Performance Improvement of Deep Learning Architectures for Phonocardiogram Signal Classification using Spectrogram

Krishanth Kumar
CB.EN.P2CEN19007
Center for excellence in Computational Engineering & Networking
Amrita Vishwa Vidyapeetham
Coimbatore, India
krishanthvs@gmail.com

Sai Kesav.R
CB.EN.P2CEN19010
Center for excellence in Computational Engineering & Networking
Amrita Vishwa Vidyapeetham
Coimbatore, India
rachakondasaikesav@gmail.com

Bhanu Prakash.M
CB.EN.P2CEN19008
Center for excellence in Computational Engineering & Networking
Amrita Vishwa Vidyapeetham
Coimbatore, India
bhanu0893@gmail.com

***Abstract*—Phonocardiogram known as PCG plays a significant role in the early diagnosis of cardiac abnormalities. Phonocardiogram can be used as an initial diagnostics tool in remote applications due to its simplicity and cost effectiveness. Instead of a disease specific approach, the proposed work aims for a single architecture that could diagnose different cardiac abnormality from the PCG signals collected from various sources. Our study also shows the effectiveness of using Spectrogram in signal processing applications. It avoids the trivial pre- processing and feature extraction mechanisms with the promising results.**

***IndexTerms*- Phonocardiogram (PCG), Spectrogram, Deep Learning.**

INTRODUCTION

The Phonocardiogram (PCG) is the graphical representation of the heart sound produced due to the mechanical activity of the heart valves and muscles. Phonocardiogram detects the heart murmurs caused by damaged valves, which cannot be detected in Electrocardiogram [1]. Phonocardiogram gives assistive diagnosis in early detection of cardiac diseases [1].The PCG signal consists of two major heart sounds s1 and s2. s1 is a low pitch sound (lub), which is longer in duration produced due to the closure of atrioventricular valves and when the blood flows from atrium to ventricle. s2 is a high pitch sound (dub) produced, when the blood flows from heart to the lung, which is shorter in duration. The beginning of early s1 to s2 is known as the systolic period. The beginning of the next s2 to s1 is known as the diastolic period. The systolic and diastolic period together known as one cardiac cycle. Additional to the main sound components, there are some abnormal sounds like murmur, extrasystole produced, if there is any cardiac abnormality.

Phonocardiogram is usually recorded in a clinical environment using a digital stethoscope. Due to the advancement in technology, the Phonocardiogram is also recorded in non clinical setup. The recorded Phonocardiogram signal contains various noises. So, the localization of the beat and classification of the Phonocardiogram signal is the challenging task.

We have chosen 3 datasets collected from various datasets, dataset-1 is collected from both clinical and non-clinical environments with two classes of normal and abnormal taken from 2016 Physionet Challenge. Dataset-2 is collected using iStethoscope Pro iPhone app with four classes normal, murmur, extrasystole, artifact. Dataset-3 is collected using the digital stethoscope DigiScope with three classes normal, murmur, extrasystole, the both Dataset- 2 and Dataset-3 are taken from AISTATS (International Conference on Artificial Intelligence and Statistics) 2012 challenge.

In this work we have used Spectrogram, which is a visual representation of the spectrum of frequencies of a signal as it varies with and is generated through a fourier transform where frequency and time are visually represented and different colors are used to show the magnitude of the spectrum.

We have structured the CNN and Skip-connection CNN for the chosen datasets with their corresponding spectrographic representation of the signal and have resulted in their respective outputs, where we found out that CNN architecture was giving better results when compared to Skip-CNN.

OBJECTIVE

Our objective is to compare the results for both CNN and the Skip-connection CNN architectures for the chosen datasets, and analyze which one is better.

RELATED WORKS

In the literature, there are many studies involved for the automatic classification of the heart sound. Further research includes the time-frequency analysis of the PCG signal, which overcomes the disadvantages of time and frequency domains. The time frequency methods include Empirical Mode Decomposition (EMD) and wavelets. In this [5], they proposed a diagnosis system using Support Vector Machines (SVM) for the classification of heart valve disease .

In [4], without any trivial pre-processing and denoising techniques, convolutional neural network (CNN) has shown encouraging results in the case of physionet databases using raw signal. Similarly, in [3] , CNN was declared by them as the better network to classify Phonocardiogram signals. Both the Physionet database and the AISTATS 2012 database are collected from different sources. The proposed work shows that there is no performance improvement in case of Skip Residual CNN for Phonocardiogram classification.

