

Article

Deep Learning-Based Heart Sound Analysis for Left Ventricular Diastolic Dysfunction Diagnosis

Yang Yang ^{1,*}, Xing-Ming Guo ^{1,*} , Hui Wang ¹ and Yi-Neng Zheng ^{2,*}

¹ Key Laboratory of Biorheology Science and Technology, Ministry of Education, College of Bioengineering, Chongqing University, Chongqing 400044, China; yangyang99@cqu.edu.cn (Y.Y.); huiw@cqu.edu.cn (H.W.)

² Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

* Correspondence: guoxm@cqu.edu.cn (X.-M.G.); yinengzheng@cqmu.edu.cn (Y.-N.Z.); Tel.: +86-23-65106820 (X.-M.G.); +86-23-89012258 (Y.-N.Z.)

Abstract: The aggravation of left ventricular diastolic dysfunction (LVDD) could lead to ventricular remodeling, wall stiffness, reduced compliance, and progression to heart failure with a preserved ejection fraction. A non-invasive method based on convolutional neural networks (CNN) and heart sounds (HS) is presented for the early diagnosis of LVDD in this paper. A deep convolutional generative adversarial networks (DCGAN) model-based data augmentation (DA) method was proposed to expand a HS database of LVDD for model training. Firstly, the preprocessing of HS signals was performed using the improved wavelet denoising method. Secondly, the logistic regression based hidden semi-Markov model was utilized to segment HS signals, which were subsequently converted into spectrograms for DA using the short-time Fourier transform (STFT). Finally, the proposed method was compared with VGG-16, VGG-19, ResNet-18, ResNet-50, DenseNet-121, and AlexNet in terms of performance for LVDD diagnosis. The result shows that the proposed method has a reasonable performance with an accuracy of 0.987, a sensitivity of 0.986, and a specificity of 0.988, which proves the effectiveness of HS analysis for the early diagnosis of LVDD and demonstrates that the DCGAN-based DA method could effectively augment HS data.

Keywords: left ventricular diastolic dysfunction; deep convolutional generative adversarial networks; heart sounds; convolutional neural network; diagnosis



Citation: Yang, Y.; Guo, X.-M.; Wang, H.; Zheng, Y.-N. Deep Learning-Based Heart Sound Analysis for Left Ventricular Diastolic Dysfunction Diagnosis. *Diagnostics* **2021**, *11*, 2349. <https://doi.org/10.3390/diagnostics11122349>

Academic Editor: Jae-Ho Han

Received: 15 November 2021

Accepted: 10 December 2021

Published: 13 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Left ventricular diastolic dysfunction (LVDD) is a clinical syndrome characterized by inadequate active relaxation and decreased cardiac output, resulting in elevated end-diastolic pressure and possibly alterations in cardiac function [1]. At present, the overall incidence of LVDD in the general population is approximately 30% [2–5] and has a positive correlation with all-cause mortality. When the manifestations of heart failure (HF) such as dyspnea, edema, and fatigue gradually appear in the LVDD population, it will move towards an irreversible stage and cause progression to HFpEF [6,7], and around half of all HF hospital admissions are accounted for by patients with HFpEF [8]. Therefore, early diagnosis of LVDD is of great significance for preventing the deterioration of cardiac function as well as timely treatment.

LVDD is the result of impaired diastolic function, which expands inadequately the heart lumen and causes insufficient blood to return to the heart. Invasive measurements, such as stiffness index, could be used to evaluate diastolic function, but they are harmful to the human body. At present, echocardiography is the most commonly used non-invasive examination in clinical practice [9], but it cannot effectively detect the early LVDD without organic diseases. Cardiac auscultation is another non-invasive method used in clinical detection of cardiovascular disease, while the application of heart sounds (HS) is limited due to the limited range of human hearing and the high requirement on physicians'

experience [10]. The electronic transducer can digitally record HS, laying the groundwork for subsequent processing and application of HS signal. Furthermore, HS can directly reflect the property of cardiac mechanical activity and provide useful information for early diagnosis of cardiac abnormalities [11].

At present, there are two types of frequently utilized approaches for evaluating cardiovascular diseases utilizing HS: traditional machine learning and deep learning methods. The former need to extract features as input, such as wavelet features [12], Mel-frequency cepstral coefficients [13,14], energy entropy [15], etc. Saraf et al. [16] offered five physiologically-motivated characteristics extracted from HS signals, which might be used to quantify LVDD using echocardiography-like criteria. These characteristics may better reflect the physical meaning of HS. However, selecting useful features requires manual interaction, and machine learning models were built using a small database. Therefore, the robustness cannot be verified. Deep learning, a machine learning extension, has been popular as artificial intelligence has progressed due to its remarkable ability to extract features automatically. For example, He et al. [17] identified normal and abnormal HS in PhysioNet/CinC Challenge 2016 based on a convolutional neural network (CNN). Also on this database, Noman et al. [18] proposed an one-dimensional (1-D) CNN model to learn deep features directly from original HS signals, indicating that CNN has shown great potential in HS classification and diagnosis.

Although the above studies show that cardiovascular diseases can be early diagnosed using HS, the relevant databases are still scarce [19]. Furthermore, the amount of data has a direct influence on the performance of deep learning models. The over-fitting problem will occur if the amount of data is small, resulting in poor performance. Data augmentation (DA) [20] is an elegant solution to this problem. Due to the rigorous requirements for experimental individuals and surroundings, collecting large-scale and high-quality HS signals of LVDD is difficult in this study. As a result, a DA strategy for LVDD diagnosis is necessary.

The generative adversarial networks (GAN) [21] model was used to generate normal HS [22], but it has problems of mode collapse and training instability [23], which often lead to meaningless outputs [24]. Therefore, some researchers have refined it to create a variety of derivatives that have been shown to be successful in generating medical images [25,26]. In addition, these derivatives also have shown potential in generating time series data, such as artificial audio [27] and electroencephalogram signals [28]. The representation of time series data in the two-dimensional (2D) form may be beneficial to machine learning tasks [29,30]. Therefore, to meet the requirements of deep learning models for the data size, we converted the HS signals into 2D spectrograms by short-time Fourier transform (STFT) [31]. Then, a deep convolutional generative adversarial networks (DCGAN) model, a combination of CNN and GAN, was proposed to automatically generate spectrograms for the HS signals.

This paper's main contributions are as follows:

- LVDD is a common pathophysiological route leading to various cardiovascular diseases [32], and therefore the analysis of HS to diagnose LVDD is a non-invasive method. In this paper, the real HS signals collected by a digital transducer are binary classified automatically using a 2D CNN model for LVDD diagnosis. In addition, the performance of the CNN model in the three datasets is compared to six deep learning models to find the best model.
- Owing to the restricted availability of relevant labels in prevailing public HS databases, there have been no substantial breakthroughs in categorizing specific forms of heart disease [33]. Therefore, this paper presented a DCGAN-based DA method to expand HS database of LVDD, which could establish an effective dataset for deep learning, drastically minimize the cost of real data collection, and improve LVDD diagnosis performance.

2. Materials and Methods

The flow diagram of the proposed method is depicted by Figure 1 for automatically diagnosing LVDD, which is explained in the following subsections.

