Synthesis of Normal Heart Sounds Using Generative Adversarial Networks and Empirical Wavelet Transform

Pedro Narváez * and Winston S. Percybrooks

Department of Electrical and Electronics Engineering, Universidad del Norte, Barranquilla 081001, Colombia; wpercyb@uninorte.edu.co

* Correspondence: pjnarvaez@uninorte.edu.co

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. 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 article, 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.

Keywords: generative adversarial network; heart sound classification; EWT; sound synthesis; machine learning; e-health

1. Introduction

Cardiovascular diseases are one of the leading causes of death in the world. According to recent reports from the World Health Organization and the American Heart Association, more than 17 million people die each year from these diseases. Most of these deaths (about 80%) occur in low- and middle-income countries [1,2]. 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].

This situation has motivated the development of computer classification models to support the identification of normal and abnormal heart sounds by non-specialized health professionals. To date, many investigations related to the analysis and synthesis of heart sounds (HSs) have been published, obtaining good results especially in the classification of normal and abnormal heart sounds [9–25].

Heart sounds are closely related to both, the vibration of the entire myocardial structure and the vibration of the heart valves during closure and opening. A recording of heart sounds is composed of series of cardiac cycles. A normal cardiac cycle, as shown in Figure 1, is composed by the S1 sound (generated by the closing of the atrioventricular valve), the S2 sound (generated by the closing of the semilunar valve), the systole (range between S1 and S2), and the diastole (range included between S2 and S1 of the next cycle) [26]. Abnormalities are represented by murmurs that usually occur in the systolic or diastolic interval [27]. Health professionals use different attributes of heart murmurs for their classification, the most common are: timing, cadence, duration, pitch, and shape of the murmur [28,29]. Therefore the examiner must have an ear sufficiently trained to identify each of these attributes.
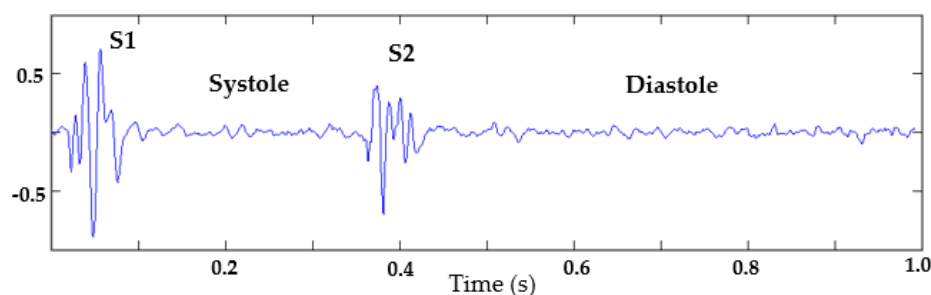

Figure 1. Example signal for a normal heart sound.

Despite the amount of work related to the classification of heart sounds, it is still difficult to statistically evaluate the robustness of these algorithms, since the number of samples used in training is not sufficient to guarantee a generalized model. Similarly, there have been no significant advances in the classification of specific types of heart murmurs, due to the limited availability of corresponding labels in current public databases [30,31]. 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. 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 [16–25]. All these works are based on mathematical models to generate the S1 and S2 sections of a cardiac cycle. 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. Table 1 presents a comparison of the different heart sound generative methods found in the Web of Science, Scopus, and IEEE Xplore databases.

Table 1. Timeline of methods for the generation of synthetic heart sounds.

Year	Author	Previous Method
1992	Tang et al. [16]	Exponentially damped sinusoidal model
1995	Trang et al. [17]	S1 and S2 as transient–linear chirp signals
1998	Zhang et al. [18]	Sum of Gaussian-modulated sinusoidal
2000	Xu et al. [19]	S1 and S2 as transient–nonlinear chirp signal
2009	Toncharoen et al. [20]	A heart-sound-like chaotic attractor
2011	Almasi et al. [21]	A dynamical model based in electrocardiogram (ECG) signal
2012	Tao et al. [22]	Amplitude and width modification of S1 and S2 sounds from real heart sounds, and combining them with noise
2013	Jablouna et al. [23]	Coupled ordinary differential equations
2018	Saederup [24]	Estimation of the second heart-sound split using windowed sinusoidal models
2018	Joseph et al. [25]	Sum of almost periodically recurring deterministic "wavelets"; S1 and S2 are modeled by two sinusoidal pulses of Gaussian modulation.

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.

Table 2. Previous methods for generation of synthetic audio.

Year	Author	Previous Method	Synthetic Signal
2015	van den Oord et al. [32]	WaveNet: Probabilistic and autoregressive model based on deep neural networks (DNNs)	Music; text-to-speech
2017	Engel et al. [33]	WaveNet and autoenconders	Musical notes
2017	Bollepalli et al. [34]	Generative adversarial network (GAN)	Glottal waveform
2018	Biagetti et al. [35]	Hidden Markov model	Text-to-speech
2019	Donahue et al. [36]	WaveGAN: GANs unsupervised	Intelligible words

According to the limitations presented in the currently proposed 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 only normal heart sounds that can be used to train machine learning models. This article is organized as follows: Section 2 presents a definition of the GAN architecture; in Section 3, the proposed method is described; Section 4 presents experimental results, using mel–cepstral distortion (MCD) and heart sound classification models; finally, the conclusions and discussions of the proposed work and experiments are presented in Section 5.