METHODOLOGY

All the existing works based on deep learning architectures for PCG signal classification use the input signal in the time domain. The proposed work converts the input PCG signal from the time domain to the frequency domain using spectrogram which is a visual representation of the spectrum of frequencies of a signal as it varies with time. A spectrogram is generated by fourier transform and uses time and frequency as x-axis and y-axis respectively and different colors to show the magnitude of the spectrum.. The frequency domain signal is fed as input to an existing CNN deep learning architecture and its results are compared to the results of an Skip connection CNN architecture, and determined which enhances the performance of PCG signal classification. The workflow of the proposed work is depicted in Figure 1.

Before the datasets can be converted into spectrograms first they have to be fixed to a particular length(here 5 seconds) and then to be normalized as we found out that dataset 1 and 3 are noisy signals as compared to that of dataset 2. Which we can see in the below figure.

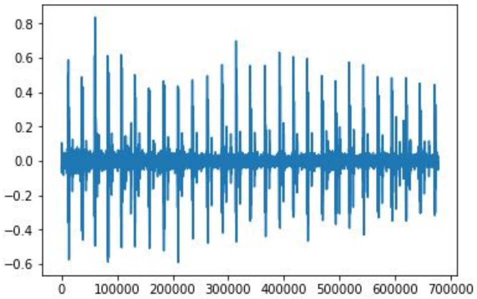


Fig.1: Dataset 1 normal signal

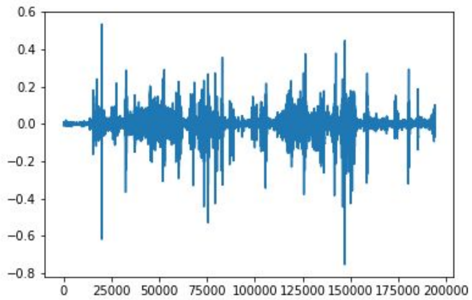


Fig.2:Dataset 3 normal signal

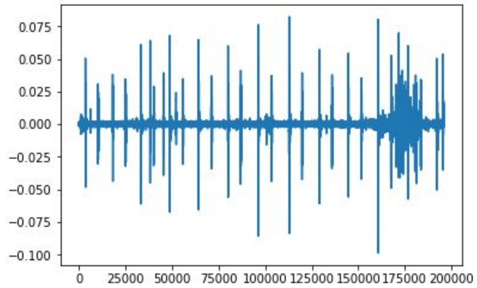


Fig.3:Dataset 2 normal signal

As we can see that there is very less noise in dataset 2. So the normalization done is dataset 1 and 3 using the min-max normalization(for which the code used is mentioned below).

```
def audio_norm(data):
    max_data = np.max(data)
    min_data = np.min(data)
    data = (data-min_data)/(max_data-min_data+0.0001)
    return data-0.5
```

Now lets compare the normalised signal for dataset 3. All the signals were trimmed to 5 seconds , but the given signal is for 3 seconds so zero padding is added.

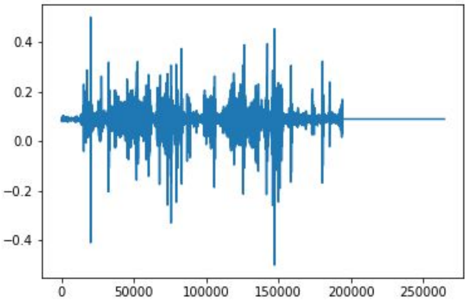


Fig.4:Dataset 3 normalized normal signal

Although it seems that there is only zero padding added to the signal when we check the corresponding amplitude values for the same the amplitude has been normalised which leads to better output.Please check the below array values of amplitude for comparison.

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 and the metrics which we achieved in using dataset-3, gave a better output when each signal fixed to a particular length and normalized and we achieved a better output without using FFT (Fast Fourier Transform) and doing 1D convolutions.

A. Input Description

In our proposed work, input to the network is Phonocardiogram signal. Phonocardiogram signal(PCG) is collected from various sources. Three datasets are considered for the study.

Dataset- 1 is available at physionet challenge 2016, which is collected from both clinical and non-clinical environments . It contains two classes namely normal and abnormal. Dataset- 2 and Dataset-3 are available at AISTATS 2012 challenge. Dataset-2 is collected using iStethoscope Pro iPhone app.It has four classes namely normal, murmur, extrasystole and artifact. Dataset-3 is collected using the digital stethoscope DigiScope. It has three classes namely normal, murmur, extrasystole.