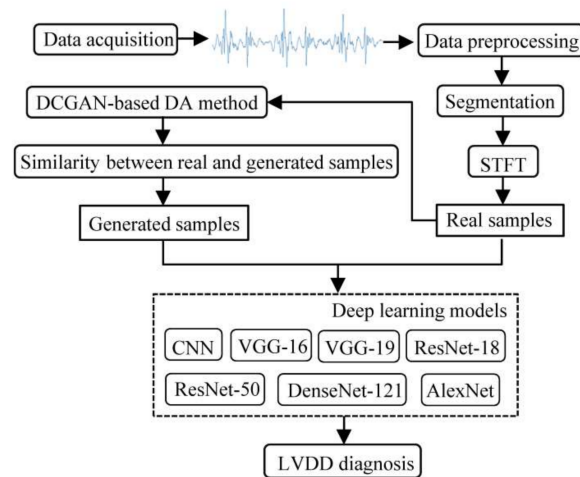


Figure 1. A flow diagram of this paper. The CNN is the proposed model, and the others are the compared models.

2.1. Data Acquisition

All subjects were assessed by cardiologists and completed informed consent forms prior to participation in this study. The day before HS signal collection, the echocardiographic examinations were performed with those patients in the supine position by professional doctors, and the equipment type is ACUSON X300 with the P2-4 phased array ultrasound probe. All subjects had their left ventricular ejection fraction (LVEF) assessed using the biplane Simpson method in the apical four-chamber view. Some Doppler echocardiography indices are shown in Table 1.

Table 1. Some Doppler echocardiography indices (mean \pm standard deviation).

Index	LVDD Group	Control Group	<i>p</i> -Value
LVEF	0.45 \pm 0.16	0.64 \pm 0.03	0.000 *
Peak E-wave velocity(cm/s)	73.81 \pm 26.60	63.25 \pm 14.07	0.034 *
Peak A-wave velocity(cm/s)	72.58 \pm 27.48	75.53 \pm 18.07	0.589
Septal e' velocity(cm/s)	4.18 \pm 1.40	6.20 \pm 2.09	0.000 *
Average E/e'	18.57 \pm 6.55	10.66 \pm 2.00	0.000 *
LA volume index (mL/m ²)	40.68 \pm 15.49	23.52 \pm 5.48	0.000 *
TR velocity(m/s)	28.74 \pm 6.62	16.79 \pm 7.26	0.000 *

* *p*-value < 0.05 demonstrates statistically significant difference.

A total of 92 HS signals were acquired from the First Affiliated Hospital of Chongqing Medical University by the multi-channel physiological signal acquisition system (RM-6240BD) and HS transducer (XJ102) with the sampling frequency at 8000 Hz, recorded on the apex position of the heart. The HS transducer utilizes piezoelectric-transducer with high-sensitivity, which converts the vibratory HS signal into electrical signal. Then, it is processed by the preamplifier, bandpass filter, 50Hz notch filter, and other circuits of the acquisition system and sent to the single-chip microcomputer for storage after analog-to-digital (A/D) conversion. Finally, the acquisition system is connected to the computer, and the waveform is exhibited on the LCD screen. The collection time was approximately five minutes.

According to the recommendations for the evaluation of left ventricular diastolic function by echocardiography from the American Society of Echocardiography and the European Association of Cardiovascular Imaging (ASE/EACVI) [34], the subjects were

divided into: (i) LVDD group, 30 subjects with LVDD, including G1, G2, G3; (ii) control group, 41 subjects with normal diastolic dysfunction, G0. The remaining 21 HS signals could not determine LVDD and were excluded from the study, GIND.

2.2. Data Preprocessing

The first HS (S1), systole (S), the second HS (S2), and diastole (D) are the four states of a cardiac period. Among them, the pitch of S2 is higher than that of S1, and the frequency range is 20–200 Hz [35]. To improve signal processing efficiency without sacrificing the key components of HS, firstly, the signals were resampled into 1000 Hz according to the Nyquist sampling theorem. The redundant noise was then removed using an improved wavelet denoising method [36]. Finally, the amplitude of the denoised signal $S(n)$ was normalized. The denoised waveform can be seen in the upper half of Figure 2.

$$S_{norm}(n) = \frac{S(n)}{\max_{n=1}^N (|S(n)|)}, \quad (1)$$

where N is the length of signal.

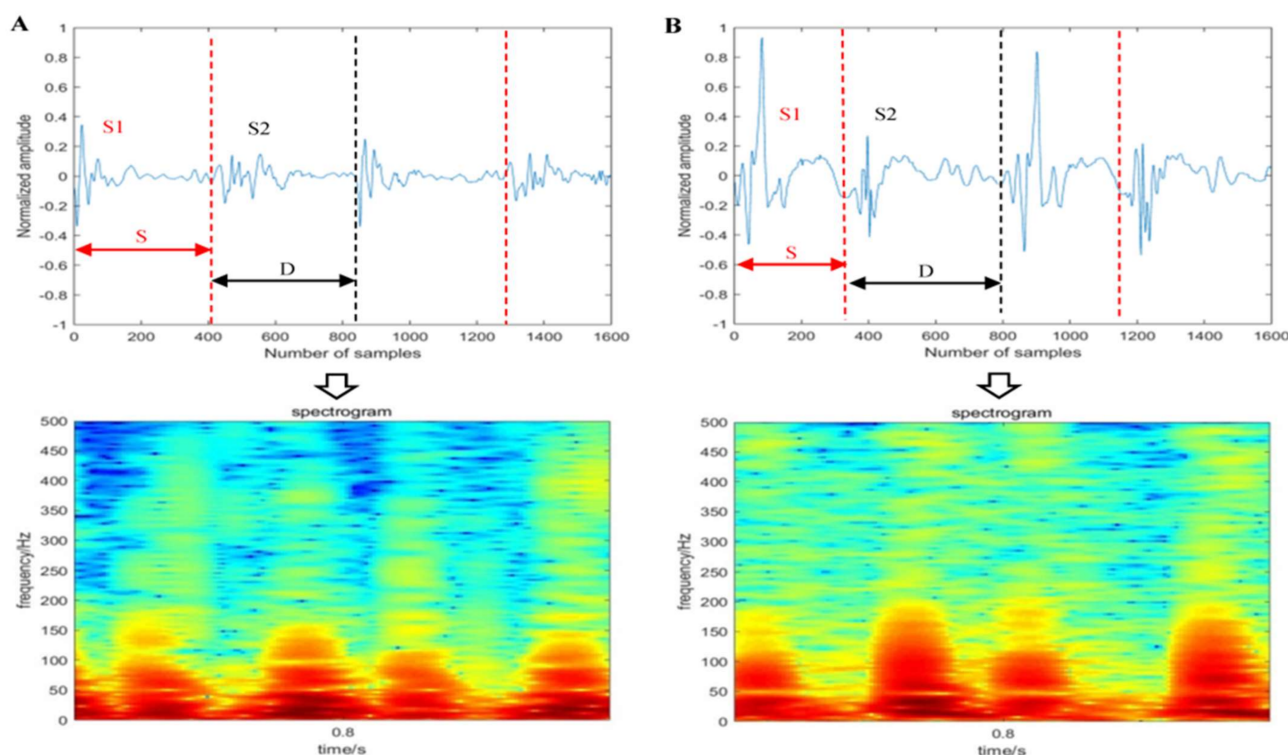


Figure 2. The examples of converting HS samples into spectrogram: (A) LVDD group; (B) control group.

