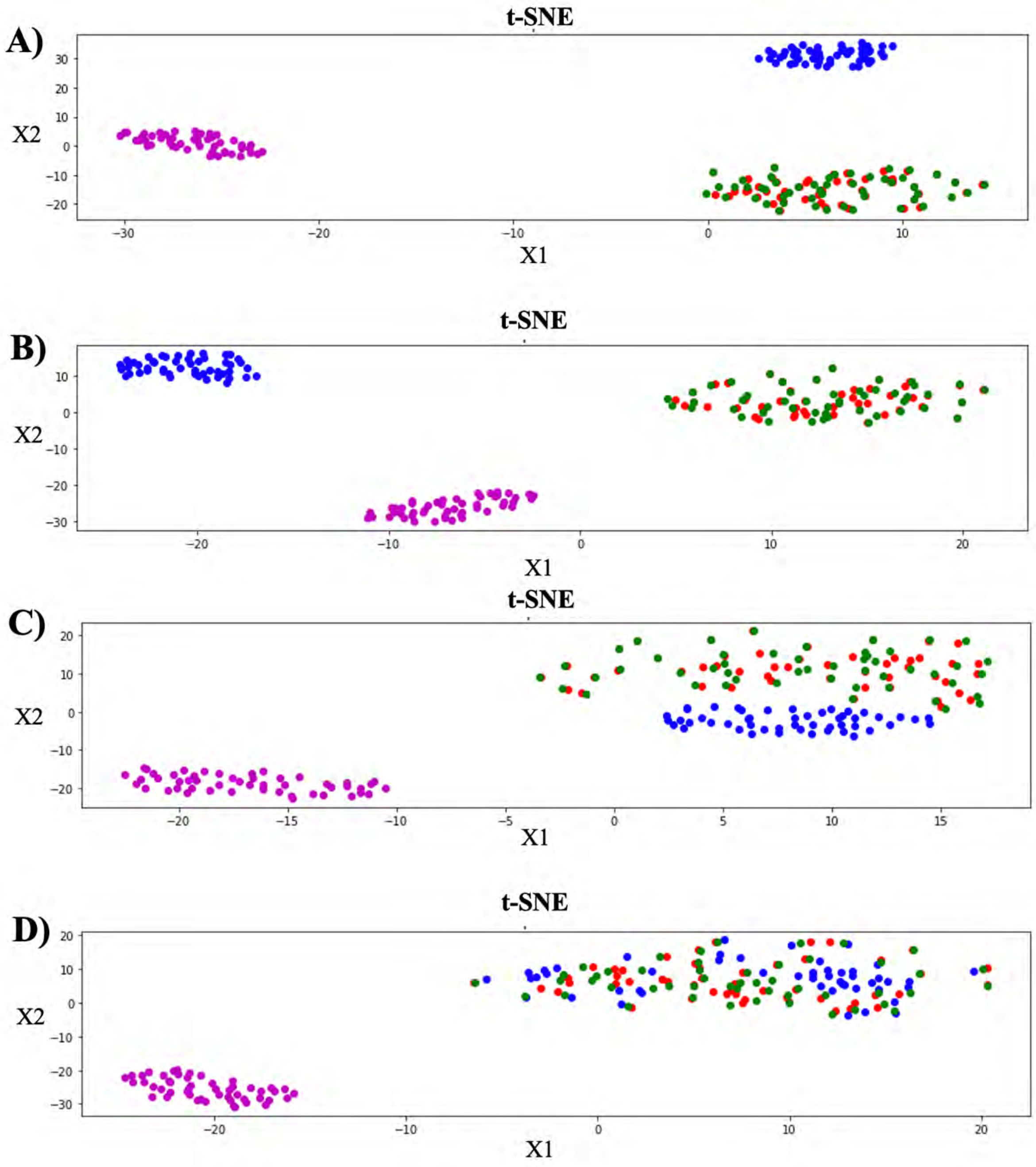
Annex 4: Result of clustering with **t-SNE** using MFCC feature segments for the generation of **normal heart sounds** at different training epochs. A) Epoch 100; B) Epoch 1000; C) Epoch 2500; B) Epoch 5000.



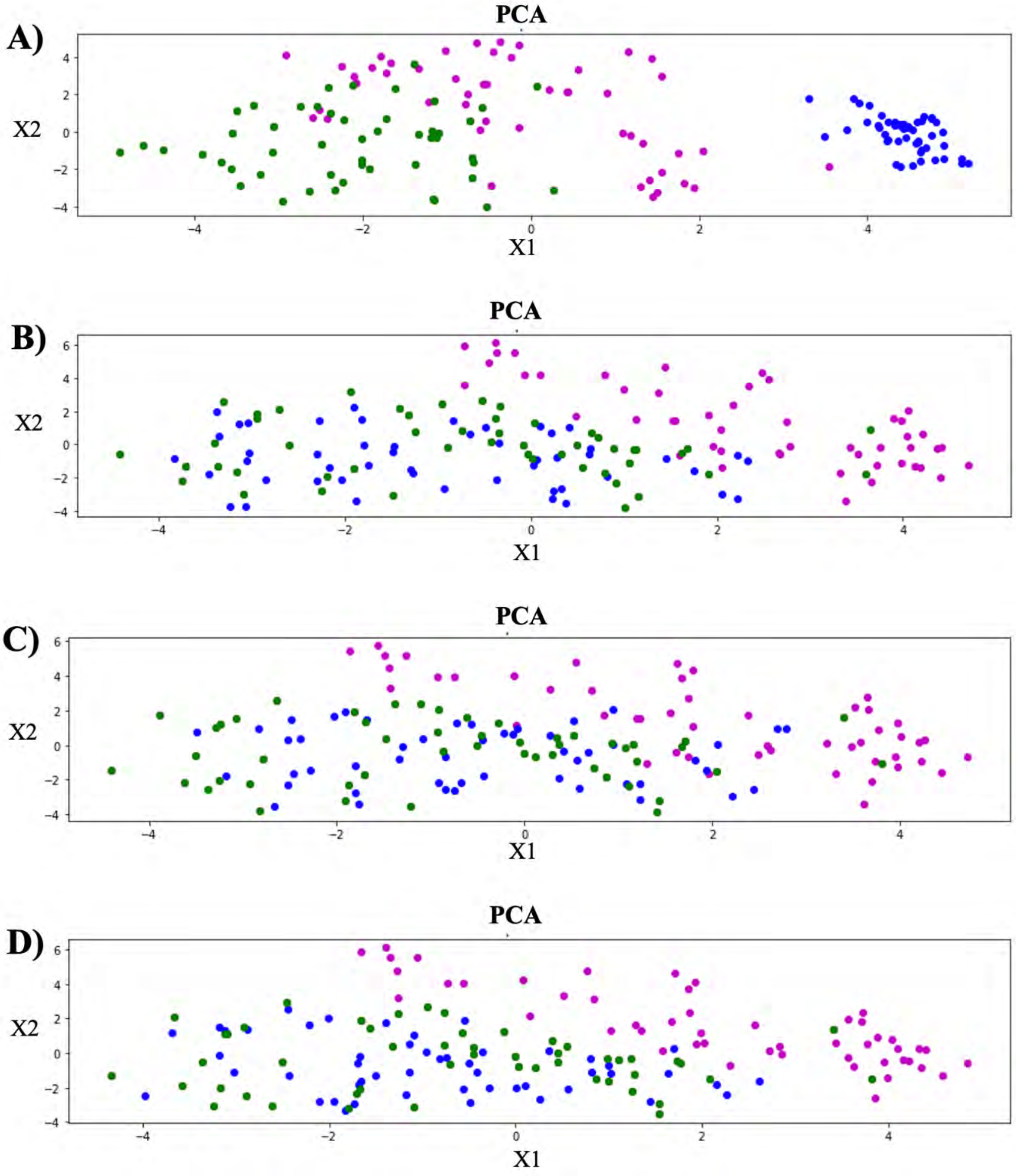
Annex 5: Result of clustering with **PCA** using concatenated signals for the generation of **murmur heart** at different training epochs. A) Epoch 100; B) Epoch 2000; C) Epoch 5000; B) Epoch 10000.