2. Generative Adversarial Networks (GANs)

Generative adversarial networks (GANs) are architectures of deep neural networks widely used in the generation of synthetic images [37]. This architecture is composed of two neural networks that face each other, called the generator and discriminator [38]. In Figure 2, a general diagram of a GAN architecture is presented.
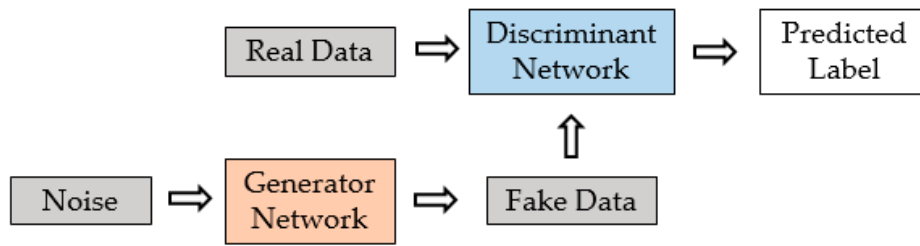
Figure 2. General diagram of a GAN.

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 realistically 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 [38].

The generator is an inverse convolutional network—that is, it takes a random noise vector and converts it into an image—unlike the discriminator, which takes an image and samples it to produce a probability. In other words, the discriminator (D) and generator (G) play the following two-player min/max game with value function L(G, D), as described in Equation (1):

$$\min_{G} \max_{D} L(D, G) = \mathbb{E}_{x \sim \rho_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim \rho_{data}(z)}[\log(1 - D(G(z)))] \tag{1}$$

where $D(x)$ represents the probability that x is estimated by the discriminator, z represents the input of random variables from the generator, and $\rho_{data}(x)$ 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.

3. Proposed Method

The proposed method is made up of two main stages, as shown in Figure 3. 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 an 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 it 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 3. General diagram of proposed method.

Figure 4 shows the diagram of the implemented GAN architecture, and each of its components is described below.

- 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 5 shows a diagram of the generating network, and the hyperparameters of each layer are specified. 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, a kernel size of 3, and a stride of 1. Finally, there is a dense layer with a 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 6 shows a diagram of the discriminator network. It begins with a dense layer with ReLu activation function, followed by four convolutional layers with filters of size 256, 128, 64, and 32, respectively; each of these layers uses Leaky ReLu activation function, a kernel size of 3, and a stride of 1; additionally, between each convolution layer there is a dropout of 0.25. Finally, there is a dense layer with a 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;

- Dataset of heart sounds: 100 normal heart sounds obtained from the Physionet database [30] were used, with a sampling frequency of 2 KHz and 1 s 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 was 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 were set;

- Loss function: a binary, cross-entropy function weas used in this work. This function computes the cross-entropy loss between true labels and predicted labels.
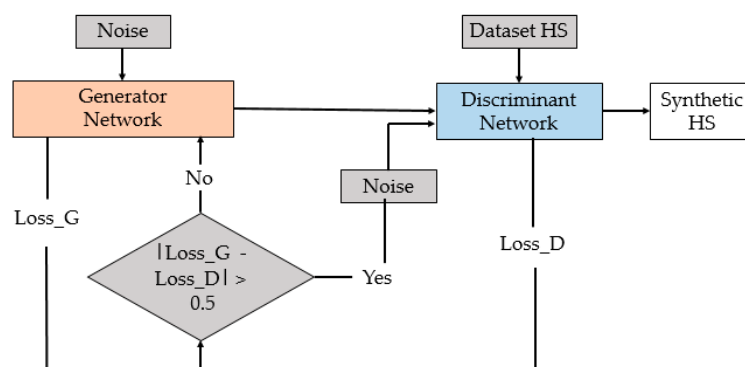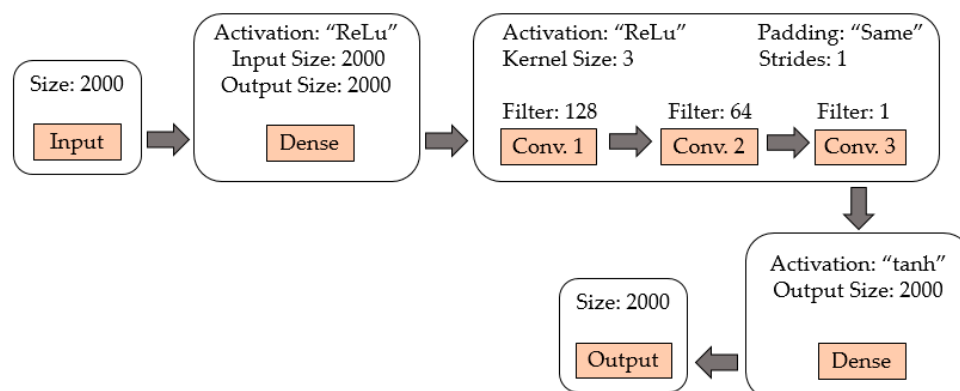

Figure 4. Proposed GAN diagram.