2.3. Segmentation

The logistic regression-based hidden semi-Markov model (LR-HSMM) is the perfect method for HS segmentation or identifying the beginning of cycles [37,38]. As a result, we chose the LR-HSMM to determine the boundary of S1 and S2. In a cardiac period, the mechanical activity of the heart can be captured [39]. Furthermore, the characteristics of each cardiac period may differ. Therefore, the frame length was set to 1.6 s (approximately two cardiac periods), which started with S1 onset [40]. The intervals between two frames were reserved about two periods to avoid the overlap [40]. Table 2 describes the number of HS samples in this paper.

Table 2. The number of HS samples in different groups.

Group	Subjects	Age (Mean \pm Square Deviation)	Samples	Sample Length (s)
LVDD group	30	Aged 24–89 (66.87 \pm 16.21)	3677	1.6
Control group	41	Aged 19–81 (58.71 \pm 13.19)	4803	1.6

2.4. Short-Time Fourier Transform

By multiplying a window function before the Fourier transform (FT), the core idea of STFT is to solve the problem that the FT could only deal with the frequency domain [31]. The window function allows STFT to represent the time and frequency domain features of the HS signal, allowing the dynamic process to be fully revealed. It is supposed that the HS signal is approximately stationary within the span of a temporal window. The 1D HS sample $s(\tau)$ is converted into the 2D spectrogram using STFT as follows:

$$S(t, f) = \sum_{\tau=0}^{L-1} s(\tau)\omega(\tau - t)e^{-j2\pi\tau f}, \quad (2)$$

where t is time, f is frequency, L is the window length, and $\omega(\tau)$ is the window function. The log values of $S(t, f)$ are represented as the 2D spectrogram.

In this paper, the Hanning window was used, and the window length was set at 256. After that, the 2D spectrogram with size of 128×128 was obtained, where the shades of color represent the change of signal energy of different frequency bands, as shown in Figure 2. It can be seen that the highest frequency of S2 in the LVDD group is about 150 Hz, while in the control group, it is about 200 Hz. The different features in the spectrograms have a high degree of distinction in judging whether to suffer from LVDD.

2.5. Convolutional Neural Network

CNN is often applied in image classification tasks, which shows better performance than traditional methods in HS signals [41–43]. The convolutional layer, pooling layer, and fully-connected layer are still present in CNN's basic structure, despite the fact that it has evolved into a variety of different forms. Layer after layer, neural nodes are used to connect the layers.

The convolutional layer is to extract features. It comprises several different convolutional kernels, each of which is used to calculate different feature maps. In addition, the convolutional operation reduces the complexity of model and makes it easier to train. Equation (3) describes the j th feature map in layer l , where x_j^l is calculated using the previous feature map x_i^{l-1} multiplied by the convolutional kernel w_{ij}^l and adding a bias parameter b_j^l .

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} \times w_{ij}^l + b_j^l \right), \quad (3)$$

where $f(\cdot)$ is activation function.

The nonlinear property gives the model the ability of uniform approximation, and the activation function can bring it into model. The expression of the ReLU is simple, as shown in Equation (4). When $x \geq 0$, the derivative of the ReLU is 1, which can solve the disappearance of gradient and the over-fitting problem.

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4)$$

The pooling layer, also called down-sampling, is used to reduce dimension. It moves the pooling window to output an element in the feature map according to the rules, such as max-pooling or mean-pooling. The max-pooling was used in this paper as follows:

$$x_{j,k}^l = \max(x_{j:s+m,k:s+n}^{l-1}), \quad (5)$$

where j and k are the locations of the current feature map x_j^l , \max is the maximum function, s stands for pooling size, $m \geq 0$, and $n \leq s$.

The fully connected layer is essentially a perceptron connecting all the feature maps between the current layer and previous layer to generate the global semantic information. After this layer, the dropout is utilized to avoid over-fitting by disregarding some neurons at random throughout the training process [44].

In the classification task, the fully connected layer is often followed by the classification function, which is used to output the classification process of target. The commonly used classification function in CNN is softmax, which obtains the classification result by calculating the probability. The performance of softmax is improved using two fully connected layers by taking into account the influence of time and frequency in this paper [45]. The relevant information of the CNN model is shown in Table 3. The Adam optimizer with a learning rate of 0.0001 is used to improve the training speed of the model.

Table 3. The detailed information of the CNN model.

#	Layer	Filter Size	Stride	Output Dimension	Activation Function
1	Input	—	—	(128, 128, 3)	—
2	Conv1	3×3	1	(128, 128, 64)	ReLU
3	Maxpool1	2×2	2	(64, 64, 64)	—
4	Conv2	3×3	1	(64, 64, 32)	ReLU
5	Maxpool2	2×2	2	(32, 32, 32)	—
6	Conv3	3×3	1	(32, 32, 16)	ReLU
7	Maxpool3	2×2	2	(16, 16, 16)	—
8	FC1	—	—	(128, 1)	—
9	FC2	—	—	(100, 1)	dropout = 0.5
10	Output	—	—	(0, 1)	softmax

Conv = convolutional layer; Maxpool = max-pooling layer; FC = fully connected layer.

2.6. DA Methods

According to the way of generating samples, DA methods are parted into two classes: non-generative and generative methods.

2.6.1. Non-Generative DA Methods

Traditional non-generative DA methods are applied to spectrogram images, such as flipping, cropping, rotation, and other operations, which will result in a large amount of information loss or distortion from spectrograms, which have no physical meaning. For audio files, five audio augmentation techniques are adopted: positive pitch shift, negative pitch shift, slow time stretches, fast time stretches, and trim silence, and then, they are converted to spectrograms, which could effectively augment audio data [46].

2.6.2. Generative DA Methods

Commonly used generative DA methods are the GAN model and its derivatives. The GAN model comprises two neural networks, called generator (G) and discriminator (D) [21], as shown in Figure 3.

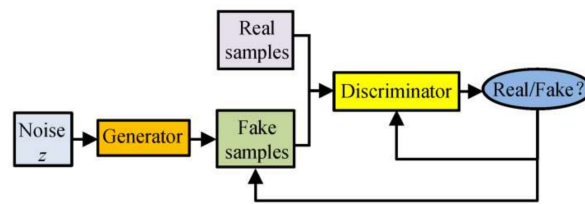


Figure 3. The structure of the GAN model.

The GAN model has the disadvantage of being too free, which results in uncontrollable and unpredictable defects when using the trained GAN model to generate samples [24]. To solve this problem, the DCGAN model has been proposed, which focus on the topology to ensure the stability in the training process [23]. In addition, since the DCGAN model can generate a large number of new samples that are closer to real samples while maintaining the validity of the semantic features [20,47], it is used to generate more samples to expand datasets in this paper. In the DCGAN model, D creates a filter that learns useful features of the target image based on CNN, while G ensures the feature quality and diversity of the generated image. The loss function of the DCGAN and GAN model is consistent, as given in Equation (6).

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log(1 - D(G(z)))], \quad (6)$$

where $P_{data}(x)$ represents the distribution of real samples x , and $P_z(z)$ represents the distribution of random noise z . G and D update parameters constantly through adversarial learning to restrict each other, and finally reach the Nash equilibrium.