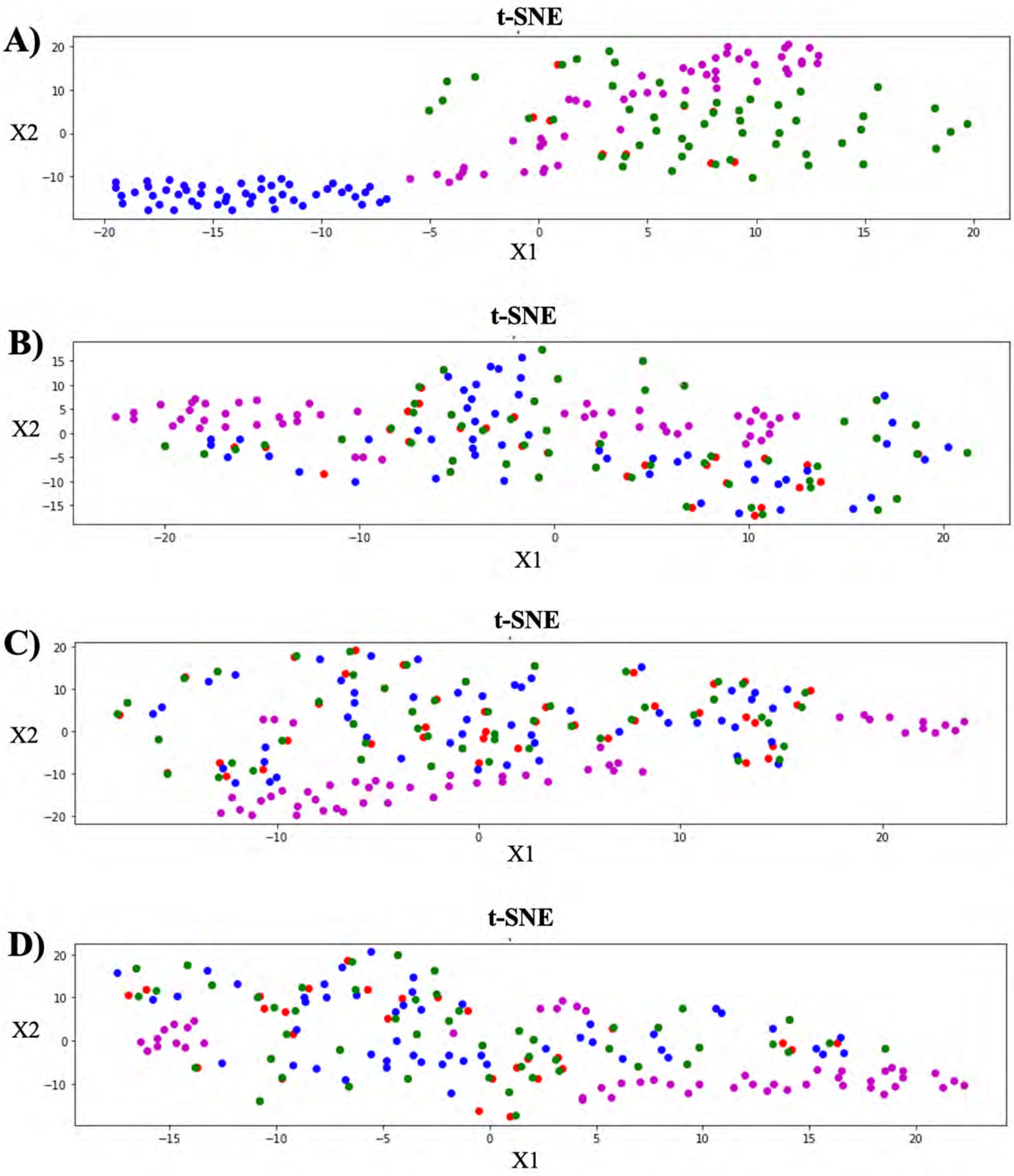
<u>Annex 6</u>: Result of clustering with **t-SNE** using concatenated signals for the generation of **murmur heart** at different training epochs. A) Epoch 100; B) Epoch 2000; C) Epoch 5000; B) Epoch 10000.



