



Performance Improvement in Deep Learning Architecture for Phonocardiogram Signal Classification Using Spectrogram

R. Sai Kesav, M. Bhanu Prakash^(✉), Krishanth Kumar, V. Sowmya, and K. P. Soman

Centre for Computational Engineering and Networks (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India
v_sowmya@cb.amrita.edu, kp_soman@amrita.edu

Abstract. Phonocardiogram (PCG) assumes a critical part in the early determination of heart irregularities. Phono-cardiogram can be utilized as an underlying diagnostics apparatus in far-off applications because of its effortlessness and cost-adequacy. The proposed work targets utilising a CNN architecture, with multiple preprocessing strategies like converting to Spectrogram or normalizing the signals, which analyze various cardiovascular anomalies from PCG signals gathered from different sources. Our study shows the viability of utilising Spectrogram and Normalization of signals in cardiac abnormalities identification. This work avoids feature extraction and trivial pre-processing mechanisms, and we have achieved promising results.

Keywords: Phonocardiogram (PCG) · Spectrogram · Deep learning

1 Introduction

The Phonocardiogram (PCG) is the graphical portrayal of the sound delivered from the cardiac muscles' cardiovascular activity inside the heart. During the heart's cardiovascular movement, murmur sounds are generated, which are unidentified by an Electrocardiogram [1] but can be recognized by a Phonocardiogram and gives an assistive analysis in the early location of cardiovascular illnesses [1]. The PCG signal comprises s1 and s2, the two significant heart sounds. s1, the low pitch sound (lub) is generated because of atrioventricular valve closure and when the blood moves from the heart chamber to the ventricle. s2, the high pitch sound (dub) is generated when blood flows through the vessels from the heart to the lung, for a limited duration. A systolic period is defined from the early s1 to s2, and a diastolic period is defined from the early s2 to s1. The systolic and diastolic periods together constitute a single heart cycle. In addition to the fundamental sound segments, there are multiple and unusual sounds like a murmur, extrasystole delivered in case of any cardiovascular abnormality [11].

Phonocardiogram is generally recorded in a clinical climate utilising an innovatively advanced stethoscope. The Phonocardiogram is additionally recorded inside a non-clinical arrangement. The recorded Phonocardiogram signal consists of different noises, making the classification of the systolic and diastolic beats generated by the heart's mechanical activity challenging to be classified.

Spectrogram has been used in this work, an electronic or visual representation of the spectrum of frequencies for a signal, which varies with time generated through a Fourier transform where frequency and time are visually represented. Different colours are used to show the spectrum's magnitude. Further, we have trained the standard CNN architecture `citeb4` using the Spectrogram. Our objective is to show the use of simple preprocessing techniques like spectrogram, and normalization on the signals can give better classification results.

2 Related Works

In the literature, a study about the automatic classification of sounds generated by the heart's mechanical activity is described [1,2]. The localization algorithm is used to localize the peaks from the input signal and constructing windows from those peaks, which are further used to classify the extracted features. Further study shows the analysis of the PCG signal's time and frequency analysis, which overcomes the frequency and time domains' anomalies. The time-frequency methods include wavelets and Empirical Mode Decomposition. In Sujadevi et al. [9] Variational Mode Decomposition (VMD) based denoising was used to remove noise from the heart sounds and using the same technique to denoise the signal and visually displaying the waveforms. In Sujit et al. [8] the abnormality produced in the heart sounds were detected by a back-end classifier that extracts the time and frequency from the PCG signal using the Ada-Boost technique and SMOTE. In Schimdt et al. [5], a diagnosis system is proposed which utilizes Support Vector Machines (SVM) to classify the diseases developed in heart valves. In Sujadevi et al. [7], different algorithms RNN, B-RNN, LSTM, B-LSTM, CNN, and GRU were used for PCG classification. 80% accuracy was achieved for CNN.

In Sujadevi et al. [4], without any denoising and trivial pre-processing techniques, a convolutional neural network (CNN) on the physionet datasets using a raw signal achieved better results. Similarly, in [3], CNN was declared by them as the better network to classify Phonocardiogram signals. Both the Physionet datasets and the AISTATS 2012 datasets were collected from multiple sources.