B. Spectrogram

In this process of solving the objective we have used spectrogram which is the visual representation of the spectrum of frequencies of a signal as it varies with time and it is generated through a fourier transform, where frequency and time are horizontal and verticals in a formed visual representation and different colours are used to show the magnitude of the spectrum.

For a given signal of x with a length of N, there are consecutive segments of signal of m, where m<<N, and the

$x \in \mathbb{R}^{m*(N-m+1)}$ where the formed matrix, rows and columns of x are indexed by time.

$\dot{x} = F*x$ and $x = (1/m)*F*\dot{x}$, of size m and the matrix F, whose columns are the DFT columns of x.where F, is the complex conjugate of F_i and F is the fourier matrix

$$F = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{i\frac{2\pi}{N}} & e^{i\frac{4\pi}{N}} & \cdots & e^{i2\pi\frac{N-1}{N}} \\ 1 & e^{i\frac{4\pi}{N}} & e^{i\frac{8\pi}{N}} & \cdots & e^{i2\pi\frac{2(N-1)}{N}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{i2\pi\frac{N-1}{N}} & e^{i2\pi\frac{2(N-1)}{N}} & \cdots & e^{i2\pi\frac{(N-1)^2}{N}} \end{bmatrix}.$$

Fig.5: n*n fourier matrix of F

Rows of ẋ are indexed by the frequency and the columns are indexed by time, where each location on it corresponds to the point in frequency and time, as the spectrogram is an visualised matrix where the matrix image with the i,j-th entry in the matrix corresponding to the intensity or color of the i,j-th pixel in the image and the in general the bright colors represent the strong frequencies in a spectrogram.

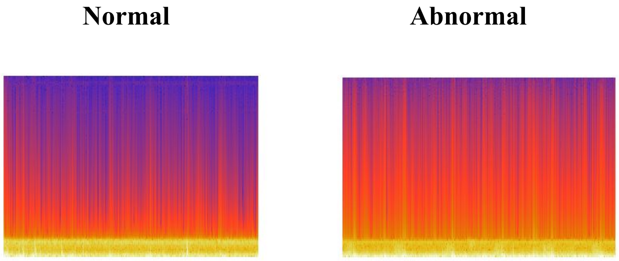


Fig.6:Spectrographic Images of Dataset-1

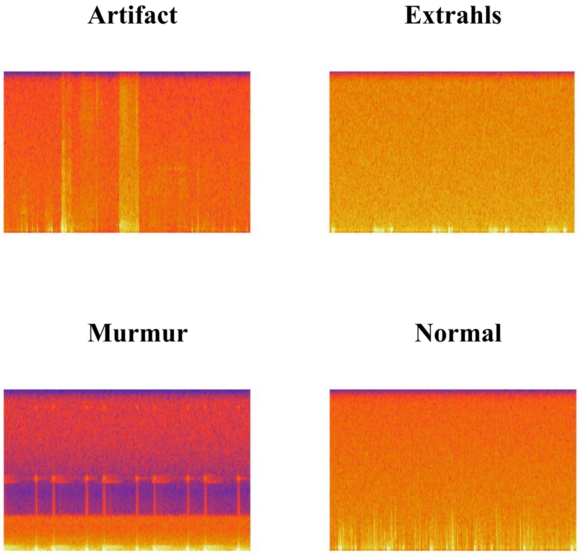


Fig.7:Spectrographic Images of Dataset-2

C. Deep Learning architectures

In our proposed method, we considered the best architectures used for raw PCG Signal classification collected from various sources like clinical and non-clinical environments. The experiment is repeated for all the considered benchmark architectures CNN and Skip-connection CNN using the three databases (Dataset 1,2 and 3). The performance comparison is done on the benchmark deep learning architectures like Convolutional Neural Network (CNN), Skip-connection CNN were used for raw PCG signal classification .

1. *Convolutional Neural Network:* Convolutional Neural Network (CNN) has four stacked Convolution layers with 64 filters. Each filter is of size 3. Every Convolution layer is followed by an average pooling layer and a ReLu activation function. Average pooling layer reduces the size of the feature map without any loss of information, and a flattening layer after the 4th convolution layer followed by 5 dense layers with the softmax activation function.