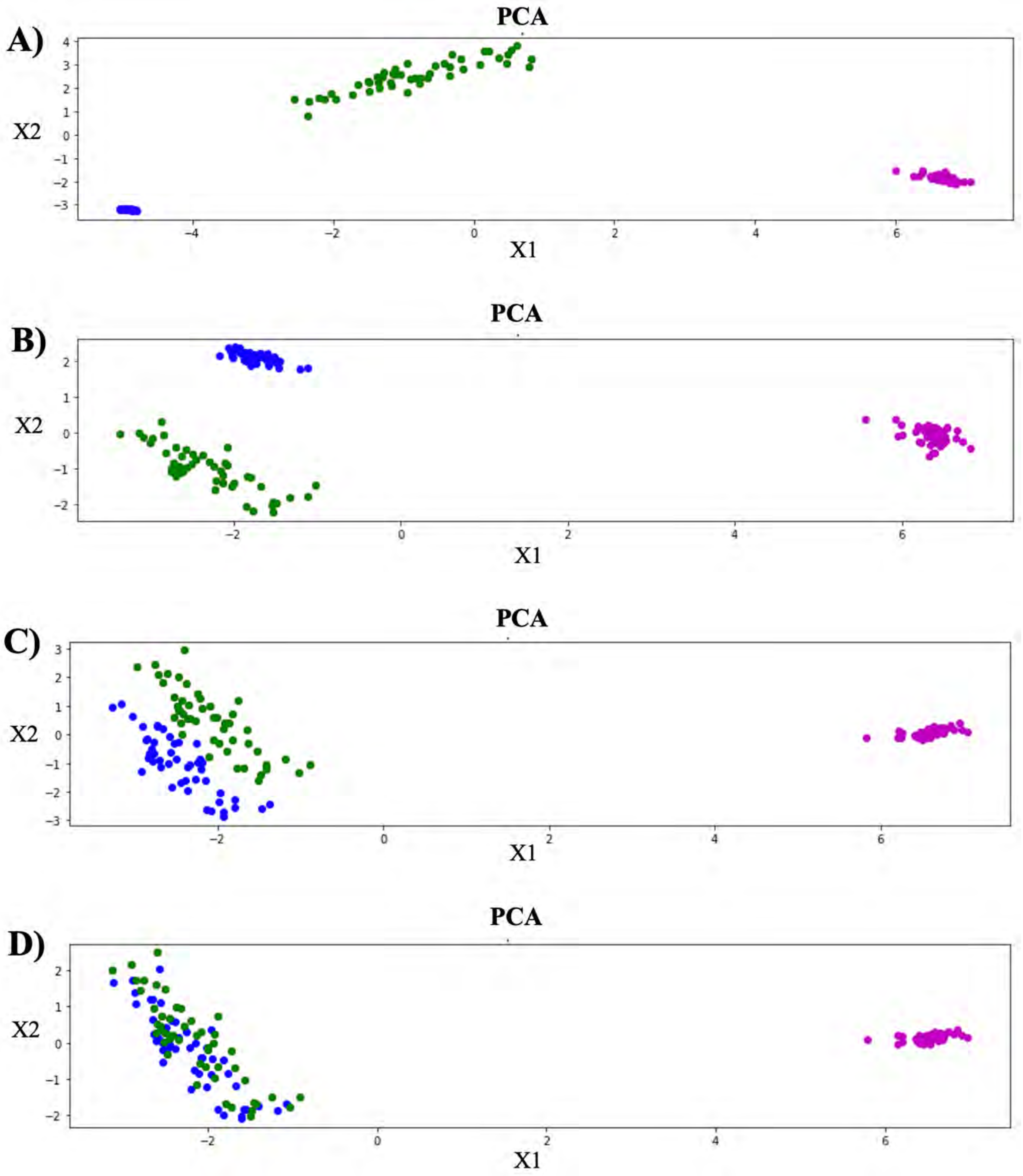
Annex 7: Result of clustering with **PCA** using FFT feature segments for the generation of **murmur heart** at different training epochs. A) Epoch 100; B) Epoch 2000; C) Epoch 5000; B) Epoch 10000.



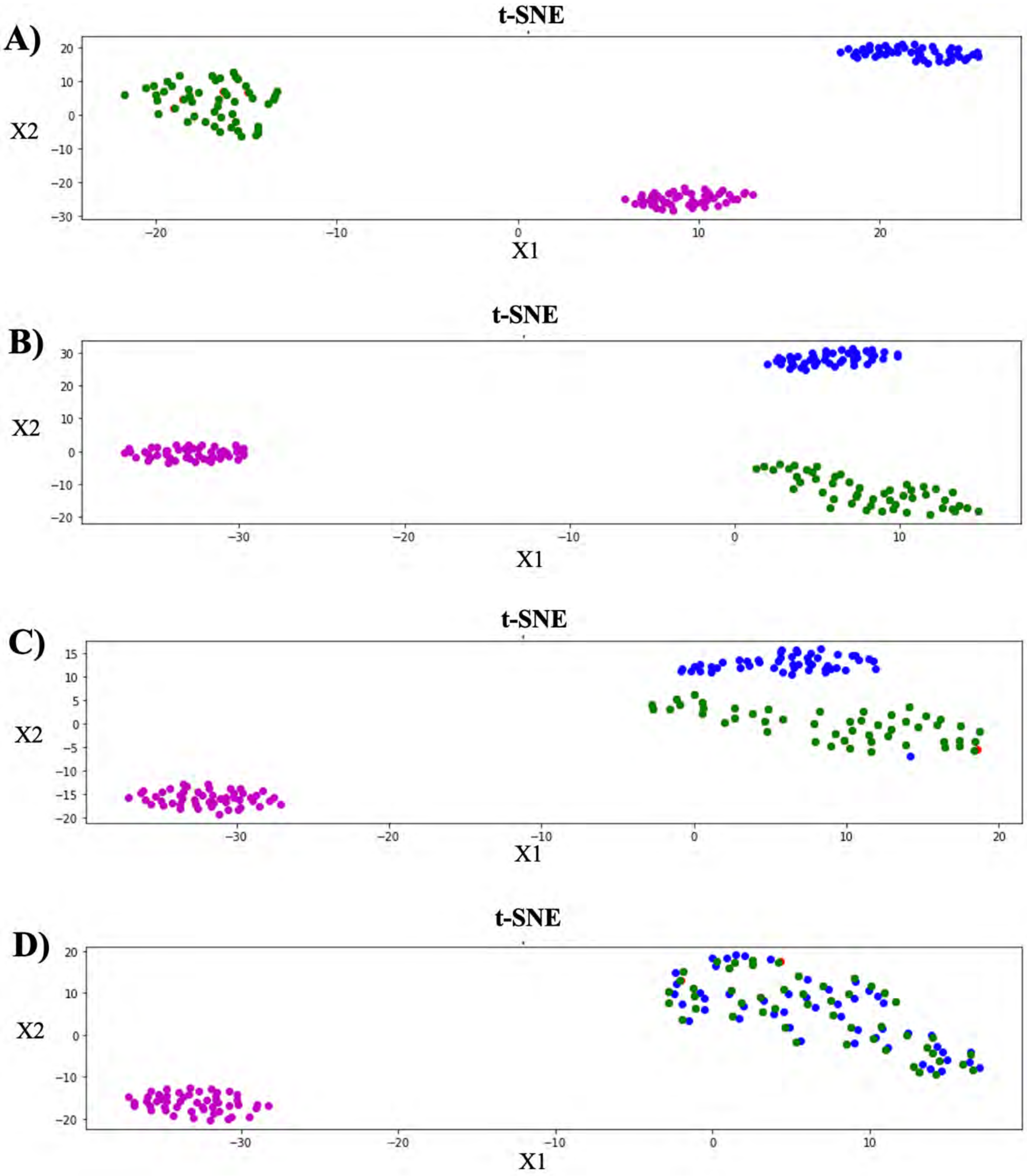
Annex 8: Result of clustering with **t-SNE** using FFT feature segments for the generation of **murmur heart** at different training epochs. A) Epoch 100; B) Epoch 2000; C) Epoch 5000; B) Epoch 10000.



Annex 9: Result of clustering with **PCA** using MFCC feature segments for the generation of **murmur heart** at different training epochs. A) Epoch 100; B) Epoch 2000; C) Epoch 5000; B) Epoch 10000.



Annex 10: Result of clustering with **t-SNE** using MFCC feature segments for the generation of **murmur heart** at different training epochs. A) Epoch 100; B) Epoch 2000; C) Epoch 5000; B) Epoch 10000.



Bibliography

[1] World Health Organization. A global brief on hypertension. 2013 Available at: http://www.who.int/cardiovascular_diseases/publications/global_brief_hypertension/en.

[2] Benjamin, E.J.; Blaha, M.J.; Chiuve, S.E.; Cushman, M.; Das, S.R.; Deo, R.; De Ferranti, S.D.; Floyd, J.; Fornage, M.; Gillespie, C.; et al. Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circ.* 2017, *135*, e146–e603, doi:10.1161/cir.0000000000000485.