3 Methodology

All the existing works utilize the input signal based on the time domain used for the PCG Signal classification using deep learning architectures. The current work is based on a spectrogram that converts the input signal from the time domain to the frequency domain through a visual portrayal of the spectrum of frequencies of a signal as it varies with time. A spectrogram is generated by

Fourier transform using the time and frequency as x-axis and y-axis respectively and different colours to show the magnitude of the spectrum. The spectrogram images are fed as input to the CNN architecture, specified in [4], where training and classification are done accordingly for the different datasets used. The CNN architecture used for this work is shown in Fig. 7.

Before converting the datasets into spectrograms, the signals should be fixed to a particular length (here 5 s) and then normalized. Normalization was done because we found that datasets 1 and 3 are noisy signals compared to dataset 2. Which are shown in Fig. 1, Fig. 2, and Fig. 3

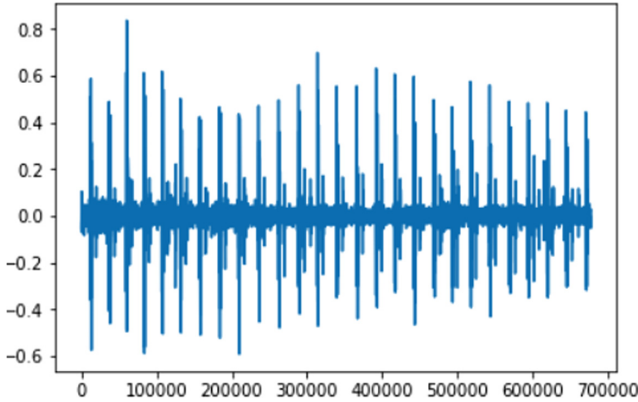


Fig. 1. Normal Signal sample was taken from dataset 1

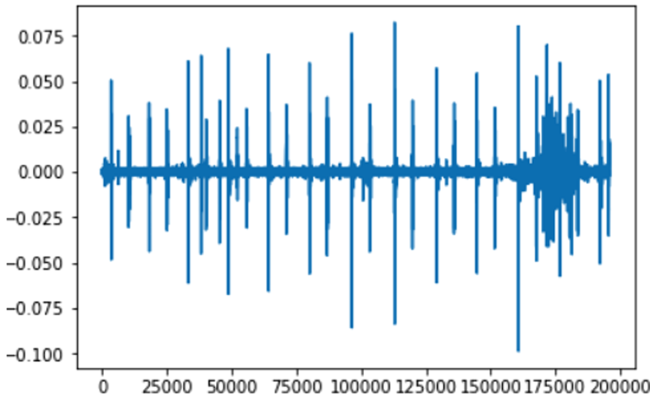


Fig. 2. Normal signal sample was taken from dataset 2

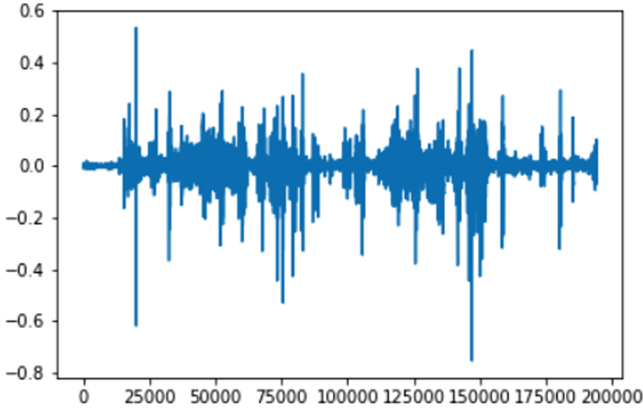


Fig. 3. Normal signal sample was taken from dataset 3

As observed from Fig. 1, Fig. 2, Fig. 3, there is very little noise in dataset 2. So the normalization is done for datasets 1 and 3 using the min-max normalization (for which the pseudo-code is shown below).

Algorithm 1. Psuedo Code for Normalization

Input: Signal;

Output: Normalized Signal;

```
def audio-norm(signal):
```

```
  max_data = max(signal)
```

```
  min_data = min(signal)
```

```
  norm_signal = (data - min_data) / (max_data - min_data + 0.0001)
```

```
  return norm_signal - 0.5
```

All the signals are trimmed to 5 s, but if a signal is less than 5 s, then zero paddings are added, which is shown in Fig. 4.