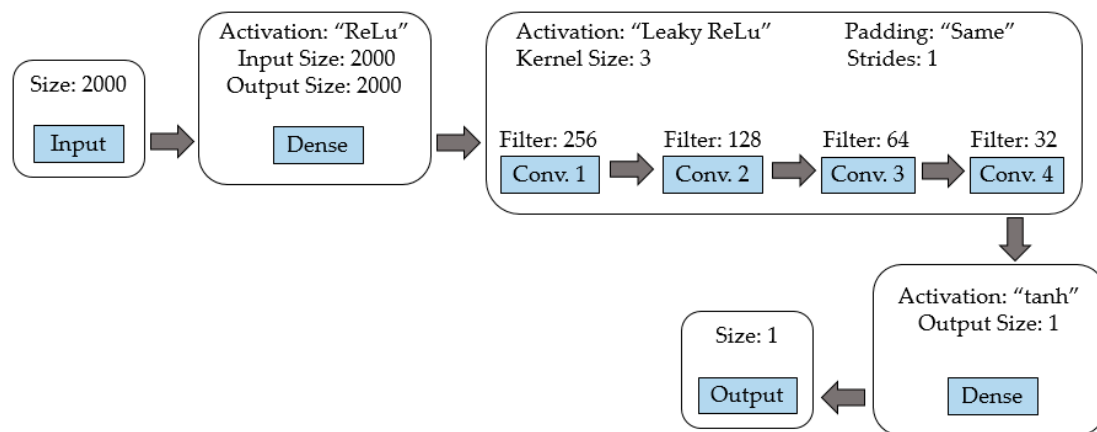**Figure 5.** Architecture of the generator model.

Figure 6. Architecture of the discriminator model.

Subsequently, the difference between generator and discriminator losses was analyzed. If this difference was greater than 0.5, the input data to the discriminator was switched to a Gaussian noise with a mean of 0 and standard deviation of 1, and not the generator output as it was otherwise. With this method, a convergence in the loss functions of the generator and discriminator could 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 [39]. 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 [40], 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 [40]. This algorithm has been used in different signal processing applications [41–43]. In [9], 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 [9]. In this work, it was decided to use as a reference the method proposed in [9] 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 [44], 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.

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 was 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. Figure 7A,B shows, respectively, a real heart sound and the same heart sound mixed with low-amplitude Gaussian noise, while Figure 7C,D shows their respective Fourier transforms (FFTs). In this last figure, the frequency of the components between 0 and 200 Hz can be seen in blue, and the frequency of components above 200 Hz caused by the Gaussian noise are in green. The signal shown in Figure 7B was then used as an input example to the proposed EWT algorithm to illustrate its de-noising action. Figure 7D shows the two frequency bands extracted with EWT using the FFT, while Figure 7E,F shows the two components extracted from the noisy signal in the time domain. As can be seen, the signal obtained in Figure 7E presents a lower noise level, and is comparable with the original cardiac signal shown in Figure 7A.
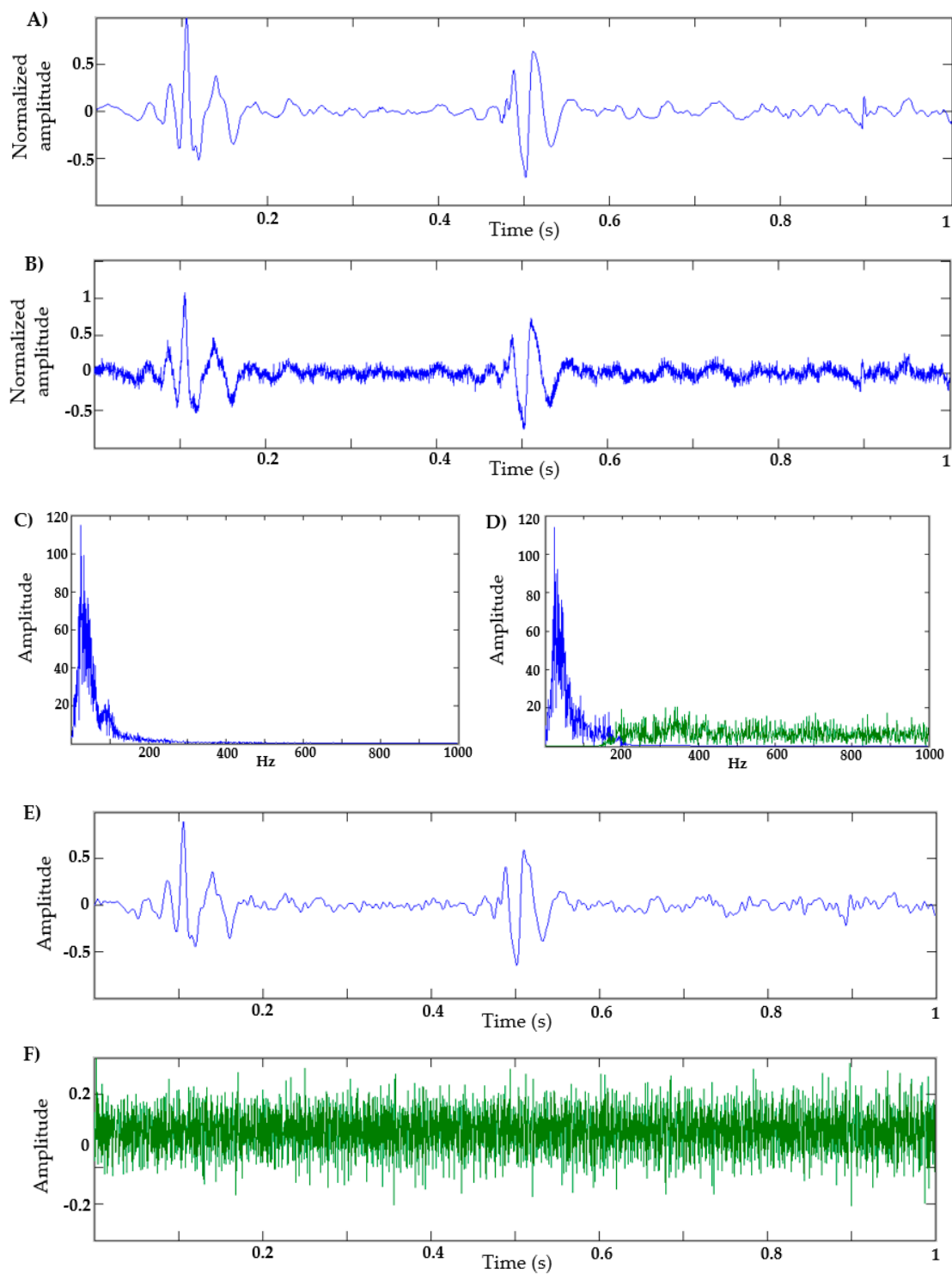
Figure 7. (**A**) Real heart sound; (**B**) real heart sound with Gaussian noise; (**C**) Fourier transform (FFT) of real heart sound; (**D**) FFT of heart sound with Gaussian noise; (**E**) empirical wavelet transform (EWT) component of the noisy signal in the frequency range of 0–200 Hz; (**F**) EWT component of the noisy signal in the frequency range greater than 200 Hz.