In this paper, spectrograms of HS samples were used as real samples. Hence, there are 3677 images in the LVDD group and 4803 images in the control group. The DCGAN model was used to generate new samples through modeling the distribution of real samples. This is essentially a two-part minimum and maximum optimization problem:

- (i) Train G : fix D and optimize the parameters of G . $D(G(z))$ indicates the probability of D to judge whether the sample generated by G is real or not. Hence, the larger $D(G(z))$, the better, implying that $V(D, G)$ should be minimized, as shown in Equation (7). D sends its gradient back to G to update the parameters.

$$\min_G V(D, G) = E_{z \sim P_z(z)}[\log(1 - D(G(z)))] \quad (7)$$

- (ii) Train D : fix G and optimize the parameters of D . To improve the discriminative capacity of D , the smaller the fake samples, the better, which means minimizing $D(G(z))$. Therefore, $V(D, G)$ should be maximized, as shown in Equation (8). The back propagation of error updates the parameters of D .

$$\max_D V(D, G) = E_{x \sim P_{data}(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log(1 - D(G(z)))] \quad (8)$$

A convolutional layer was added to the original DCGAN model to construct a model with an output of 128×128 in this paper.

- (i) The generative model of the DCGAN model is a neural network similar to deconvolutional (Deconv), whose information is shown in Table 4. Its input layer is a random vector with uniform distribution. Secondly, the pooling layer is replaced by Deconv to learn spatial upsampling. Then, the stride is 2, the number of channels is cut in half, and the size is doubled. Batch normalization (BN) operation applies a transformation that maintains the means of the convolved features close to 0 and the variances of those close to 1 [48]. Adding BN operation to each layer except the output layer can solve initialization problems.

Table 4. The detailed information of the generative model.

Layer	Filter Size	Stride	Output Dimension	Activation Function	BN
Input	—	—	(1, 1, 100)	ReLU	Yes
Deconv1	5×5	2	(8, 8, 512)	ReLU	Yes
Deconv2	5×5	2	(16, 16, 256)	ReLU	Yes
Deconv3	5×5	2	(32, 32, 128)	ReLU	Yes
Deconv4	5×5	2	(64, 64, 64)	ReLU	Yes
Deconv5	5×5	2	(128, 128, 3)	ReLU	Yes
Output	—	—	(128, 128, 3)	Tanh	No

Deconv = deconvolutional layer.

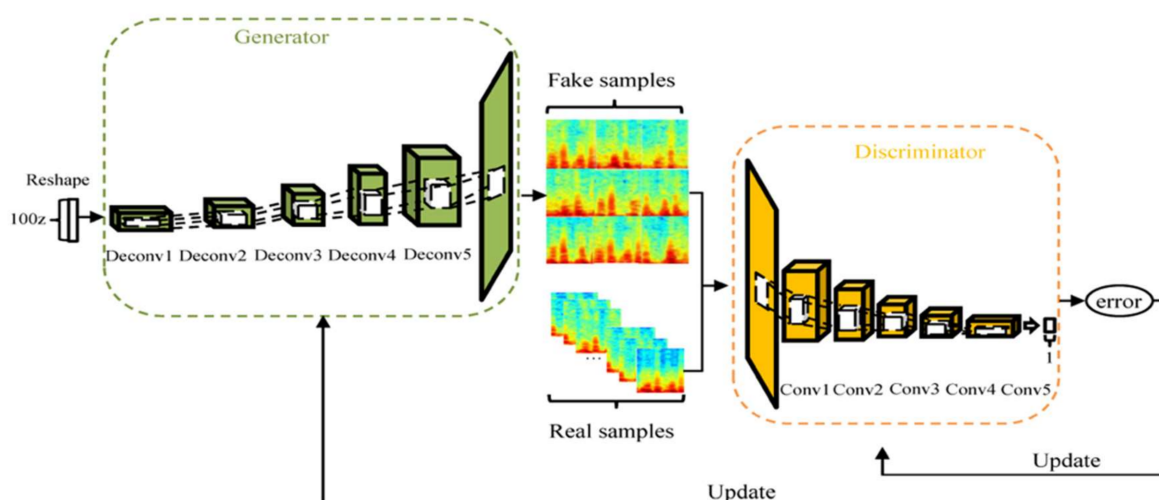
- (ii) The discriminant model of the DCGAN model is CNN, whose information is shown in Table 5. Similar to the generative model, the BN operation is carried out on each layer except the input layer, which alleviates the disappearance of gradient. Leaky ReLU, an activation function based on the ReLU, assigns a non-zero slope to all negative values for preventing the sparseness of gradient [49]. A scalar with a range of 0 to 1 is output through a fully connected layer to indicate the probability that input data belong to a real sample rather than a generated sample.

Table 5. The detailed information of the discriminant model.

Layer	Filter Size	Stride	Output Dimension	Activation Function	BN
Input	—	—	(128, 128, 3)	Leaky ReLU	No
Conv1	5×5	2	(64, 64, 64)	Leaky ReLU	Yes
Conv2	5×5	2	(32, 32, 128)	Leaky ReLU	Yes
Conv3	5×5	2	(16, 16, 256)	Leaky ReLU	Yes
Conv4	5×5	2	(8, 8, 512)	Leaky ReLU	Yes
Conv5	5×5	2	(4, 4, 1024)	Leaky ReLU	Yes
Output	—	—	(0, 1)	sigmoid	Yes

Conv = convolutional layer.

Figure 4 depicts the structure of the DCGAN model, using Adam as the optimizer with momentum of 0.5 to help to stabilize the training; the learning rate is 0.0002, batch size is 64, epoch is 300, and the slope of Leaky ReLU is 0.2.

**Figure 4.** The structure of the DCGAN model.

2.7. Similarity between Real and Generated Samples

For the generative result of the DCGAN model, how to evaluate whether the generated samples are qualified or not is a problem. Although the generated samples can be displayed as waveforms or spectrograms, it is difficult to detect the subtle changes visually. We

presented an indirect way to measure the similarity between the samples generated by the DCGAN model and the real samples to tackle this problem [50]. To begin, the DCGAN dataset, which consists of samples generated by the DCGAN model in the LVDD and control groups, was used as a training set to train the CNN model, which learns the features that generated samples. The RS dataset, which is composed of real samples, is then utilized to test the fully trained CNN model. Finally, the test accuracy of the CNN model reflects the similarity of two types of samples. This method could be used to compare the real samples with the generated samples in order to determine the differences between the two types of samples.

3. Results

Matlab (Version: R2018a) software was used to implement the signal preprocessing, segmentation, and STFT. The deep learning models were built with Python (Version: 3.7.0) and the Tensorflow package (Version: 2.0.0). All experiments were implemented on a computer with an Intel Core i5-9500K processor running at 3.70 GHz and a GTX1660Ti GPU with 64 GB of RAM.

3.1. Result of DA

As for the results of non-generative DA methods, this paper adopted the above five audio augmentation methods to augment HS, and then converted to spectrograms through STFT. Therefore, there are 18,385 images in the LVDD group and 24,015 images in the control group, called the non-generative augmentation dataset (NG dataset), which are mixed with the RS dataset as the input of the model for LVDD diagnosis.

In the DCGAN model, the cross-entropy loss function was utilized to define the loss of G and D, respectively. When the epoch was 300, the outputs of G and D were relatively stable, as well as the loss value was close to 1 and then started to oscillate, as shown in Figure 5. D can no longer tell the difference between real and fake samples at 300th epochs. Therefore, the epoch is set to 300. Figure 6 shows the examples of the generated spectrogram using the DCGAN model in the LVDD group. First, it is an original spectrogram and then is split in intervals of 50 epochs successively, where the profile and feature of the generated spectrogram becomes clearer and more obvious with an increasing number of epochs.