[3] Camic, P.M.; Knight, S.J. *Clinical Handbook of Health Psychology: A Practical Guide to Effective Interventions*; Hogrefe & Huber Publishers: Cambridge, MA, USA, 2004; pp. 31–32.

[4] Alvarez, C.; Patiño, A. State of emergency medicine in Colombia. *Int. J. Emerg. Med.* 2015, *8*, 1–6.

[5] Shank, J. *Auscultation Skills: Breath & Heart Sounds*, 5th ed.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2013.

[6] Alam, U.; Asghar, O.; Khan, S.; Hayat, S.; Malik, R. Cardiac auscultation: An essential clinical skill in decline. *Br. J. Cardiol.* 2010, *17*, 8.

[7] Roelandt, J.R.T.C. The decline of our physical examination skills: Is echocardiography to blame? *Eur. Heart J. Cardiovasc. Imaging* 2014, *15*, 249–252.

[8] Clark, D.; Ahmed, M.; Dell'Italia, L.; Fan, P.; McGiffin, D. An argument for reviving the disappearing skill of cardiac auscultation. *Clevel. Clin. J. Med.* 2012, *79*, 536–537.

[9] Shank, J. Auscultation skills: Breath & Heart sounds. Publisher: Lippincott Williams & Wilkins. 5th Edition. 2013

[10] Brown, E., Leung, T., Collis, W. and Salmon, A. "Heart Sounds Made Easy". Churchill Livingstone Elsevier. Edition 2. 2008.

[11] Y. Etoon, S. Ratnapalan. "Evaluation of Children With Heart Murmurs". Clinical Pediatrics 53(2). DOI: 10.1177/0009922813488653. May 2013.

[12] Johnson W., Moller J. Pediatric cardiology: The essential pocket guide. Editorial: Wiley-Blackwell. 2008

[13] University of Michigan. Heart sound and Murmur library. Available in: https://open.umich.edu/find/open-educational-resources/medical/heart-sound-murmur-library. Last visit: 31-05-2018.

[14] University of Washington. Heart sound and murmur. Available in: https://depts.washington.edu/physdx/heart/demo.html. Last visit: 31-05-2018.

[15] Thinklabs. Heart sounds library. Available in: http://www.thinklabs.com/heart-sounds. Last visit: 31-05-2018.

[16] PhysioNet/Computing in Cardiology Challenge. "Classification of Normal/Abnormal Heart Sound Recordings". Available in: https://www.physionet.org/challenge/2016/. Last visit: 31-05-2018

[17] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, R. Getz. "Classifying Heart Sounds Callenge [online]". Available in: http://www.peterjbentley.com/heartchallenge/#downloads. Last visit: 03-05-2018.

[18] Tang, Y.; Danmin, C.H.; Durand, L.G. The synthesis of the aortic valve closure sound on the dog by the mean filter of forward and backward predictor. IEEE Trans. Biomed. Eng. 1992, 39, 1–8. [PubMed]

[19] Tran, T.; Jones, N.B.; Fothergill, J.C. Heart sound simulator. Med. Biol. Eng. Comput. 1995, 33, 357–359. [PubMed]

[20] Zhang, X.; Durand, L.G.; Senhadji, L.; Lee, H.C.; Coatrieux, J.L. Analysis—synthesis of the phonocardiogram based on the matching pursuit method. IEEE Trans. Biomed. Eng. 1998, 45, 962–971. [PubMed]

[21] Xu, J.; Durand, L.; Pibarot, P. Nonlinear transient chirp signal modelling of the aortic and pulmonary components of the second heart sound. IEEE Trans. Biomed. Eng. 2000, 47, 1328–1335.

[22] Toncharoen, C.; Srisuchinwong, B. A heart-sound-like chaotic attractor and its synchronization. In Proceedings of the 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON, Pattaya, Thailand, 6 May 2009.

[23] Almasi, A.; Shamsollahi, M.B.; Senhadji, L. A dynamical model for generating synthetic phonocardiogram signals. In Proceedings of the 33rd Annual International Conference of the IEEE EMBS, Boston, MA, USA, 30 August–3 September 2011.

[24] Tao, Y.W.; Cheng, X.F.; He, S.Y.; Ge, Y.P.; Huang, Y.H. Heart sound signal generator Based on LabVIEW. Appl. Mech. Mater. 2012, 121, 872–876.

[25] Jablouna, M.; Raviera, P.; Buttelli, O.; Ledeea, R.; Harbaa, R.; Nguyenb, L. A generating model of realistic synthetic heart sounds for performance assessment of phonocardiogram processing algorithms. Biomed. Signal Process. Control 2013, 8, 455–465.

[26] Sæderup, R.G.; Hoang, P.; Winther, S.; Boettcher, M.; Struijk, J.J.; Schmidt, S.E.; Ostergaard, J. Estimation of the second heart sound split using windowed sinusoidal models. Biomed. Signal Process. Control 2018, 44, 229–236.

[27] Joseph, A.; Martínek, R.; Kahankova, R.; Jaros, R.; Nedoma, J.; Fajkus, M. Simulator of Foetal Phonocardiographic Recordings and Foetal Heart Rate Calculator. J. Biomim. Biomater. Biomed. Eng. 2018, 39, 57–64.

[28] Van den oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. Available online: https://arxiv.org/abs/1609. 03499 (accessed on 15 September 2020).

[29] Engel, J.; Resnick, C.; Roberts, A.; Dieleman, S.; Eck, D.; Simonyan, K.; Norouzi, M. Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017.

[30] Bollepalli, B.; Juvela, L.; Alku, P. Generative Adversarial Network-Based Glottal Waveform Model for Statistical Parametric Speech Synthesis. In Proceedings of the Interspeech 2017, Stockholm, Sweden, 20–24 August 2017.

[31] Biagetti, G.; Crippa, P.; Falaschetti, L.; Turchetti, C. HMM speech synthesis based on MDCT representation. *Int. J. Speech Technol.* **2018**, *21*, 1045–1055.

[32] Chrism, D.; Julian, M.; Miller, P. Adversarial Audio Synthesis. Available online: https://arxiv.org/abs/1802. 04208 (accessed on 15 September 2020).

[33] H. Liang, S. Lukkarinen, I. Hartimo. Heart Sound Segmentation Algorithm Based on Heart Sound Envelolgram. *Computers in Cardiology 1997*. **1997**, pp. 105–108.

[34] Moukadem, A.; Dieterlen, A.; Hueber, N.; Brandt, C. A robust heart sounds segmentation module based on S-transform. *Biomed. Signal Process. Control.* **2013**, *8*, 273–281, doi:10.1016/j.bspc.2012.11.008.

[35] Huiying, L.; Sakari, L.; Iiro, H. A heart sound segmentation algorithm using wavelet decomposition and reconstruction. In Proceedings of the Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No.97CH36136); Chicago, IL, USA, 30 Oct.–2 Nov 1997; Vol. 4, pp. 1630–1633.

[36] Alexander, B.; Nallathambi, G.; Selvaraj, N. Screening of Heart Sounds Using Hidden Markov and Gammatone Filterbank Models. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA); Orlando, FL, USA,17–20 Dec 2018; pp. 1460–1465.