4. Experiments and Results

The proposed GAN model was trained for a total of 2000 epochs. Figure 8 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 12,000 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 8E. Therefore, it was decided to generate the synthetic signals up to 2000 epochs, and subsequently apply the proposed EWT algorithm. Figure 8F,G shows the results of a synthetic signal generated with 12,000 training epochs without applying an EWT algorithm or a natural signal, respectively.
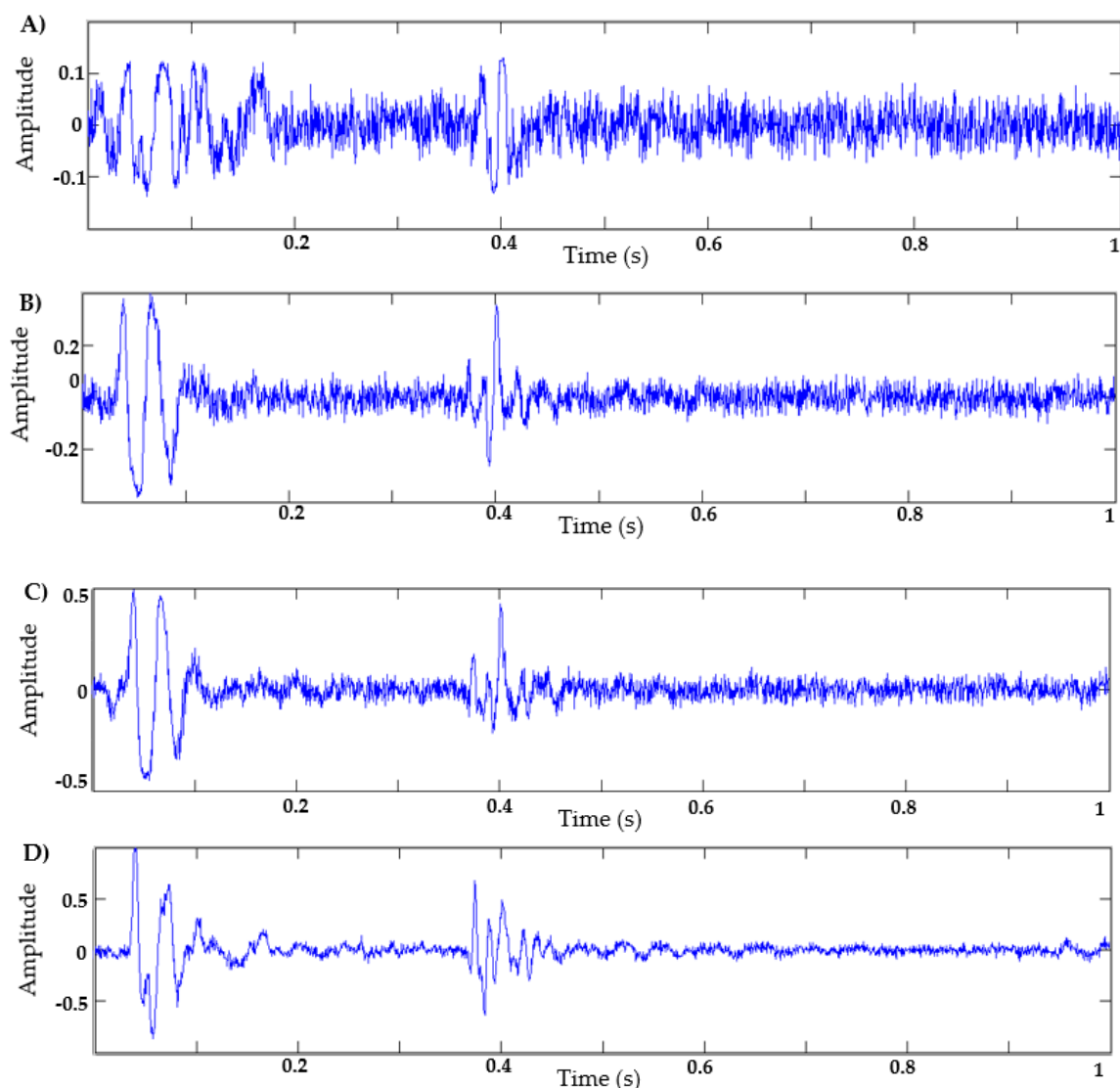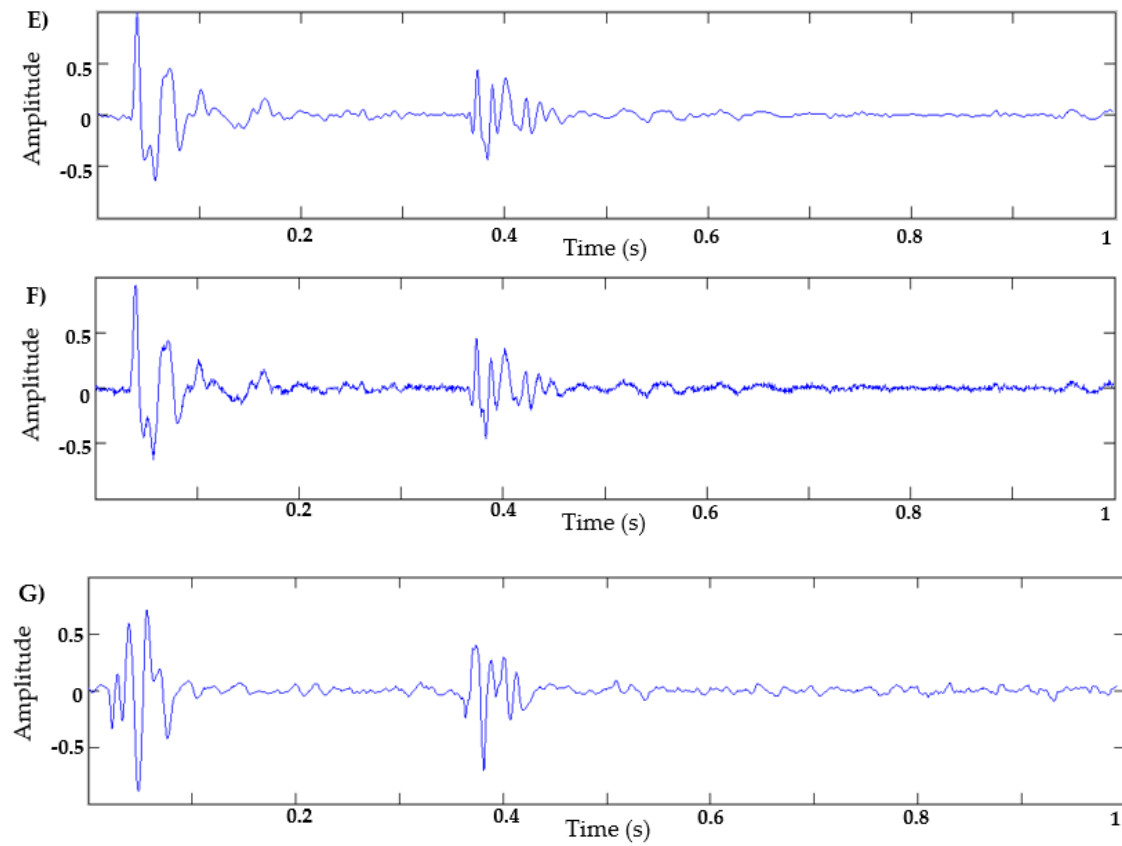