From Fig. 4, it has been observed that there is zero-padding added to the signal. When we compare the original signal's corresponding amplitude values with the normalized signal, the amplitude has been normalized, giving better classification results. Amplitude for comparison is shown below:

Original amplitude values: [-0.00806781 -0.00920942 -0.00991806 ... -0.03190301 -0.02225797]

Normalized amplitude values: [0.08240861 0.08149707 0.08093125 ... 0.08885056 0.08885056 0.08885056]

When compared to the previously achieved output in [3] and the metrics which was achieved here for dataset-3, as seen in the results in Table 6, an improvement in output is observed when the proposed method of fixing to a particular length and normalization is used.

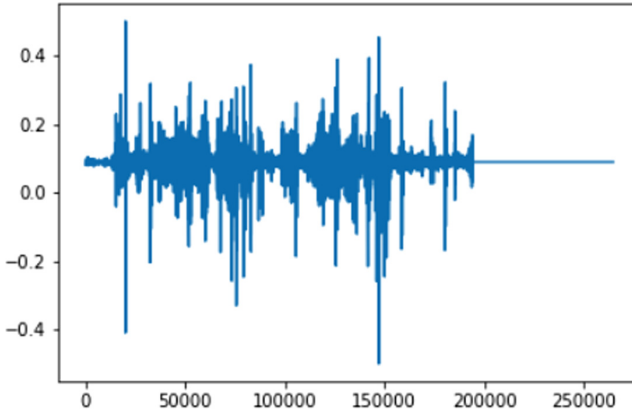


Fig. 4. Dataset 3 normalized normal signal

3.1 Input Description

In our current work, the Phonocardiogram signal(PCG) is gathered from different sources, and a total of three datasets are used for this work. Dataset-1, collected from clinical and non-clinical conditions, is accessible from physionet challenge 2016 and consisting of classes: normal and abnormal. Dataset-2 and Dataset-3 are accessible from AISTATS 2012 test. Dataset-2 was gathered utilising the iStethoscope Pro iPhone application and consisted of four classes: extrasystole, murmur, artifact, and normal. Dataset-3 was gathered using a digital stethoscope – DigiScope application and consisted of three classes: normal, murmur, extrasystole.

3.2 Spectrogram

A spectrogram is generated through a Fourier transform which visually represents the spectrum of frequencies of a given signal as it varies with time. Frequency and time are horizontal and verticals in a formed visual representation and different colours show the spectrum's magnitude [7].

For a given signal of x with a length of N , there are consecutive segments of the signal of m , where $m \leq N$, and the $x \in$

$$R^{m \times (N-m+1)}$$

where, in the formed matrix, rows and columns of x are indexed by time.

$\hat{x} = F * x$ and $x = (1/m) * F * \hat{x}$, of size m and matrix F , which are DFT columns of x , and F is the Fourier matrix with F_i , being its complex conjugate.

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{i2\pi} & e^{i\frac{4\pi}{N}} & \dots & e^{i2\pi\frac{(N-1)}{N}} \\ 1 & e^{i\frac{4\pi}{N}} & e^{i\frac{8\pi}{N}} & \dots & e^{i2\pi\frac{2(N-1)}{N}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & e^{i2\pi\frac{(N-1)}{N}} & e^{i2\pi\frac{2(N-1)}{N}} & \dots & e^{i2\pi\frac{(N-1)^2}{N}} \end{bmatrix} \tag{1}$$

Rows and columns of \hat{x} are indexed by the frequency and time, respectively, and their location corresponds to the point in frequency and time. The spectrogram is a visualised matrix where the matrix image with the i th and j th entry in the matrix, corresponds to the intensity or colour of the i th, and j th pixel in the visually represented image general the bright colours denote the strong frequencies in a spectrogram. Figure 5 and Fig. 6 portray the spectrographic images of datasets 1 and 2.