[37] Springer, D.B.; Tarassenko, L.; Clifford, G.; D.B., S.; L., T.; G.D., C. Logistic Regression-HSMM-based Heart Sound Segmentation. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 1, doi:10.1109/tbme.2015.2475278.

[38] Liu, C.; Springer, D.; Clifford, G.D. Performance of an open-source heart sound segmentation algorithm on eight independent databases. *Physiol. Meas.* **2017**, *38*, 1730–1745, doi:10.1088/1361-6579/aa6e9f.

[39] Y. Deng, P. J. Bentley. A Robust Heart Sound Segmentation and Classification Algorithm using Wavelet Decomposition and Spectrogram. *Workshop Classifying Heart Sounds. La Palma, Canary Islands*, 24 April 2012; pp. 1–6.

[40] Mubarak, Q.-U.-A.; Akram, M.U.; Shaukat, A.; Hussain, F.; Khawaja, S.G.; Butt, W.H. Analysis of PCG signals using quality assessment and homomorphic filters for localization and classification of heart sounds. *Comput. Methods Programs Biomed.* **2018**, *164*, 143–157, doi:10.1016/j.cmpb.2018.07.006.

[41] Schmidt, S.E.; Holst-Hansen, C.; Graff, C.; Toft, E.; Struijk, J.J. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiol. Meas.* 2010, *31*, 513–529, doi:10.1088/0967-3334/31/4/004.

[42] E.F. Gomes, E. Pereira. Classifying heart sounds using peak location for segmentation and feature construction. *Workshop Classifying Heart Sound. La Palma, Canary Islnd.* 24 April 2012; pp. 480–92.

[43] C. Fatima, J. Abdelilah, N. Chafik, H. Ahmed, H. Amir. Detection and Identification Algorithm of the S1 and S2 Heart Sounds. In Proceeding of *2016 International Conference on Electrical and Information Technologies (ICEIT)*. Tangier, Morocco, 4–7 May 2016.

[44] PhysioNet/Computing in Cardiology Challenge. Classification of Normal/Abnormal Heart Sound Recordings. Available Online: https://www.physionet.org/challenge/2016/. Access on: 31-05-2018.

[45] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, R. Getz. Classifying Heart Sounds Callenge [online]. Available Online: http://www.peterjbentley.com/heartchallenge/#downloads. Access on: 03-05-2018.

[46] Renna, F.; Oliveira, J.H.; Coimbra, M. Deep Convolutional Neural Networks for Heart Sound Segmentation. *IEEE J. Biomed. Heal. Informatics* 2019, *23*, 2435–2445, doi:10.1109/jbhi.2019.2894222.

[47] P. J. Bently. "Abstract", in Pascal Workshop Classifying Heart Sounds. 2012.

[48] Oliveira, J.; Renna, F.; Mantadelis, T.; Coimbra, M. Adaptive Sojourn Time HSMM for Heart Sound Segmentation. *IEEE J. Biomed. Heal. Informatics* 2018, *23*, 642–649, doi:10.1109/jbhi.2018.2841197.

[49] Yaseen; Son, G.-Y.; Kwon, S. Classification of Heart Sound Signal Using Multiple Features. *Appl. Sci.* 2018, *8*, 2344, doi:10.3390/app8122344.

[50] Arora, V.; Leekha, R.; Singh, R.; Chana, I. Heart sound classification using machine learning and phonocardiogram. *Mod. Phys. Lett. B* 2019, *33*, doi:10.1142/s0217984919503214.

[51] Noman, F.; Ting, C.-M.; Salleh, S.-H.; Ombao, H. Short-segment Heart Sound Classification Using an Ensemble of Deep Convolutional Neural Networks. In Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Institute of Electrical and Electronics Engineers (IEEE), 2019; pp. 1318–1322.

[52] Raza, A.; Mehmood, A.; Ullah, S.; Ahmad, M.; Choi, G.S.; On, B.-W. Heartbeat Sound Signal Classification Using Deep Learning. Sensors 2019, 19, 4819, doi:10.3390/s19214819.

[53] Abdollahpur, M.; Ghaffari, A.; Ghiasi, S.; Mollakazemi, M.J. Detection of pathological heart sounds. Physiol. Meas. 2017, 38, 1616–1630, doi:10.1088/1361-6579/aa7840.

[54] Maknickas, V.; Maknickas, A. Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients. Physiol. Meas. 2017, 38, 1671–1684, doi:10.1088/1361-6579/aa7841.

[55] Homsi, M.N.; Warrick, P. Ensemble methods with outliers for phonocardiogram classification. Physiol. Meas. 2017, 38, 1631–1644, doi:10.1088/1361-6579/aa7982.

[56] Plesinger, F.; Viscor, I.; Halamek, J.; Jurco, J.; Jurak, P. Heart sounds analysis using probability assessment. Physiol. Meas. 2017, 38, 1685–1700, doi:10.1088/1361-6579/aa7620.

[57] Meintjes, A.; Lowe, A.; Legget, M. Fundamental Heart Sound Classification using the Continuous Wavelet Transform and Convolutional Neural Networks. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Institute of Electrical and Electronics Engineers (IEEE), 2018; pp. 409–412.

[58] Nogueira, D.M.; Ferreira, C.A.; Gomes, E.; Jorge, A.M. Classifying Heart Sounds Using Images of Motifs, MFCC and Temporal Features. J. Med Syst. 2019, 43, 168, doi:10.1007/s10916-019-1286-5.

[59] Kay, E.; Agarwal, A. DropConnected neural networks trained on time-frequency and inter-beat features for classifying heart sounds. Physiol. Meas. 2017, 38, 1645–1657, doi:10.1088/1361-6579/aa6a3d.

[60] Li, L.; Wang, X.; Du, X.; Liu, Y.; Liu, C.; Qin, C.; Li, Y. Classification of heart sound signals with BP neural network and logistic regression. 2017 Chinese Automation Congress (CAC) 2017, 7380–7383, doi:10.1109/cac.2017.8244111.

[61] Hamidi, M.; Ghassemian, H.; Imani, M. Classification of heart sound signal using curve fitting and fractal dimension. Biomed. Signal Process. Control. 2018, 39, 351–359, doi:10.1016/j.bspc.2017.08.002.

[62] Juniati, D.; Khotimah, C.; Wardani, D.E.K.; Budayasa, I.K. Fractal dimension to classify the heart sound recordings with KNN and fuzzy c-mean clustering methods. J. Physics: Conf. Ser. 2018, 953, 12202, doi:10.1088/1742-6596/953/1/012202.

[63] Zhang, W.; Han, J.; Deng, S. Heart sound classification based on scaled spectrogram and tensor decomposition. *Expert Syst. Appl.* **2017**, *84*, 220–231, doi:10.1016/j.eswa.2017.05.014.

[64] Huang, He & S. Yu, Phillip & Wang, Changhu. An Introduction to Image Synthesis with Generative Adversarial Nets. 2018.

[65] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in neural information processing systems, 2014, pp. 2672–2680.