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

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

Where y_i^p is the predicted values, and y_i are the original values.

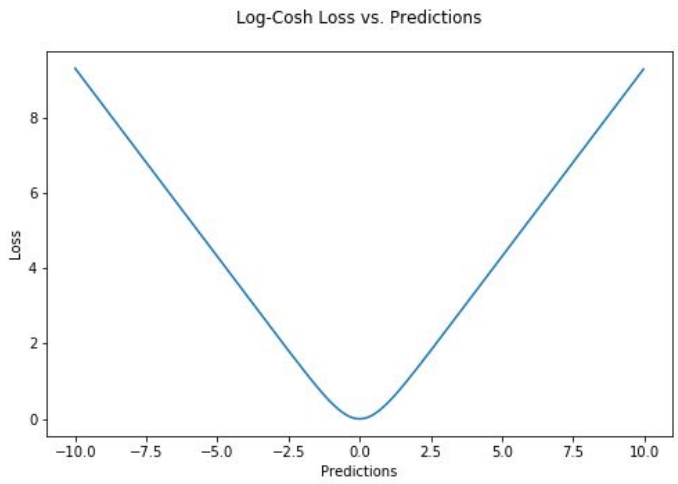


Fig.8: Log-Cosh Loss Function's plot

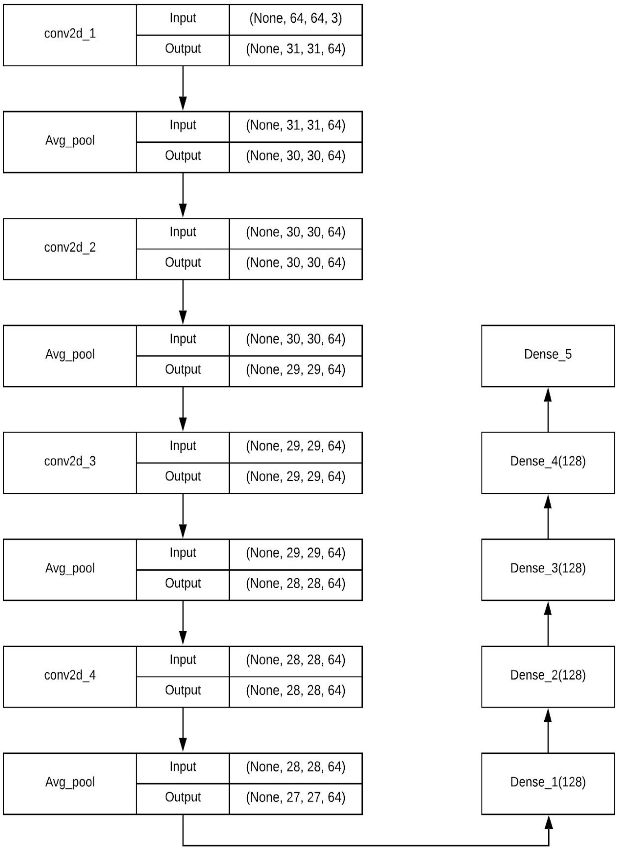
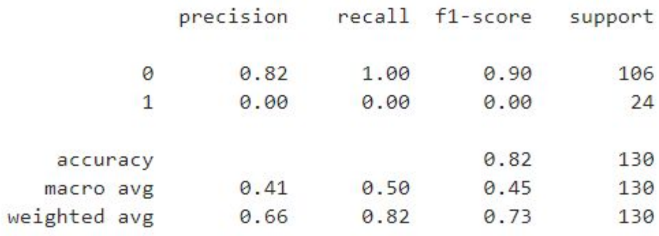


Fig.9: Architecture of CNN

Below are the metrics achieved by using the above architecture.

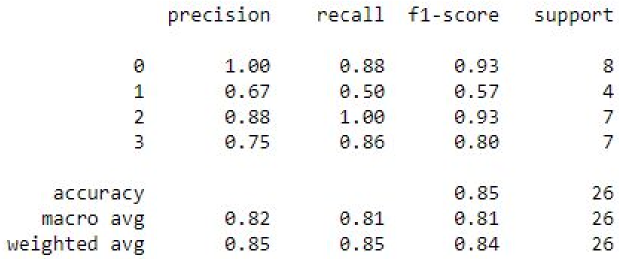
For dataset 1 using spectrogram


```
              precision    recall  f1-score   support

           0       0.82      1.00      0.90       106
           1       0.00      0.00      0.00        24

    accuracy                           0.82       130
   macro avg       0.41      0.50      0.45       130
weighted avg       0.66      0.82      0.73       130
```