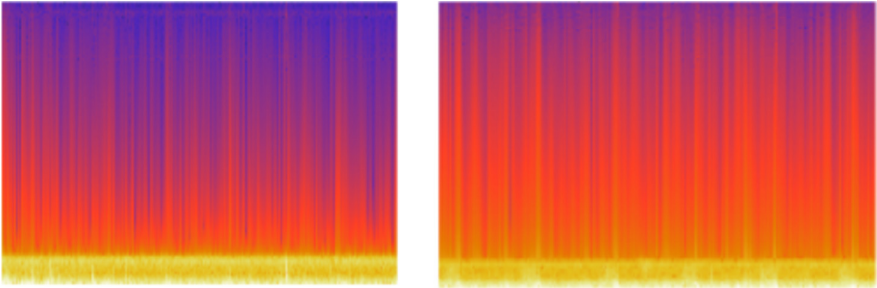


Fig. 5. Spectrographic images of dataset-1 showing normal and abnormal classes

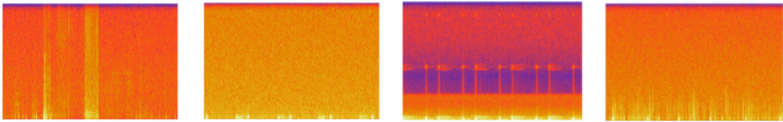


Fig. 6. Spectrographic images of dataset-2 showing normal, murmur, extrahls and artifact classes

3.3 Deep Learning Architecture

In [4], a distinguished classification for the PCG model was implemented. In this work, similar topologies and hyperparameters for crude PCG Signal classification are gathered from different clinical and non-clinical conditions. The experiments

are done using the benchmark architecture of CNN and using the three datasets (Dataset 1,2, and 3). The architecture of CNN used can be seen in Fig. 7

Convolutional Neural Network (CNN): The CNN architecture used here comprises of four Convolution layers stacked along with 64 filters, each of filter size 3, and followed by an average pooling layer and a ReLu activation function. The average pooling layer reduces the size of the feature without losing any information and a flattening layer after the 4th convolution layer followed by five dense layers with the softmax activation function. The architecture details are mentioned in Table 1.

The loss function and optimizer used in this architecture are Logcosh and ADAM optimizer. Whose formula can be seen as:

$$L(y, y^P) = \sum_{i=1}^n \log(\cosh(y_i^P - y_i)) \quad (2)$$

where y_i^P is the predicted values, and y_i are the original values.

As we have achieved less accuracy for dataset 3 using spectrogram, we implemented 1D convolution layers instead of using spectrogram images and have used the normalized signal for their classification.

The loss function and optimizer used for dataset 3 are Categorical Cross-Entropy and ADAM optimizer. Whose formula can be seen as

$$CCE(p, t) = - \sum_{c=1}^C t_{0,c} \log(p_{0,c}) \quad (3)$$

C is no. of classes

t is the binary indicator for the correctly classified observation o.

p is o's predicted probability in class c.

4 Experimental Results

4.1 Dataset Description

Phonocardiogram (PCG) Data Gathered from both Clinical and Non-Clinical Environments (Dataset 1): This information source is a segment from physionet challenge 2016, which is gathered from healthy and unhealthy patients worldwide in clinical and non-clinical conditions. Detailed information regarding datasets is clarified underneath in Table 2, which sums up the total number of samples considered from each class. It is purposefully used as training and testing input for the model from 665 abnormal and 2575 normal signals.