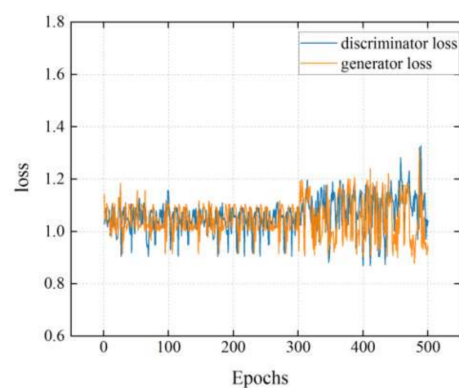


Figure 5. The loss value of the DCGAN model.

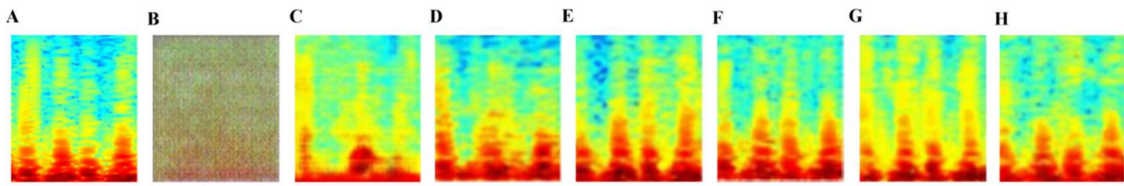


Figure 6. The examples of the spectrogram using the DCGAN model in the LVDD group with an increasing number of epochs: (A) original image; (B) epoch = 0; (C) epoch = 50; (D) epoch = 100; (E) epoch = 150; (F) epoch = 200; (G) epoch = 250; (H) epoch = 300.

3.2. Result of Similarity between Real and Generated Samples

The experiment was repeated by the method mentioned in Section 2.7, and the average test accuracy was 0.869. This indicates that the generated samples retain the most features of the real samples. Moreover, the DCGAN model could ensure the high validity and quality of the generated samples, regarding as a good method for DA. However, there are also some differences that indicate the diversity and unknown values of the generated samples. In view of this, it is also possible to mix the DCGAN dataset with the RS dataset and use them as the input of the CNN model (or other models) to identify LVDD. The structure of the CNN model is depicted in Figure 7.

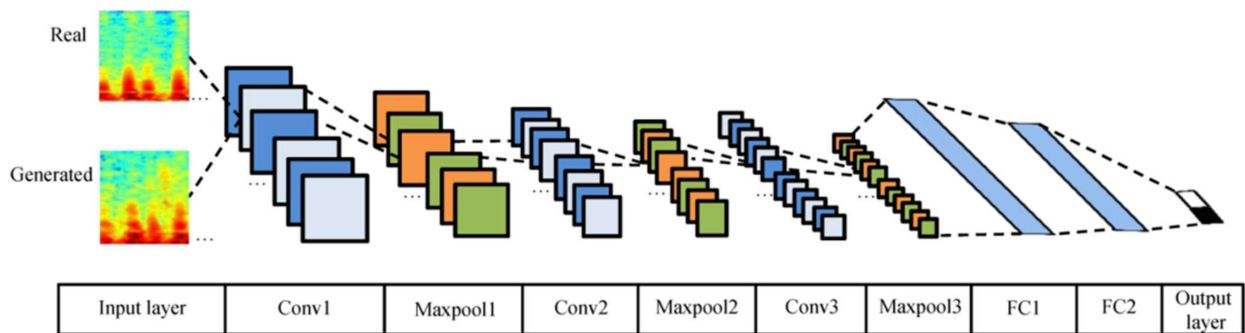


Figure 7. The structure of the CNN model.

3.3. Identification Performance

The performance of the proposed method was calculated from the testing set. All the samples were split into ten equal-sized subsets. Eight subsets were selected as the training set, and one subset was used as the validation set to tune the parameters of training process, and the rest was selected as the testing set, while specifying that the samples from the same subject cannot appear in the different subsets of the 10-fold cross-validation scheme. This process was carried out ten times. The model was evaluated in this paper using the typical evaluation metrics of accuracy (Acc), sensitivity (Se), and specificity (Sp).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}; \tag{9}$$

$$Se = \frac{TP}{TP + FN}; \tag{10}$$

$$Sp = \frac{TN}{TN + FP}; \tag{11}$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

3.3.1. Expansion Coefficient in the DCGAN Dataset

The expansion coefficient represents the average number of images generated by DA from each spectrogram of the RS dataset. According to Figure 8, when the expansion coefficient ranges 0 to 8 times, the performance is significantly improved, and the over-

fitting problem is gradually solved with the expansion of samples and the use of dropout operation. When the expansion coefficient in the DCGAN dataset is set to 8 times, the CNN model achieves good performance with an Acc of 0.981, a Se of 0.980, and a Sp of 0.982. As samples are further expanded, when the coefficient is in the range 8 to 20 times, the performance does not increase, reaching a state of saturation. Therefore, there are 29,416 images in the LVDD group and 38,424 images in the control group in the DCGAN dataset.

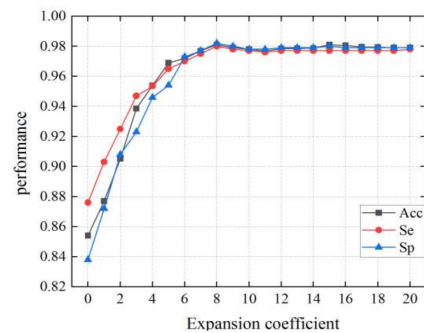


Figure 8. The performance of the CNN model in the DCGAN dataset with different coefficients.

3.3.2. Performance Comparison

In the three different datasets, we tested the performance of the CNN model. Moreover, the CNN model was compared against VGG-16, VGG-19, ResNet-18, ResNet-50, DenseNet-121, and AlexNet to determine the best model for the diagnosis of LVDD. The number of parameters of these models is shown in Table 6. The global average pooling layer was adjusted before the fully connected layer to avoid dimension mismatches, and the hyperparameters of each model remain consistent.

Table 6. Characteristics of the CNNs' architectures used in this paper.

Models	Parameters	Epochs
VGG-16	138,357,544	500
VGG-19	20,483,904	500
ResNet-18	63,470,656	500
ResNet-50	46,159,168	500
DenseNet-121	62,378,344	500
AlexNet	82,378,344	500
Proposed CNN	559,396	500

The CNN model was trained for 500 epochs, and the gradient of backpropagation was also calculated using the cross-entropy loss function. After expanding the samples, the dropout operation in the CNN model achieved a certain effect of regularization, accelerating the convergence speed.

According to Tables 7–9, our proposed CNN model achieves the best classification performance in the RS dataset + DCGAN dataset, with an average Acc of 0.987, Se of 0.986, and Sp of 0.988 among the seven models. The CNN model has the fewest parameters, so it takes less time to train. It can be seen that the length of training time has little impact on model performance. Furthermore, the DCGAN-based generative DA method outperforms the non-generative DA method in terms of classification performance. Compared with the RS dataset, the Acc, Se, and Sp of the CNN model in the RS dataset + DCGAN dataset are improved by 0.071, 0.054, and 0.093, respectively, which shows that the DCGAN model could effectively expand HS dataset of LVDD, thus improving the performance for LVDD diagnosis.

Table 7. The results of different models on the testing set in the RS dataset during the 10-fold cross-validation (mean \pm standard deviation).