[66] Radford, Alec & Metz, Luke & Chintala, Soumith. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016.

[67] Pang, Yutian & Liu, Yongming. Conditional Generative Adversarial Networks (CGAN) for Aircraft Trajectory Prediction considering weather effects. Doi: 10.2514/6.2020-1853. (2020).

[68] Zhang, Zhaoyu & Li, Mengyan & Yu, Jun. On the Convergence and Mode Collapse of GAN. doi: 1-4. 10.1145/3283254.3283282. (2018).

[69] Bhagyashree, V. Kushwaha and G. C. Nandi, "Study of Prevention of Mode Collapse in Generative Adversarial Network (GAN)," 2020 IEEE 4th Conference on Information & Communication Technology (CICT), 2020, pp. 1-6, doi: 10.1109/CICT51604.2020.9312049.

[70] Jenni, Simon & Favaro, Paolo. On Stabilizing Generative Adversarial Training With Noise. doi: 12137-12145. 10.1109/CVPR.2019.01242. (2019).

[71] Feng, Ruili & Zhao, Deli & Zha, Zhengjun. On Noise Injection in Generative Adversarial Networks. International Conference on Learning Representations, ICLR 2020, Vienna, Austria, May 4, 2020.

[72] Gujar, Sujit. Generative Adversarial Networks (GANs): The Progress So Far In Image Generation. Computer Science. (2019).

[73] Lee, Je-Yeol & Choi, Sang-Il. (2020). Improvement of Learning Stability of Generative Adversarial Network Using Variational Learning. Applied Sciences. 10. 4528. 10.3390/app10134528.

[74] Piacentino, E.; Guarner, A.; Angulo, C. Generating Synthetic ECGs Using GANs for Anonymizing Healthcare Data. Electronics 2021, 10, 389. https://doi.org/10.3390/electronics10040389

[75] Sarkar, Pritam & Etemad, Ali. (2020). CardioGAN: Attentive Generative Adversarial Network with Dual Discriminators for Synthesis of ECG from PPG.

[76] Antczak, K.. "A Generative Adversarial Approach To ECG Synthesis And Denoising." ArXiv abs/2009.02700 (2020): n. pag.

[77] W. Naren, W. Wang, P. Sun, K. Wang, Y. Xia and H. Zhang. "Generating electrocardiogram signals by deep learning." Neurocomputing 404 (2020): 122-136.

[78] Brophy, Eoin. (2020). Synthesis of Dependent Multichannel ECG using Generative Adversarial Networks. 3229-3232. 10.1145/3340531.3418509.

[79] Golany, Tomer, Daniel Freedman and Kira Radinsky. "SimGANs: Simulator-Based Generative Adversarial Networks for ECG Synthesis to Improve Deep ECG Classification." ICML (2020).

[80] Delaney, Anne Marie, Eoin Brophy and Tomas E. Ward. "Synthesis of Realistic ECG using Generative Adversarial Networks." ArXiv abs/1909.09150 (2019): n. pag.

[81] F. Ye, F. Zhu, Y. Fu and B. Shen, "ECG Generation With Sequence Generative Adversarial Nets Optimized by Policy Gradient," in IEEE Access, vol. 7, pp. 159369-159378, 2019, doi: 10.1109/ACCESS.2019.2950383.

[82] Golany, T., & Radinsky, K. (2019). PGANs: Personalized Generative Adversarial Networks for ECG Synthesis to Improve Patient-Specific Deep ECG Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 557–564. doi:10.1609/aaai.v33i01.3301557

[83] Zhu, Fei & Fei, Ye & Fu, Yuchen & Liu, Quan & Shen, Bairong. (2019). Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. Scientific Reports. 9. 10.1038/s41598-019-42516-z.

[84] Fei, Ye & Zhu, Fei & Fu, Yuchen & Shen, Bairong. (2019). ECG Generation With Sequence Generative Adversarial Nets Optimized by Policy Gradient. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2950383.

[85] Harada, Shota & Hayashi, Hideaki & Uchida, Seiichi. (2019). Biosignal Generation and Latent Variable Analysis With Recurrent Generative Adversarial Networks. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2934928.

[86] Yang, Jun & Yu, Huijuan & Shen, Tao & Song, Yaolian & Chen, Zhuangfei. (2021). 4-Class MI-EEG Signal Generation and Recognition with CVAE-GAN. Applied Sciences. 11. 1798. 10.3390/app11041798.

[87] K. Gautam, C. Tran, M. Carnahan, Y. Han and A. H. Tewfik. “Generating EEG features from Acoustic features.” 2020 28th European Signal Processing Conference (EUSIPCO) (2021): 1100-1104.

[88] Luo TJ, Fan Y, Chen L, Guo G, Zhou C. EEG Signal Reconstruction Using a Generative Adversarial Network With Wasserstein Distance and Temporal-Spatial-Frequency Loss. Front Neuroinform. 2020;14:15. Published 2020 Apr 30. doi:10.3389/fninf.2020.00015

[89] Aznan, Nik & Atapour Abarghouei, Amir & Bonner, Stephen & Connolly, Jason & Al Moubayed, Noura & Breckon, Toby. (2019). Simulating Brain Signals: Creating Synthetic EEG Data via Neural-Based Generative Models for Improved SSVEP Classification. 1-8. 10.1109/IJCNN.2019.8852227.

[90] F. Fahimi, Z. Zhang, W. B. Goh, K. K. Ang and C. Guan, "Towards EEG Generation Using GANs for BCI Applications," 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2019, pp. 1-4, doi: 10.1109/BHI.2019.8834503.

[91] Hartmann, Kay Gregor, Robin Tibor Schirrmeister and Tonio Ball. “EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals.” ArXiv abs/1806.01875 (2018): n. pag.

[92] D. Kiyasseh et al., "PlethAugment: GAN-Based PPG Augmentation for Medical Diagnosis in Low-Resource Settings," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 11, pp. 3226-3235, Nov. 2020, doi: 10.1109/JBHI.2020.2979608.

[93] E. Campbell, J. A. D. Cameron and E. Scheme, "Feasibility of Data-driven EMG Signal Generation using a Deep Generative Model," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 3755-3758, doi: 10.1109/EMBC44109.2020.9176072.

[94] Zanini, Rafael Anicet and E. Colombini. “Parkinson's Disease EMG Data Augmentation and Simulation with DCGANs and Style Transfer.” Sensors 20 (2020): 2605.

[95] Hazra, Debapriya & Byun, Yungcheol. (2020). SynSigGAN: Generative Adversarial Networks for Synthetic Biomedical Signal Generation. Biology. 9. 441. 10.3390/biology9120441.

[96] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220; 2000 (June 13). PMID: 10851218; doi: 10.1161/01.CIR.101.23.e215

[97] Pimentel, M. A.; Johnson, A. E.; Charlton, P. H.; Birrenkott, D.; Watkinson, P. J.; Tarassenko, L.; and Clifton, D. A. 2016. Toward a robust estimation of respiratory rate from pulse oximeters. IEEE Transactions on Biomedical Engineering 64(8): 1914–1923.