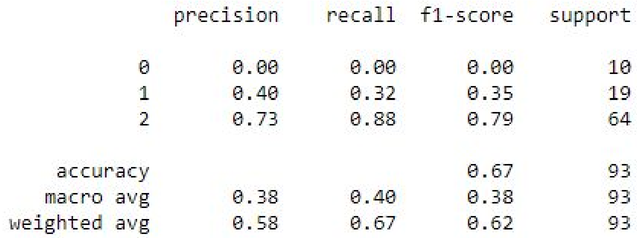
For dataset 2 using spectrogram


```
              precision    recall  f1-score   support

           0       1.00      0.88      0.93         8
           1       0.67      0.50      0.57         4
           2       0.88      1.00      0.93         7
           3       0.75      0.86      0.80         7

    accuracy                           0.85        26
   macro avg       0.82      0.81      0.81        26
weighted avg       0.85      0.85      0.84        26
```

Dataset 3 using spectrogram


```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        10
           1       0.40      0.32      0.35        19
           2       0.73      0.88      0.79        64

    accuracy                           0.67        93
   macro avg       0.38      0.40      0.38        93
weighted avg       0.58      0.67      0.62        93
```

As we got less accuracy for dataset 3 using spectrogram, we then tried normalizing the signals and used 1D convolutions without fft .

The loss function and optimizer used for dataset 3 are Categorical_Cross_Entropy and ADAM optimizer. Whose formula can be seen as:

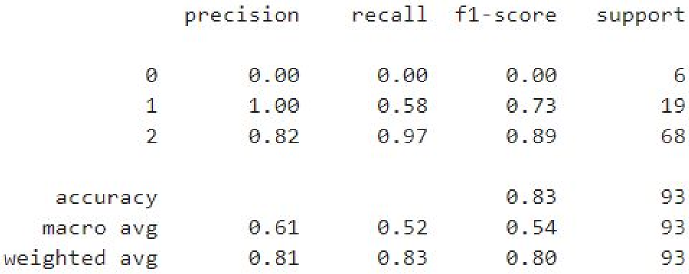
$$CCE(p,t) = -\sum_{c=1}^{C} t_{o,c}\log\left(p_{o,c}\right)$$

C is no.of classes

t is the binary indicator if the class has been classified correctly for observation o.

p is the predicted probability observation o in class c.

Dataset 3 without spectrogram and fft


```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         6
           1       1.00      0.58      0.73        19
           2       0.82      0.97      0.89        68

    accuracy                           0.83        93
   macro avg       0.61      0.52      0.54        93
weighted avg       0.81      0.83      0.80        93
```

2. *Skip-connection CNN:* The structured network is stacked with four convolutional layers with 64 filters having a filter size of 3, and similar to that of the convolutional layer architecture we have an average pooling layer, and a ReLu activation function, and a flattening layer after the 4th convolution layer followed by the 5 dense layers with the activation function. The final architecture of skip-connection CNN is, there are two skip connections in this shallow architecture from 1st convolutional layer to the third convolutional layer and from 2nd

convolutional layer to the 4th convolutional layer. These skip connections are the extra connections between nodes in different layers of the CNN that are skipped for one layer and in certain cases for more than one layer for non-linear processing, They also accelerate the traverse information of the feature throughout the network faster than the CNN architectures.

Before implementing this architecture we have tried different skip connections initially and got poor results. Those of which can be seen in the below figures.