Figure 8. *Cont.*

Figure 8. (**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 12,000 epochs, and (**G**) natural signal of heart sound.

In this work, a comparison is made between the proposed method and a mathematical model proposed in [21], in order to determine which method generates a cardiac signal realistic enough to be used by a classification model. The mathematical method [21] was inspired by a dynamic model that generates a synthetic electrocardiographic signal, as described [45]. Therefore, this model is based on the morphology of the phonocardiographic signal, and has been used as a reference in other proposed methods [23]. The equation for the reference model [21] is described below:

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

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 [21] are summarized in Table 3.

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 3, we obtain the graph in Figure 9A, which represents the S1 and S2 sounds of a cardiac cycle. Figure 9B shows a natural signal of heart sound.

Table 3. Parameters used in [21] 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)	$3\pi/4$	$9\pi/11$	$7\pi/8$	$3\pi/4$
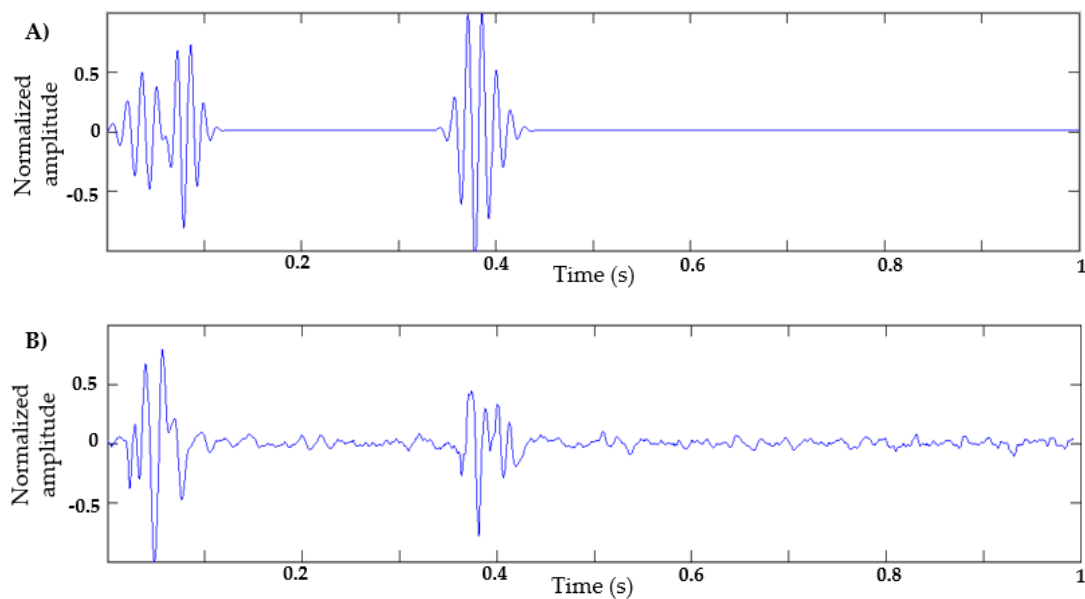
Figure 9. (**A**) Synthetic heart sound generated by the model [21]; (**B**) natural signal of the heart sound.

4.1. Results Using Mel–Cepstral Distortion (MCD)

Mel–cepstral distortion (MCD) is a metric widely used to objectively evaluate audio quality [46], 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 [46]. 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 cepstra 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 [47], different ways of calculating this distortion are presented. In Equation (3), the formula used in [46] is defined, where C_{MFCC} and $C_{\hat{MFCC}}$ 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.

$$MCD_{FRAME} = \sum_{l=1}^{L} \left(C_{MFCC}[l] - C_{\hat{MFCC}}[l] \right)^2 \tag{3}$$

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 [9–12].

A set of 400 natural normal heart sounds taken from the Physionet [30] and Pascal [31] databases were used. Each signal was cut to a single cardiac cycle, with a 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 [21]. In the case of the model [21], 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 [21]. 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 10 shows a schematic of the procedure to compute the MCD. Figure 11 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 [21] (green color) using the MCD method.
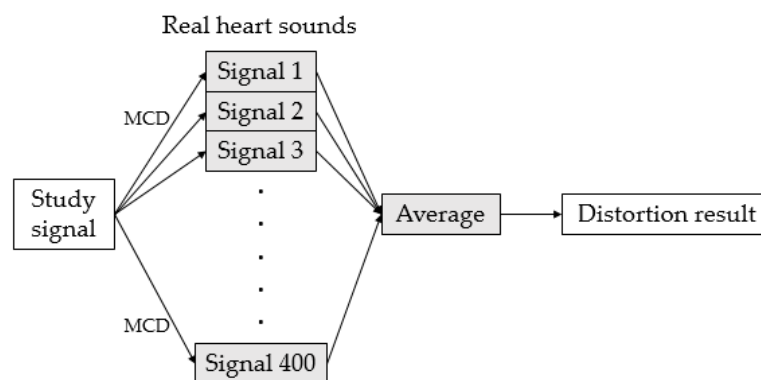

Figure 10. General diagram of the procedure to calculate the signal distortion.
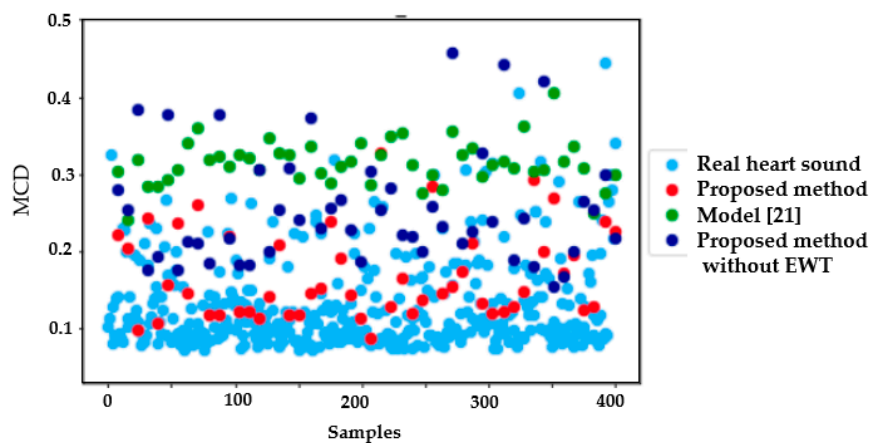

Figure 11. Result of mel–cepstral distortions (MCDs).