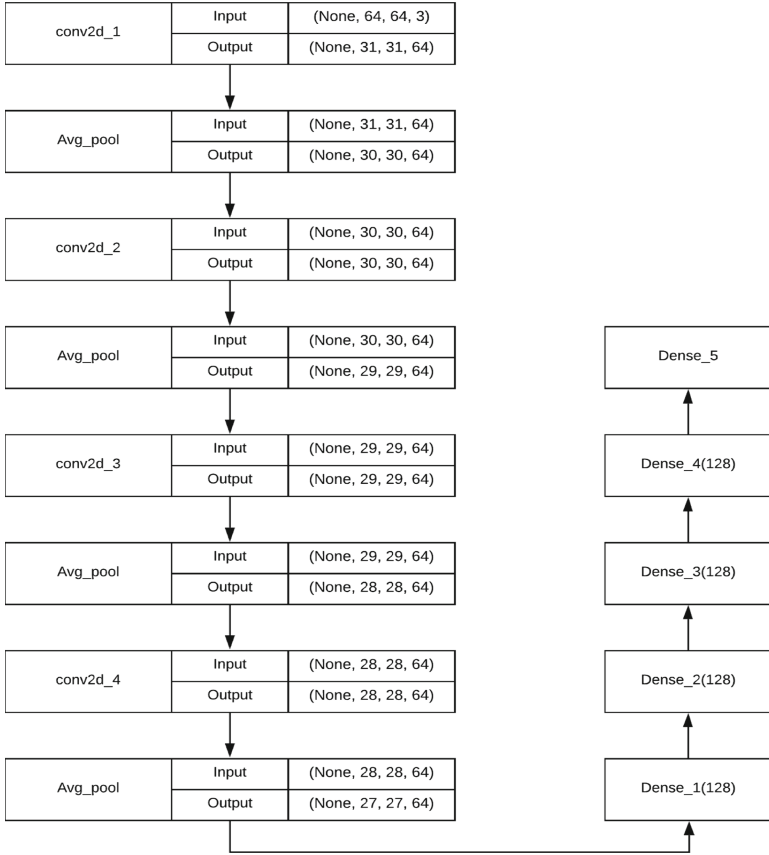


Fig. 7. Architecture of CNN used for classification

Table 1. The Architectural details of the PCG Signal Classification Collected from multiple sources with distinct Cardiac Abnormalities

Architecture details		CNN
No. of neurons in the stacked layers		64
Zeropaddings (Border mode)		Yes
Filter size		3
Pool length		2
Number of neurons in dense layers		128
Output layer	Dataset 1	2
	Dataset 2	4
	Dataset 3	3

Table 2. The Summary of PCG Datasets used for the PCG Classification

Dataset	Class name	Class number	Number of Signals		
			Train	Test	Total
1	Normal	0	2060	515	2575
	Abnormal	1	532	133	665
	Total		2592	828	3240
2	Artifact	3	32	8	40
	Extrasystole	2	15	4	19
	Murmur	1	27	7	34
	Normal	0	25	6	31
	Total		99	25	124
3	Extrasystole	2	36	10	46
	Murmur	1	76	19	95
	Normal	0	255	64	319
	Total		367	93	460

PCG Information Gathered Utilising Ipro Phone Application and Digiscope (Dataset 2 and 3): The information source is from an event, AISTATS 2012, supported by PASCAL, and two distinct sorts of datasets are accessible. Dataset 2 is gathered utilising the Ipro phone application utilising iStethoscope. Dataset 3 is gathered from a clinical environment using a computerized stethoscope. Table 2 portrays the outline of the dataset accessible in the AISTATS 2012 challenge. A dataset split of training and testing signals based on the strategy proposed by AISTATS 2012.

Table 3. Hyper-parameter set for the CNN

Hyper-parameters	CNN
Batch-size	32
Learning rate	0.1
No. of hidden layers	8
Optimizer	Adam
No. of epochs	50

Table 4. The summary of the results obtained using the above Architecture and Spectrogram images

Metrics/Datasets	Precision	Accuracy	Recall	f1-score
Dataset I	0.66	0.82	0.82	0.73
Dataset II	0.85	0.85	0.85	0.84
Dataset III	0.58	0.67	0.67	0.62

Table 5. The Summary of the results obtained for Dataset III using 1D Convolution without Spectrogram

Metrics/Datasets	Precision	Accuracy	Recall	f1-score
Dataset III	0.81	0.83	0.83	0.80

4.2 Result Analysis

In Sujadevi et al. [6], for dataset 1, when trained with raw signals promising results were obtained using CNN. They have concluded that the architecture, as seen in Fig. 7, gave better results. For benchmarking the optimum value, multiple experiments were done with various configurations. In this work, the same hyper-parameters used in [6] are considered. The learning rate and batch size are fixed at 0.1 and 32 for the deep learning architecture. All the parameters used are shown in Table 3.

The model's performance was evaluated using precision, recall, F1-score, and accuracy. The results obtained when spectrogram is used shown in Table 4. And the results of dataset 3 without using spectrogram are shown in Table 5.