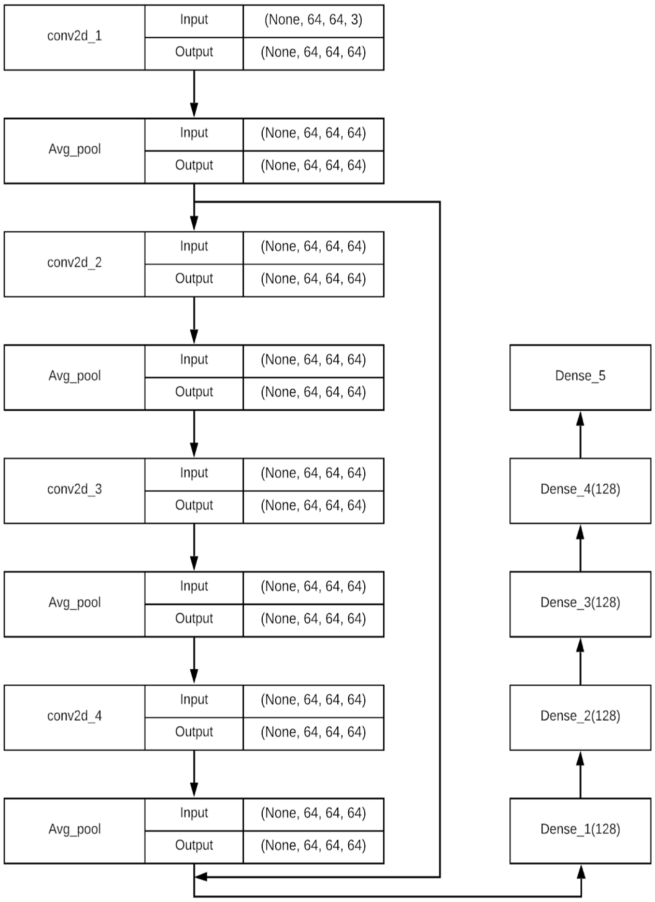
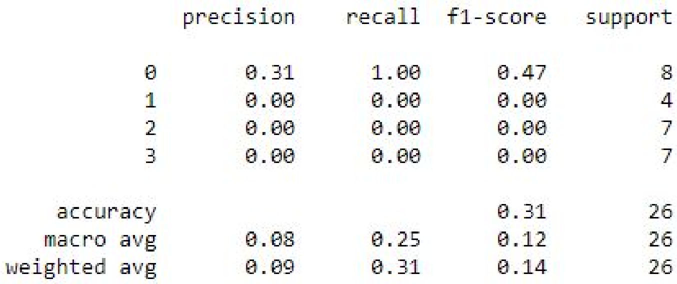

conv2d_1 | Input (None, 64, 64, 3) | Output (None, 64, 64, 64)
Avg_pool | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
conv2d_2 | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
Avg_pool | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
conv2d_3 | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
Avg_pool | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
conv2d_4 | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
Avg_pool | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
Dense_1(128)
Dense_2(128)
Dense_3(128)
Dense_4(128)
Dense_5

Fig.10: Initial Architecture of Skip-connection CNN

The above image is the initial skip connection which was tried and below are the metrics for this.


```
              precision    recall  f1-score   support

           0       0.31      1.00      0.47         8
           1       0.00      0.00      0.00         4
           2       0.00      0.00      0.00         7
           3       0.00      0.00      0.00         7

    accuracy                           0.31        26
   macro avg       0.08      0.25      0.12        26
weighted avg       0.09      0.31      0.14        26
```

Then we have changed the skip connections of the above architecture and made a connection from layer 1 to layer 3.

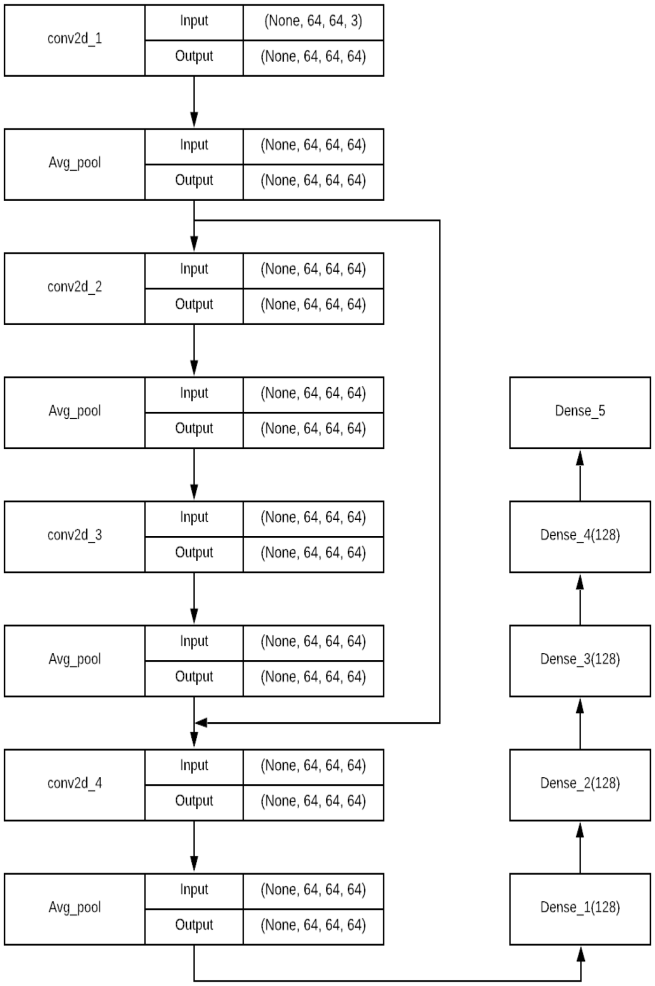
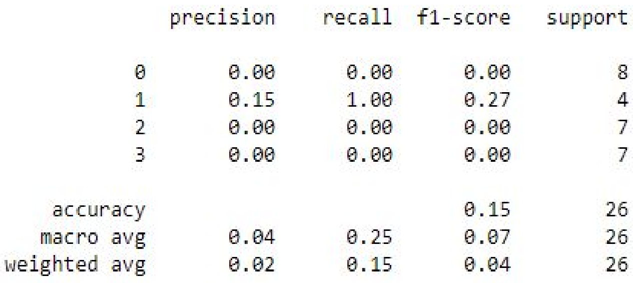