Models	Evaluation Metrics			Training Time (mins:secs)
	Acc	Se	Sp	
VGG-16	0.913 \pm 0.018	0.929 \pm 0.018	0.892 \pm 0.019	443:68
VGG-19	0.894 \pm 0.022	0.908 \pm 0.021	0.876 \pm 0.022	112:37
ResNet-18	0.861 \pm 0.021	0.883 \pm 0.020	0.864 \pm 0.020	251:26
ResNet-50	0.883 \pm 0.022	0.899 \pm 0.021	0.871 \pm 0.020	178:13
DenseNet-121	0.842 \pm 0.019	0.856 \pm 0.022	0.825 \pm 0.021	224:79
AlexNet	0.879 \pm 0.017	0.897 \pm 0.019	0.857 \pm 0.019	317:25
Proposed CNN	0.916 \pm 0.015	0.932 \pm 0.017	0.895 \pm 0.018	65:44

RS dataset = 3677 images in the LVDD group and 4803 images in the control group.

Table 8. The results of different models on the testing set in the RS dataset + NG dataset during the 10-fold cross-validation (mean \pm standard deviation).

Models	Evaluation Metrics			Training Time (mins:secs)
	Acc	Se	Sp	
VGG-16	0.949 \pm 0.009	0.958 \pm 0.009	0.931 \pm 0.010	535:25
VGG-19	0.928 \pm 0.013	0.933 \pm 0.012	\pm 0.912 \pm 0.013	243:14
ResNet-18	0.902 \pm 0.011	0.911 \pm 0.010	0.901 \pm 0.011	361:12
ResNet-50	0.925 \pm 0.012	0.934 \pm 0.011	0.914 \pm 0.011	302:57
DenseNet-121	0.887 \pm 0.011	0.896 \pm 0.012	0.867 \pm 0.012	332:13
AlexNet	0.919 \pm 0.009	0.930 \pm 0.009	0.898 \pm 0.010	435:79
Proposed CNN	0.955 \pm 0.005	0.966 \pm 0.007	0.947 \pm 0.008	179:26

RS dataset + NG dataset = 22,062 images in the LVDD group and 28,818 images in the control group.

Table 9. The results of different models on the testing set in the RS dataset + DCGAN dataset during the 10-fold cross-validation (mean \pm standard deviation).

Models	Evaluation Metrics			Training Time (mins:secs)
	Acc	Se	Sp	
VGG-16	0.981 \pm 0.003	0.978 \pm 0.004	0.979 \pm 0.003	494:56
VGG-19	0.964 \pm 0.004	0.956 \pm 0.005	0.961 \pm 0.005	194:78
ResNet-18	0.957 \pm 0.004	0.941 \pm 0.006	0.949 \pm 0.005	314:34
ResNet-50	0.968 \pm 0.005	0.952 \pm 0.006	0.951 \pm 0.006	263:71
DenseNet-121	0.936 \pm 0.005	0.922 \pm 0.007	0.913 \pm 0.006	286:45
AlexNet	0.962 \pm 0.004	0.959 \pm 0.005	0.955 \pm 0.005	381:24
Proposed CNN	0.987 \pm 0.001	0.986 \pm 0.002	0.988 \pm 0.002	130:13

RS dataset + DCGAN dataset = 33,093 images in the LVDD group and 43,227 images in the control group.

Figure 9 demonstrates how the performance of the CNN model in the RS dataset + DCGAN dataset improves as the number of epochs increases during training. Since the 350th epoch, the CNN model has steadily converged, with the accuracy and loss of validation set approaching those of the training set, indicating that the training epoch of 500 is sufficient.

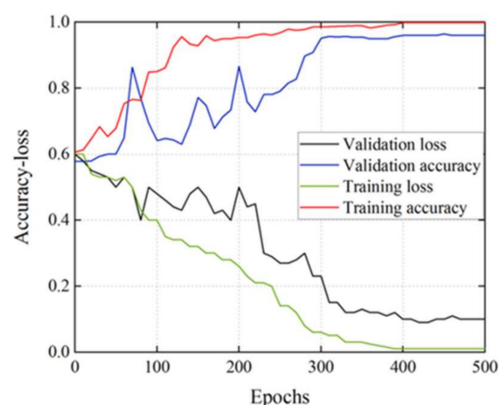


Figure 9. The training and validation performance of the CNN model in the RS dataset + DCGAN dataset at 500 epochs.

4. Discussion

Over the years, many studies on the diagnosis of LVDD involved biomarkers, echocardiographic diagnostic indices, epidemiology, and pathogenesis. For example, Mocan et al. [51] found that IL-6 as an inflammatory biomarker has the best predictive capacity for LVDD, but such an invasive method is not suitable for pervasive application. HS signal classification, a non-invasive method, has potential to identify cardiovascular diseases and has been extensively studied, but the objects of these researches are different [52,53]. However, the study for LVDD diagnosis based on deep learning using HS has not been studied. This paper could therefore be an efficient supplement for the diagnosis of LVDD.

Due to the available HS databases are insufficient to train deep learning models, Narváez and Percybrooks [22] developed a GAN-based DA method that could only generate normal HS with a single type and participate in classification tasks. In the GAN model, G and D are in healthy competition for concurrent learning. The loss function is derived using the output of D, and therefore it updates parameters more quickly, whereas G does not always converge. Another drawback of the GAN model is that, for diverse random inputs, G produces similar outputs [54]. In the original DCGAN model, CNN architecture is integrated into unsupervised learning training to alleviate this problem and boost generative effect [55]. In this paper, the DCGAN model was applied to the analysis of HS, which could solve the problem of sparsity and imbalance of samples using the DA of HS dataset. A DCGAN model with the output of 128×128 was constructed to automatically generate spectrograms of HS frames, which could be used as the input of the different deep learning models for training. Compared with the non-generative DA method, the DCGAN-based DA method improves the classification performance of different models. Moreover, it is also beneficial to enhance the speed and quality of the generated samples, which shows the feasibility and effectiveness of the DCGAN model for DA in the HS field.

The benefits of CNN can be shown in the following areas: it could learn features from datasets automatically, and it has convolutional invariant, and it could obtain features straight from the convolutional layer, overcoming the limitations of hand-crafted feature extraction and providing high reliability [56]. Furthermore, transforming the HS signal into a 2D spectrogram could better describe the time and frequency aspects of the signal [57]. Based on the above advantages, we proposed a 2D CNN model for the diagnosis of LVDD. The performance of this model is compared with six common deep learning models, as shown in Tables 7–9. In the RS dataset + DCGAN dataset, the CNN model has the best classification performance and stability, which is a wonderful solution to the problem of small samples. The method classifies spectrograms generated using the DCGAN model and decreases the influence of out-of-distribution inputs, indicating it the preferred method for the diagnosis of LVDD using HS.

This paper also has two limitations. Firstly, the samples generated by the DCGAN model have a low resolution of 128×128 , which may affect the classification results of deep learning models [58]. Secondly, the number of collected HS signals of LVDD is small, which makes the features in the spectrograms insufficient to continue DA effectively to improve model performance. We could collect more HS signals of LVDD to train the super resolution generative adversarial networks (SRGAN) model to generate high-resolution spectrograms, so as to have a better performance of deep learning models.

The DA method proposed in this paper could be utilized in the future to establish HS databases for the other cardiovascular diseases in order to expand the application of deep learning in the different fields. It provides important auxiliary diagnosis for medical professionals and forms a non-invasive, low-cost, and reusable early diagnosis method to avoid tedious examination.