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 12 shows the resulting MCD values, with and without the EWT postprocessing. It can be seen that in none of the cases is there an MCD value equal to approaching zero; therefore, the generated signal is not a copy of any of the training inputs.
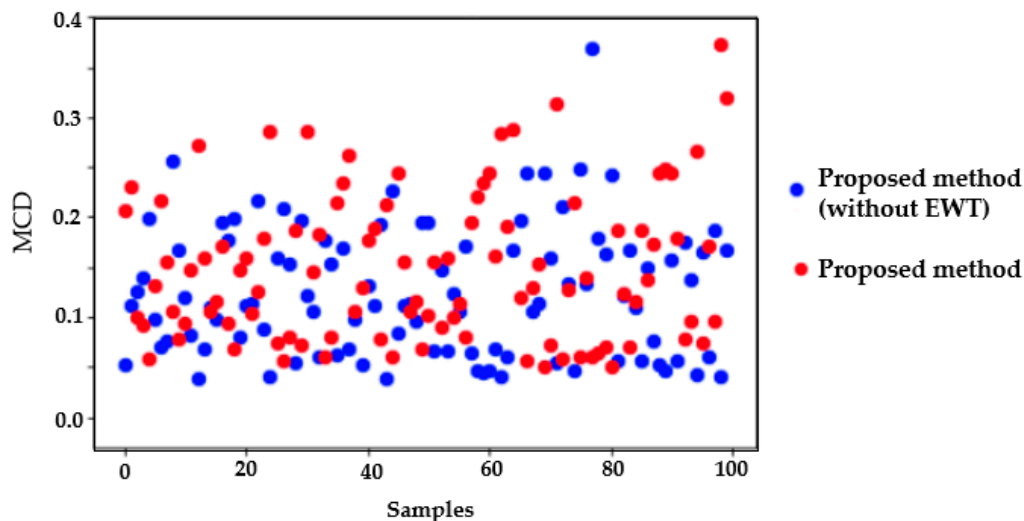

Figure 12. Result of MCD distortion using one synthetic signal with training dataset.

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 [21].

4.2. Results Using Classification Models

In this section, different heart sound classification models published in the state of the art are tested [9–12]. 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 [30], Pascal challenge database [31], Database of the University of Michigan [48], Database of the University of Washington [49], Thinklabs database (digital stethoscope) [50], and 3M database (digital stethoscope) [51].

Table 4 presents the different characteristics extracted in the proposed classification methods [9–12]. These characteristics belong to the domains of time, frequency, time–frequency, and perception.

Table 4. Features extracted in models [9–12].

Reference	Features
[9]	Six power values (three in systole and three in diastole)
[10]	Nineteen mel–frequency cepstral coefficients (MFCCs) and 24 discrete wavelet transform features
[11]	Statistical domain: mean value, median value, standard deviation, mean absolute deviation, quartile 25, quartile 75, inter-quartile range (IQR), skewness, kurtosis, and coefficient of variation. Frequency domain: entropy, dominant frequency value, dominant frequency magnitude, and dominant frequency ratio. Perceptual domain: 13 MFCCs.
[12]	Six linear prediction coefficient (LPCs) + 14 MFCCs 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 (KNNs), random forest (RF), and multilayer perceptron (MLP).

In Table 5, 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 [9].

Table 5. Accuracy results of the methods proposed in [9–12], taken from article [9].

Feature Extraction	Classifier			
	SVM	KNN	RF	MLP
EWT + power [9]	92.42%	99.25%	99.00%	98.63%
MFCC + DWT [10]	90.68%	91.18%	91.42%	91.55%
Statistical, frequency, and perceptual [11]	84.47%	93.66%	93.66%	92.54%
LPC + MFCC [12]	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 6. The same procedure was done with the synthetic signals without applying the EWT algorithm, with the accuracy results presented in Table 7.

Table 6. Accuracy results of synthetic signals, using the trained models proposed in articles [9–12].

Feature Extraction	Classifier			
	SVM	KNN	RF	MLP
EWT + power [9]	100%	100%	100%	100%
MFCC + DWT [10]	80%	90%	78%	82%
Statistical, frequency, and perceptual [11]	98%	78%	78%	60%
LPC + MFCC [12]	98%	96%	96%	82%

Table 7. Accuracy results of synthetic signals without applying an EWT algorithm, using the trained models proposed in articles [9–12].

Feature Extraction	Classifier			
	SVM	KNN	RF	MLP
EWT + power [9]	100%	100%	100%	100%
MFCC + DWT [10]	85%	88%	80%	90%
Statistical, frequency, and perceptual [11]	90%	78%	75%	60%
LPC + MFCC [12]	95%	90%	90%	78%

The best results were obtained with the power characteristics proposed in [9]. However, in several combinations of characteristics and ML models, precision results greater than 90% were obtained, as was the case of the combination of LPC and MFCC proposed in [12]. 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. Conclusions and Future Work

A GAN-based architecture was 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 us 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 [21]. Two evaluation tests were carried out: the first was 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, 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 [21].

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 [9] 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 4.

According to the results obtained using the MCD distortion metric, and the performance of different ML models, a strong indication can be seen that synthetic signals can be used to improve the performance of heart sound classification models, since the number of samples in the training could be increased.

As future work, we are implementing a GAN-based model that can generate abnormal types of heart sounds, delving into the generation of heart murmurs, since it is very difficult to acquire many samples of a specific type of abnormality. By creating a database of normal synthetic heart sounds and types of abnormalities, we hope to improve the performance of classification systems and advance the detection of types of heart abnormalities, generating significant support for healthcare professionals.