[98] Karlen, W.; Raman, S.; Ansermino, J. M.; and Dumont, G. A. 2013. Multiparameter respiratory rate estimation from the photoplethysmogram. IEEE Transactions on Biomedical Engineering 60(7): 1946–1953.

[99] Reiss, A.; Indlekofer, I.; Schmidt, P.; and Van Laerhoven, K. 2019. Deep PPG: large-scale heart rate estimation with convolutional neural networks. Sensors 19(14): 3079.

[100] Schmidt, P.; Reiss, A.; Duerichen, R.; Marberger, C.; and Van Laerhoven, K. 2018. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the International Conference on Multimodal Interaction, 400–408.

[101] P. Wagner et al., "PTB-XL, a large publicly available electrocardiography dataset," Sci. Data, vol. 7, no. 1, p. 154, Dec. 2020, doi: 10.1038/s41597-020-0495-6.

[102] R. Mark, G. Moody. MIT-BIH Arrhythmia Database Directory. Massachusetts Institute of Technology, Cambridge, 1988.

[103] A. Bagnall et al., "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," Data Mining and Knowledge Discovery, vol. Online First, 2016.

[104] Zhang, Haihong & Guan, Cuntai & Ang, Kai & Wang, Chuanchu. (2012). BCI Competition IV – Data Set I: Learning Discriminative Patterns for Self-Paced EEG-Based Motor Imagery Detection. Frontiers in neuroscience. 6. 7. 10.3389/fnins.2012.00007.

[105] G. Krishna, C. Tran, Y. Han, M. Carnahan, and A. Tewfik, "Speech synthesis using eeg," in Acoustics, Speech and Signal Processing (ICASSP), 2020 IEEE International Conference on. IEEE, 2020.

[106] Luo, T.-J., Lv, J., Chao, F., and Zhou, C. (2018b). Effect of different movement speed modes on human action observation: an EEG study. *Front. Neurosci*. 12:219. doi: 10.3389/fnins.2018.00219

[107] Luciw, M. D., Jarocka, E., and Edin, B. B. (2014). Multi-channel EEG recordings during 3,936 grasp and lift trials with varying weight and friction. *Sci. Data* 1:140047. doi: 10.1038/sdata.2014.47

[108] Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci*. 6:55. doi: 10.3389/fnins.2012.00055

[109] M. S. Treder, A. Bahramisharif, N. M. Schmidt, M. A. van Gerven, and B. Blankertz, "Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention," Journal of NeuroEngineering and Rehabilitation, journal article vol. 8, no. 1, p. 24, May 2011.

[110] Y. Liang, Z. Chen, G. Liu, and M. Elgendi, "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in china," Scientific data, vol. 5, p. 180020, 2018.

[111] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215–e220, 2000.

[112] Pinheiro, W.C.; Bittencourt, B.E.; Luiz, L.B.; Marcello, L.A.; Antonio, V.F.; de Lira, P.H.A.; Stolf, R.G.; Castro, M.C.F. Parkinson's Disease Tremor Suppression. In Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017), Porto, Portugal,21 23 February 2017; pp. 149–155.

[113] Atzori, M.; Gijsberts, A.; Castellini, C.; Caputo, B.; Hager, A.G.M.; Elsig, S.; Giatsidis, G.; Bassetto, F.; Müller, H. Electromyography data for non-invasive naturally - controlled robotic hand prostheses. Sci. Data. 2014,1, 140053.

[114] Detti, P.; Vatti, G.; Zabalo Manrique de Lara, G. EEG Synchronization Analysis for Seizure Prediction: A Study on Data of Noninvasive Recordings. Processes **2020**, 8, 846.

[115] Detti, P. Siena Scalp EEG Database (Version 1.0.0), PhysioNet 2020. Available online: **https://doi.org/10.13026/5d4a-j060** (accessed on 1 February 2022).

[116] Kemp, B.; Zwinderman, A.H.; Tuk, B.; Kamphuisen, H.A.; Oberye, J.J. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* **2000**, *47*, 1185–1194.

[117] Pimentel, M.A.; Johnson, A.E.; Charlton, P.H.; Birrenkott, D.; Watkinson, P.J.; Tarassenko, L.; Clifton, D.A. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Trans. Biomed. Eng.* **2016**, *64*, 1914–1923.

[118] Gilles, J. Empirical Wavelet Transform. *IEEE Trans. Signal Process.* **2013**, *61*, 3999–4010, doi:10.1109/tsp.2013.2265222.

[119] Oung, Q.W.; Muthusamy, H.; Basah, S.N.; Lee, H.; Vijean, V. Empirical Wavelet Transform Based Features for Classification of Parkinson’s Disease Severity. *J. Med Syst.* **2017**, *42*, 29, doi:10.1007/s10916-017-0877-2.

[120] Qin, C.; Wang, D.; Xu, Z.; Tang, G. Improved Empirical Wavelet Transform for Compound Weak Bearing Fault Diagnosis with Acoustic Signals. *Appl. Sci.* **2020**, *10*, 682, doi:10.3390/app10020682.

[121] Alegria, O.C.; Valtierra-Rodriguez, M.; Amezquita-Sanchez, J.P.; Millan-Almaraz, J.R.; Rodriguez, L.M.; Moctezuma, A.M.; Dominguez-Gonzalez, A.; Cruz-Abeyro, J.A. Empirical Wavelet Transform-based Detection of Anomalies in ULF Geomagnetic Signals Associated to Seismic Events with a Fuzzy Logic-based System for Automatic Diagnosis. *Wavelet Transform and Some of Its Real-World Applications* **2015**, doi:10.5772/61163.

[122] Debbal, S.M.; Bereksi-Reguig, F. Computerized heart sounds analysis. *Comput. Boil. Med.* **2008**, *38*, 263–280, doi:10.1016/j.compbiomed.2007.09.006.

[123] Choi, S.; Jiang, Z. Comparison of envelope extraction algorithms for cardiac sound signal segmentation. *Expert Syst. Appl.* **2008**, *34*, 1056–1069, doi:10.1016/j.eswa.2006.12.015.

[124] Martínez-Alajarín, J.; Merino, R.R. Efficient method for events detection in phonocardiographic signals. *Microtechnologies for the New Millennium 2005* **2005**, *5839*, 398–409, doi:10.1117/12.608203.

[125] Deshpande, N. Assessment of systolic and diastolic cycle duration from speech analysis in the state of anger and fear; Academy and Industry Research Collaboration Center (AIRCC), 2012;.

[126] K.R. Aida–Zade, C. Ardil and S.S. Rustamov. Investigation of Combined use of MFCC and LPC Features in Speech Recognition Systems. International journal of signal processing volume 3 number 1. 2006.

[127] Features P.J.Arnott. G.W.Pfeiffer. M.E.Tavel. Spectral analysis of heart sounds: Relationships between some physical characteristics and frequency spectra of first and second heart sounds in normals and hypertensives. Journal of Biomedical Engineering Volume 6, Issue 2. 1984.

[128] Deng, L., O'Shaughnessy, D. “Analysis of Discrete-Time Speech Signal”. In Speech Processing: A Dynamic and Optimization-Oriented Approach. CRC Press, 2003, Chapter 2.