5. Conclusions

Early diagnosis of LVDD can help guide treatment and provide important insights into the progression and severity of HFpEF. We proposed a CNN model for the diagnosis of LVDD using HS in this paper that does not rely on hand-crafted feature extraction, as well as a DCGAN-based DA method for expanding the HS dataset of LVDD. The experimental results support the efficacy of our proposed model. Furthermore, the suggested method is not constrained by small datasets and offers a viable solution for HS signal analysis using deep learning, which is a potential non-invasive technology for the identification of LVDD.

Author Contributions: Conceptualization, Y.Y., X.-M.G. and Y.-N.Z.; methodology, Y.Y. and Y.-N.Z.; validation, Y.Y., X.-M.G., Y.-N.Z. and H.W.; formal analysis, Y.Y. and Y.-N.Z.; investigation, Y.Y., X.-M.G., H.W. and Y.-N.Z.; resources, X.-M.G. and Y.-N.Z.; data curation, Y.Y., X.-M.G., H.W. and Y.-N.Z.; writing—original draft preparation, Y.Y. and Y.-N.Z.; supervision, X.-M.G. and Y.-N.Z.; project administration, X.-M.G. and Y.-N.Z.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number: 31800823, 31870980 and 31570003).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the First Affiliated Hospital of Chongqing Medical University (Protocol no: CYYYYLL2018-092, 15 January 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wan, S.-H.; Vogel, M.W.; Chen, H.H. Pre-clinical diastolic dysfunction. *J. Am. Coll. Cardiol.* **2014**, *63*, 407–416. [[CrossRef](#)]
2. Redfield, M.M.; Jacobsen, S.J.; Burnett, J.C.; Mahoney, D.W.; Bailey, K.R.; Rodeheffer, R.J. Burden of systolic and diastolic ventricular dysfunction in the community—appreciating the scope of the heart failure epidemic. *Jama* **2003**, *289*, 194–202. [[CrossRef](#)] [[PubMed](#)]
3. Abhayaratna, W.P.; Marwick, T.H.; Smith, W.T.; Becker, N.G. Characteristics of left ventricular diastolic dysfunction in the community: An echocardiographic survey. *Heart* **2006**, *92*, 1259–1264. [[CrossRef](#)] [[PubMed](#)]
4. Lam, C.S.P.; Lyass, A.; Kraigher-Krainer, E.; Massaro, J.M.; Lee, D.S.; Ho, J.E.; Levy, D.; Redfield, M.M.; Pieske, B.M.; Benjamin, E.J.; et al. Cardiac dysfunction and noncardiac dysfunction as precursors of heart failure with reduced and preserved ejection fraction in the community. *Circulation* **2011**, *124*, 24–30. [[CrossRef](#)] [[PubMed](#)]
5. Mureddu, G.F.; Agabiti, N.; Rizzello, V.; Forastiere, F.; Latini, R.; Cesaroni, G.; Masson, S.; Cacciatore, G.; Colivicchi, F.; Ugucconi, M.; et al. Prevalence of preclinical and clinical heart failure in the elderly. A population-based study in Central Italy. *Eur. J. Heart Fail.* **2012**, *14*, 718–729. [[CrossRef](#)]
6. Prasad, S.B.; Guppy-Coles, K.B.; Holland, D.; Stanton, T.; Krishnasamy, R.; Whalley, G.; Atherton, J.J.; Thomas, L. Echocardiographic predictors of all-cause mortality in patients with left ventricular ejection fraction >35%: Value of guideline-based assessment of diastolic dysfunction. *Int. J. Cardiol. Hearth Vasc.* **2019**, *24*, 1–6. [[CrossRef](#)]

7. Ashcheulova, T.V.; Demydenko, H.V.; Herasymchuk, N.M.; Rezunencko, Y.K.; Kochubei, O.A.; Ambrosova, T.M. Diastolic dysfunction: From the discovery to the latest updates. *Zaporozhye Med. J.* **2019**, *21*, 538–545.
8. Jeong, E.M.; Dudley, S.C. Diastolic dysfunction: Potential new diagnostics and therapies. *Circ. J.* **2015**, *79*, 470–477. [[CrossRef](#)]
9. Kossaify, A.; Nasr, M. Diastolic dysfunction and the new recommendations for echocardiographic assessment of left ventricular diastolic function: Summary of guidelines and novelties in diagnosis and grading. *J. Diagn. Med. Sonogr.* **2019**, *35*, 317–325. [[CrossRef](#)]
10. Das, S.; Pal, S.; Mitra, M. Acoustic feature based unsupervised approach of heart sound event detection. *Comput. Biol. Med.* **2020**, *126*, 103990. [[CrossRef](#)] [[PubMed](#)]
11. Wu, C.; Herman, B.A.; Retta, S.M.; Grossman, L.W.; Liu, J.S.; Hwang, N.H.C. On the closing sounds of a mechanical heart valve. *Ann. Biomed. Eng.* **2005**, *33*, 743–750. [[CrossRef](#)]
12. Eslamzadeh, G.; Barati, R. Heart murmur detection based on wavelet transformation and a synergy between artificial neural network and modified neighbor annealing methods. *Artif. Intell. Med.* **2017**, *78*, 23–40. [[CrossRef](#)] [[PubMed](#)]
13. Chauhan, S.; Wang, P.; Sing Lim, C.; Anantharaman, V. A computer-aided MFCC-based HMM system for automatic auscultation. *Comput. Biol. Med.* **2008**, *38*, 221–233. [[CrossRef](#)]
14. Son, G.Y.; Kwon, S. Classification of heart sound signal using multiple features. *Appl. Sci.* **2018**, *8*, 2344. [[CrossRef](#)]
15. Zheng, Y.; Guo, X.; Ding, X. A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification. *Expert Syst. Appl.* **2015**, *42*, 2710–2721. [[CrossRef](#)]
16. Saraf, K.; Baek, C.I.; Wasko, M.H.; Zhang, X.; Zheng, Y.; Borgstrom, P.H.; Mahajan, A.; Kaiser, W.J. Assessment of left ventricular diastolic function using phonocardiogram signals: A comparison with echocardiography. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Montreal, QC, Canada, 20–24 July 2020; pp. 4067–4070.
17. He, Y.; Li, W.; Zhang, W.; Zhang, S.; Pi, X.; Liu, H. Research on segmentation and classification of heart sound signals based on deep learning. *Appl. Sci.* **2021**, *11*, 651. [[CrossRef](#)]
18. 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 44th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1318–1322.
19. Dong, F.; Qian, K.; Ren, Z.; Baird, A.; Li, X.; Dai, Z.; Dong, B.; Metze, F.; Yamamoto, Y.; Schuller, B.W. Machine listening for heart status monitoring: Introducing and benchmarking HSS—The Heart Sounds Shenzhen Corpus. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2082–2092. [[CrossRef](#)]
20. Salamon, J.; Bello, J.P. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Process. Lett.* **2017**, *24*, 279–283. [[CrossRef](#)]
21. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Aaron Courville, Y.B. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
22. Narváez, P.; Percybrooks, W.S. Synthesis of normal heart sounds using generative adversarial networks and empirical wavelet transform. *Appl. Sci.* **2020**, *10*, 7003. [[CrossRef](#)]
23. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2016**, arXiv:1511.06434.
24. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training GANs. *arXiv* **2016**, arXiv:1606.03498v1.
25. Schlegl, T.; Seebock, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. *arXiv* **2017**, arXiv:1703.05921v1.
26. Choi, E.; Biswal, S.; Malin, B.; Duke, J.; Stewart, W.F.; Sun, J. Generating multi-label discrete patient records using generative adversarial networks. *arXiv* **2017**, arXiv:1703.06490.
27. Donahue, C.; McAuley, J.; Puckette, M. Synthesizing audio with GANs. In Proceedings of the 6th International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018; pp. 1–6.
28. Yun Luo, B.-L.L. EEG Data augmentation for emotion recognition using a conditional wasserstein GAN. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* **2018**, *2018*, 2535–2538.
29. Wu, Y.; Yang, F.; Liu, Y.; Zha, X.; Yuan, S. A comparison of 1-D and 2-D deep convolutional neural networks in ECG classification. *arXiv* **2018**, arXiv:1810.07088.
30. Zhao, J.; Mao, X.; Chen, L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomed. Signal Process. Control* **2019**, *47*, 312–323.
31. Griffin, D.W.; Lim, J.S. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust.* **1984**, *32*, 236–243. [[CrossRef](#)]
32. Palmiero, P.; Zito, A.; Maiello, M.; Cameli, M.; Modesti, P.A.; Muiesan, M.L.; Novo, S.; Saba, P.S.; Scicchitano, P.; Pedrinelli, R. Left ventricular diastolic function in hypertension: Methodological considerations and clinical implications. *J. Clin. Med. Res.* **2015**, *7*, 137–144. [[CrossRef](#)]
33. Jeong, Y.; Kim, J.; Kim, D.; Kim, J.; Lee, K. Methods for improving deep learning-based cardiac auscultation accuracy: Data augmentation and data generalization. *Appl. Sci.* **2021**, *11*, 4544. [[CrossRef](#)]