References

1. World Health Organization. A Global Brief on Hypertension. Available online: http://www.who.int/cardiovascular_diseases/publications/global_brief_hypertension/en/ (accessed on 15 September 2020).
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. *Circulation* **2017**, *135*, 146–603. [CrossRef]
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. Narváez, P.; Gutierrez, S.; Percybrooks, W. Automatic Segmentation and Classification of Heart Sounds using Modified Empirical Wavelet Transform and Power Features. *Appl. Sci.* **2020**, *10*, 4791. [CrossRef]
10. Yaseen; Gui-Young, S.; Kwon, S. Classification of Heart Sound Signal Using Multiple Features. *Appl. Sci.* **2018**, *8*, 2344. [CrossRef]
11. Arora, V.; Leekha, R.; Singh, R.; Chana, I. Heart sound classification using machine learning and phonocardiogram. *Mod. Phys. Lett. B.* **2019**, *33*, 1950321. [CrossRef]
12. 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.

13. Noman, F.; Ting, C.; Salleh, S.; Ombao, H. Short-segment heart sound classification using an ensemble of deep convolutional neural networks. In Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Brighton, UK, 12 May 2019.

14. Raza, A.; Mehmood, A.; Ullah, S.; Ahmad, M.; Sang, G.; Byung-Won, O. Heartbeat Sound Signal Classification Using Deep Learning. *Sensors* **2019**, *19*, 4819. [CrossRef]

15. Abdollahpur, M.; Ghaffari, A.; Ghiasi, S.; Mollakazemi, M.J. Detection of pathological heart sound. *Physiol. Meas.* **2017**, *38*, 1616–1630.

16. 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]

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

18. 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]

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. McConnell M., E.; Branigan, A. *Pediatric Heart Sounds*; Springer: London, UK, 2008. [CrossRef]

27. Brown, E.; Leung, T.; Collis, W.; Salmon, A. *Heart Sounds Made Easy*, 2nd ed.; Churchill Livingstone Elsevier: Philadelphia, PA, USA, 2008.

28. Etoom, Y.; Ratnapalan, S. Evaluation of Children With Heart Murmurs. *Clin. Pediatr.* **2013**, *53*, 111–117. [CrossRef]

29. Johnson, W.; Moller, J. *Pediatric Cardiology: The Essential Pocket Guide*; Wiley-Blackwell: Oxford, UK, 2008.

30. PhysioNet/Computing in Cardiology Challenge. Classification of Normal/Abnormal Heart Sound Recordings. Available online: https://www.physionet.org/challenge/2016/1.0.0/ (accessed on 15 September 2020).

31. Bentley, P.; Nordehn, G.; Coimbra, M.; Mannor, S.; Getz, R. Classifying Heart Sounds Callenge. Available online: http://www.peterjbentley.com/heartchallenge/#downloads (accessed on 15 September 2020).

32. 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).

33. 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.

34. 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. [CrossRef]

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

36. Chrism, D.; Julian, M.; Miller, P. Adversarial Audio Synthesis. Available online: https://arxiv.org/abs/1802.04208 (accessed on 15 September 2020).

37. Huang, H.; Yu, P.S.; Wang, C. An Introduction to Image Synthesis with Generative Adversarial Nets. *arXiv* **2018**, arXiv:1803.04469.

38. Goodfellow, J.I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Neural Information Processing Systems 27, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.

39. Hany, J.; Walters, G. *Hands-On Generative Adversarial Networks with PyTorch 1.x.*; Pckt publishing Ltd.: Birmingham, UK, 2019.

40. Gilles, J. Empirical wavelet transform. *IEEE Trans. Signal Process.* **2013**, *61*, 3999–4010.

41. Oung, Q.; Muthusamy, H.; Basah, S.; Lee, H.; Vijean, V. Empirical Wavelet Transform Based Features for Classification of Parkinson's Disease Severity. *J. Med. Syst.* **2017**, *42*, 29.

42. 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.

43. Chavez, O.; Dominguez, A.; Valtierra-Rodriguez, M.; Amezquita-Sanchez, J.P.; Mungaray, A.; Rodriguez, L.M. Empirical Wavelet Transform-based Detection of Anomalies in ULF Geomagnetic Signals Associated to Seismic Events with a Fuzzy Logic-based System for Automatic Diagnosis. In *Wavelet Transform and Some of Its Real-World Applications*; InTech: Rijeka, Croatia, 2015.

44. Debbal, S.M.; Bereksi-Reguig, F. Computerized Heart Sounds Analysis. *Comput. Biol. Med.* **2008**, *38*, 263–280. [CrossRef]

45. McSharry, P.; Clifford, G.; Tarassenko, L.; Smith, L. A Dynamical Model for Generating Synthetic Electrocardiogram Signals. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 289–294. [PubMed]

46. Di Persia, L.; Yanagida, M.; Rufiner, H.; Milone, D. Objective quality evaluation in blind source separation for speech recognition in a real room. *Signal Process.* **2007**, *87*, 1951–1965.

47. Vasilijevic, A.; Petrinovic, D. Perceptual significance of cepstral distortion measures in digital speech processing. *Automatika* **2011**, *52*, 132–146.

48. University of Michigan. Heart Sound and Murmur Library. Available online: https://open.umich.edu/find/open-educational-resources/medical/heart-sound-murmur-library (accessed on 15 September 2020).

49. University of Washington. Heart Sound and Murmur. Available online: https://depts.washington.edu/physdx/heart/demo.html (accessed on 15 September 2018).

50. Thinklabs. Heart Sounds Library. Available online: http://www.thinklabs.com/heart-sounds (accessed on 15 September 2020).

51. Littmann Stethoscope. Heart Sounds Library. Available online: www.3m.com/healthcare/littmann/mmm-library.html (accessed on 15 September 2020).