[129] VOICEBOX: Speech Processing Toolbox for MATLAB. Available in: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html. Last visit: 21-05-2017

[130] Machine Learning Group at the University of Waikato. Weka 3: Data Mining Software in Java [online]. Available in: http://www.cs.waikato.ac.nz/ml/weka/. Last visit: 21-05-2017

[131] Salima O., Asri N., Hamid H. Machine Learning Techniques for Anomaly Detection: An Overview. International Journal of Computer Applications, Volume 79 – No.2. 2013.

[132] J. Zhang, M. Zulkernine and A. Haque, "Random-Forests-Based Network Intrusion Detection Systems," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 5, pp. 649-659, Sept. 2008.

[133] Lee, S., Verri, A. Pattern Recognition with Support Vector Machines. Springer, 2002.

[134] Wang, L. Support Vector Machines: Theory and Applications. Springer Science & Business Media, 2005.

[135] Rajaguru, H., Kumar, S. “KNN Classifier”. In KNN Classifier and K-Means Clustering for Robust Classification of Epilepsy from EEG Signals. A Detailed Analysis. 2017.

[136] Breiman, L. Random Forests. Machine Learning (2001) 45: 5. doi:10.1023/A:1010933404324.

[137] Zhu, Wen & Zeng, Nancy & Wang, Ning. Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS ® Implementations. *North East SAS users group, health care and life sciences*. **2010**, *19*, 67

[138] Her, H., & Chiu, H. Using Time-Frequency Features to Recognize Abnormal Heart Sounds. Computing in Cardiology Conference (CinC). 2016.

[139] Probo, L., Nugroho, H., Wulandari, M. Feature Extraction and Classification of Heart Sound based on Autoregressive Power Spectral Density (AR-PSD). 1st International Conference on Information Technology, Computer and Electrical Engineering. 2014.

[140] Hadi, H. M., Mashor, M. Y., Mohamed, M. S., & Tat, K. B. Classification of Heart Sounds Using Wavelets and Neural Networks. 5th International Conference on Electrical Engineering, Computing Science and Automatic Control. 2008.

[141] Phatiwuttipat, P., Kongprawechon, W., Tungpimolrut, K., & Yuenyong, S. (2011). Cardiac auscultation analysis system with neural network and SVM technique. ECTI-CON 2011 - 8th Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI) Association of Thailand **-** Conference 2011, 1027–1030. https://doi.org/10.1109/ECTICON.2011.5948018

[142] Bouril, A., Aleinikava, D., Guillem, M. S., Mirsky, G. M., Llc, S., & València, U. Automated Classification of Normal and Abnormal Heart Sounds using Support Vector Machines. Computing in Cardiology Conference (CinC). 2016.

[143] Art Azmy, M. M. Classification of Normal and Abnormal Heart Sounds Using New Mother Wavelet and Support Vector Machines. 4th International Conference on Electrical Engineering (ICEE). 2015.

[144] Narvaez, P.; Vera, K.; Bedoya, N.; Percybrooks, W.S. Classification of heart sounds using linear prediction coefficients and mel-frequency cepstral coefficients as acoustic features. In Proceedings of the 2017 IEEE Colombian Conference on Communications and Computing (COLCOM); Institute of Electrical and Electronics Engineers (IEEE), 2017; pp. 1–6.

[145] Littmann Stethoscope. Heart sounds library. Available Online: http://solutions.3mae.ae/wps/portal/3M/en_AE/3M-Littmann-EMEA/stethoscope/littmann-learning institute/heart-lung-sounds/. Access on: 31-05-2018.

[146] Etoom, Y.; Ratnapalan, S. Evaluation of Children With Heart Murmurs. *Clin. Pediatr.* **2013**, *53*, 111–117, doi:10.1177/0009922813488653.

[147] Johnson W., Moller J. Pediatric cardiology: The essential pocket guide. Editorial: Wiley-Blackwell: Hoboken, NJ, USA. 2008.

[148] Gordon, E. Signal and Linear System Analysis. Editorial: Allied Publishers Limited: New Delhi, India. Page 386. 1994.

[149] Haikyn, S. Neural Networks and Learning Machines. Prentice Hall: Upper Saddle River, NJ, USA, Edition 3td. 2009.

[150] Johnson, E.M.; Cowie, B.; De Lange, W.P.; Falloon, G.; Hight, C.; Khoo, E. Adoption of innovative e-learning support for teaching: A multiple case study at the University of Waikato. *Australas. J. Educ. Technol.* **2011**, *27*, doi:10.14742/ajet.957.

[151] Narváez, P., Gutierrez, S., Percybrooks, W. “Automatic Segmentation and Classification of Heart Sounds using Modified Empirical Wavelet Transform and Power Features”. Applied Science MDPI. 2020.

[152] J. Hany and G. Walters. Hands-On Generative Adversarial Networks with PyTorch 1.x. 2019. Published by Packt publishing Ltda.

[153] M. Jablouna, P. Raviera, O. Buttelli, R. Ledeea, R. Harbaa, L. Nguyenb. A generating model of realistic synthetic heart sounds for performance assessment of phonocardiogram processing algorithms. Biomedical Signal Processing and Control. 2013.

[154] P. McSharry, G. Clifford, L. Tarassenko, and L. Smith. “A Dynamical Model for Generating Synthetic Electrocardiogram Signals”. IEEE Transactions on Biomedical Engineering, vol. 50, no. 3. March 2003.

[155] Di Persia, L., Yanagida, M., Rufiner, H., Milone,D. Objective quality evaluation in blind source separation for speech recognition in a real room. Signal Processing. August 2007.

[156] Vasilijevic, A., Petrinovic, D. Perceptual significance of cepstral distortion measures in digital speech processing. Automatika. 2011.

[157] Jeong, Dong Hyun & Jeong, Caroline & Ziemkiewicz, William & Ribarsky, Remco & Chang. Understanding Principal Component Analysis Using a Visual Analytics Tool. 2010.

[158] van der Maaten, Laurens & Hinton, Geoffrey. Visualizing data using t-SNE. Journal of Machine Learning Research. 9. 2579-2605. (2008).

[159] Melit Devassy, Binu & George, Sony. Dimensionality Reduction and Visualisation of Hyperspectral Ink Data Using t-SNE. Forensic Science International. 311. 110194. 10.1016/j.forsciint.2020.110194. (2020).

[160] Bakurov, Illya & Buzzelli, Marco & Schettini, Raimondo & Castelli, Mauro & Vanneschi, Leonardo. Structural Similarity Index (SSIM) Revisited: a Data-Driven Approach. Expert Systems with Applications. Doi: 189. 10.1016/j.eswa.2021.116087. (2021)

[161] Scikit-Image Python. https://scikit-image.org/docs/dev/auto_examples/transform/plot_ssim.html.

[162] Rix, Antony & Beerends, John & Kim, Doh-Suk & Kroon, Peter & Ghitza, Oded. (2006). Objective Assessment of Speech and Audio Quality—Technology and Applications. Audio, Speech, and Language Processing, IEEE Transactions on. 14. 1890 - 1901. 10.1109/TASL.2006.883260.