The comparison of our results and previous results obtained in [3] are displayed in Table 6. Here, an improvement is observed in the classification performance of dataset 2 using spectrogram when compared with the results of the existing methodology used.

Utilising dataset 1, work done has been almost a replica of the previous existing classification performance accuracy of 82%. For dataset 2, the proposed work improved resulted in achieving better classification performance than the previously existing one, from accuracy 82% to 85%. In the case of dataset 3, we have achieved better results than the previously mentioned performance. We have normalized the signal and used 1D convolutions in place of Fast Fourier Transform.

Table 6. Performance comparison of the PCG Classification using CNN for Existing work and Proposed work

Data	Metrics	Existing	Proposed
1	Precision	0.83	0.66
	Accuracy	0.82	0.82
	Recall	0.82	0.82
	F1-score	0.83	0.73
2	Precision	0.85	0.85
	Accuracy	0.80	0.85
	Recall	0.79	0.85
	F1-score	0.82	0.84
3	Precision	0.81	0.81
	Accuracy	0.75	0.83
	Recall	0.76	0.83
	F1-score	0.79	0.80

5 Conclusion

In this work, it was found that data preprocessing is an essential task for training. As seen in dataset 2, it has given a very good output. And class imbalance problems can lead to overfitting of data of the majority class, as seen in dataset 1. Therefore downsampling was used to overcome the imbalance. As it was noticed that the challenge in which dataset 3 was given is for feature extraction, we used fixed signal length and normalized the signals, and performed 1D convolutions with them, which gave better classification results than that of FFT results. The current work can be extended to make further advancements in cardiac diseases classification.

References

1. Amit, G., Gavriely, N., Intrator, N.: Cluster analysis and classification of heart sounds. *Biomed. Signal Process. Control* **4**(1), 26–36 (2009)
2. Maglogiannis, I., Loukis, E., Zafiropoulos, E., Stasis, A.: Support vectors machine-based identification of heart valve diseases using heart sounds. *Comput. Methods Prog. Biomed.* **95**(1), 47–61 (2009)
3. Gopika, P., Sowmya, V., Gopalakrishnan, E.A., Soman, K.P.: Performance improvement of deep learning architectures for phonocardiogram signal classification using fast fourier transform. In: *International Conference on Intelligent Computing and Communication Technologies, (ICICCT-2019)*. Springer, Heidelberg (2019)

4. Sujadevi, V.G., Soman, K.P., Vinayakumar, R., Prem Sankar, A.U.: Anomaly detection in phonocardiogram employing deep learning. In: Behera, H.S., Nayak, J., Naik, B., Abraham, A. (eds.) Computational Intelligence in Data Mining. AISC, vol. 711, pp. 525–534. Springer, Singapore (2019). https://doi.org/10.1007/978-981-10-8055-5_47
5. Schmidt, S.E., Holst-Hansen, C., Graff, C., Toft, E., Struijk, J.J.: Segmentation of heart sound recordings by a duration-dependent hidden markov model. *Physiol. Meas.* **31**(4), 513 (2010)
6. Sujadevi, V., Soman, K.P., Vinayakumar, R., Sankar, A.P.: Deep models for phonocardiography (PCG) classification. In: The 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT), pp. 211–216. IEEE (2017)
7. Sujadevi, V.G., Soman, K.P., Vinayakumar, R., Sankar, A.U.P.: Anomaly detection in phonocardiogram employing deep learning. In: Advances in Intelligent Systems and Computing, vol. 711, pp. 525–534 (2019)
8. Sujit, N.R., Kumar C.S., Rajesh, C.B.: Improving the performance of cardiac abnormality detection from PCG signal. In: AIP Conference Proceedings, vol. 1715 (2016)
9. Sujadevi, V.G., Soman, K.P., Kumar, S.S., Mohan, N., Arunjith, A.S.: Denoising of phonocardiogram signals using variational mode decomposition. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, pp. 1443–1446 (2017). <https://doi.org/10.1109/ICACCI.2017.8126043>
10. <https://www.princeton.edu/~cuff/ele201/files/spectrogram.pdf>
11. <https://www.ncbi.nlm.nih.gov/books/NBK333/>