34. Nagueh, S.F.; Smiseth, O.A.; Appleton, C.P.; Byrd, B.F.; Dokainish, H.; Edvardsen, T.; Flachskampf, F.A.; Gillebert, T.C.; Klein, A.L.; Lancellotti, P. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: An update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J. Am. Soc. Echocardiogr.* **2016**, *29*, 277–314. [[CrossRef](#)]
35. Debbal, S.M.; Bereksi-Reguig, F. Time-frequency analysis of the first and the second heartbeat sounds. *Appl. Math. Comput.* **2007**, *184*, 1041–1052. [[CrossRef](#)]
36. Liu, Y.; Guo, X.; Zheng, Y. An automatic approach using ELM classifier for HFpEF identification based on heart sound characteristics. *J. Med. Syst.* **2019**, *43*, 285. [[CrossRef](#)]
37. 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. [[CrossRef](#)]
38. Springer, D.B.; Tarassenko, L.; Clifford, G.D. Logistic regression-HSMM-based heart sound segmentation. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 822–832. [[CrossRef](#)] [[PubMed](#)]
39. Deng, S.W.; Han, J.Q. Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. *Futur. Gener. Comput. Syst.* **2016**, *60*, 13–21. [[CrossRef](#)]
40. Gao, S.; Zheng, Y.; Guo, X. Gated recurrent unit-based heart sound analysis for heart failure screening. *Biomed. Eng. Online* **2020**, *19*, 1–17. [[CrossRef](#)] [[PubMed](#)]
41. Demir, F.; Şengür, A.; Bajaj, V.; Polat, K. Towards the classification of heart sounds based on convolutional deep neural network. *Health Inf. Sci. Syst.* **2019**, *7*, 1–9. [[CrossRef](#)]
42. Alqudah, A.M.; Alquran, H.; Qasmieh, I.A. Classification of heart sound short records using bispectrum analysis approach images and deep learning. *Netw. Model. Anal. Health Inform. Bioinform.* **2020**, *9*, 66. [[CrossRef](#)]
43. Dhar, P.; Dutta, S.; Mukherjee, V. Cross-wavelet assisted convolution neural network (AlexNet) approach for phonocardiogram signals classification. *Biomed. Signal Process. Control* **2021**, *63*, 102142. [[CrossRef](#)]
44. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv* **2012**, arXiv:1207.0580.
45. Tabar, Y.R.; Halici, U. A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* **2017**, *14*, 016003. [[CrossRef](#)] [[PubMed](#)]
46. Mushtaq, Z.; Su, S.F.; Tran, Q.V. Spectral images based environmental sound classification using CNN with meaningful data augmentation. *Appl. Acoust.* **2021**, *172*, 107581. [[CrossRef](#)]
47. Cui, X.; Goel, V.; Kingsbury, B. Data augmentation for deep convolutional neural network acoustic modeling. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing, South Brisbane, QLD, Australia, 19–24 April 2015; pp. 4545–4549.
48. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on machine learning, Lille, France, 7–9 July 2015; pp. 448–456.
49. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; p. 28.
50. Shao, S.; Wang, P.; Yan, R. Generative adversarial networks for data augmentation in machine fault diagnosis. *Comput. Ind.* **2019**, *106*, 85–93. [[CrossRef](#)]
51. Mocan, M.; Mocan Hognogi, L.D.; Anton, F.P.; Chiorescu, R.M.; Goidescu, C.M.; Stoia, M.A.; Farcas, A.D. Biomarkers of inflammation in left ventricular diastolic dysfunction. *Dis. Markers* **2019**, *2019*, 1–14. [[CrossRef](#)]
52. Beritelli, F.; Capizzi, G.; Lo Sciuto, G.; Napoli, C.; Scaglione, F. Automatic heart activity diagnosis based on gram polynomials and probabilistic neural networks. *Biomed. Eng. Lett.* **2018**, *8*, 77–85. [[CrossRef](#)]
53. Ren, Z.; Cummins, N.; Pandit, V.; Han, J.; Qian, K.; Schuller, B. Learning image-based representations for heart sound classification. In Proceedings of the 2018 International Conference on Digital Health, Lyon, France, 23–26 April 2018; pp. 143–147.
54. Rath, A.; Mishra, D.; Panda, G.; Satapathy, S.C. Heart disease detection using deep learning methods from imbalanced ECG samples. *Biomed. Signal Process. Control* **2021**, *68*, 102820. [[CrossRef](#)]
55. Öcal, A.; Özbakır, L. Supervised deep convolutional generative adversarial networks. *Neurocomputing* **2021**, *449*, 389–398. [[CrossRef](#)]
56. Chen, W.; Sun, Q.; Chen, X.; Xie, G.; Wu, H.; Xu, C. Deep learning methods for heart sounds classification: A systematic review. *Entropy* **2021**, *23*, 667. [[CrossRef](#)]
57. Dwivedi, A.K.; Imtiaz, S.A.; Rodriguez-Villegas, E. Algorithms for automatic analysis and classification of heart sounds—A systematic review. *IEEE Access* **2019**, *7*, 8316–8345. [[CrossRef](#)]
58. Yip, M.Y.T.; Lim, G.; Lim, Z.W.; Nguyen, Q.D.; Chong, C.C.Y.; Yu, M.; Bellemo, V.; Xie, Y.; Lee, X.Q.; Hamzah, H.; et al. Technical and imaging factors influencing performance of deep learning systems for diabetic retinopathy. *NPJ Digit. Med.* **2020**, *3*, 31–34. [[CrossRef](#)] [[PubMed](#)]