conv2d_1 | Input (None, 64, 64, 3) | Output (None, 64, 64, 64)
Avg_pool | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
conv2d_2 | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
Avg_pool | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
conv2d_3 | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
Avg_pool | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
conv2d_4 | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
Avg_pool | Input (None, 64, 64, 64) | Output (None, 64, 64, 64)
Dense_1(128)
Dense_2(128)
Dense_3(128)
Dense_4(128)
Dense_5

Fig.11: Second Architecture of Skip-connection CNN

Metrics for the above architecture are posted below.


```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         8
           1       0.15      1.00      0.27         4
           2       0.00      0.00      0.00         7
           3       0.00      0.00      0.00         7

    accuracy                           0.15        26
   macro avg       0.04      0.25      0.07        26
weighted avg       0.02      0.15      0.04        26
```

As we can see that even for this architecture the results were not good. So we went with adding two skip connections (one from layer 1 to 3 and other from layer 2 to 4). The architecture for which can be seen below.

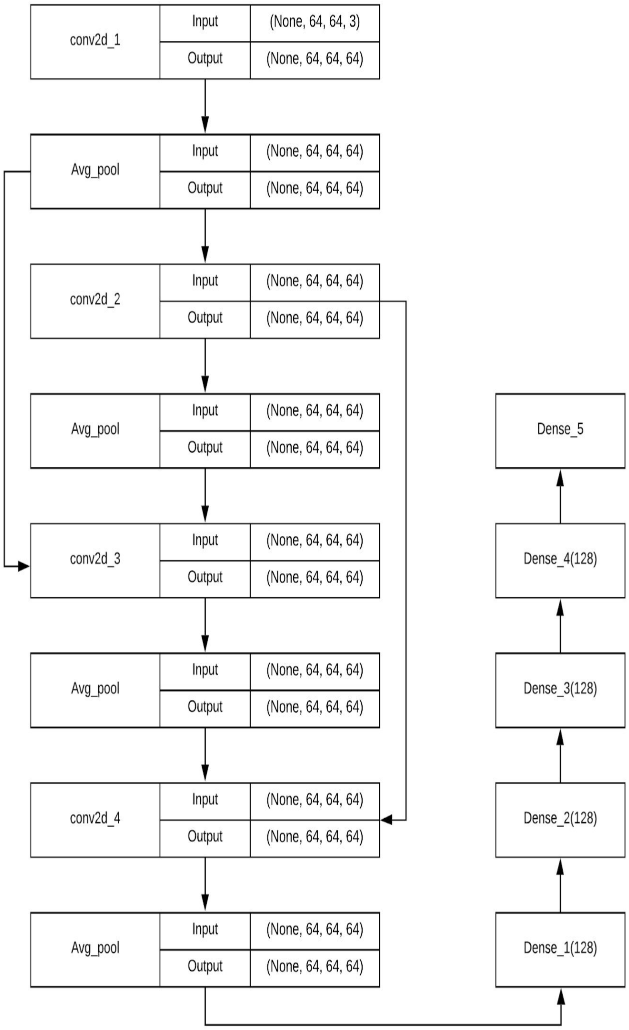
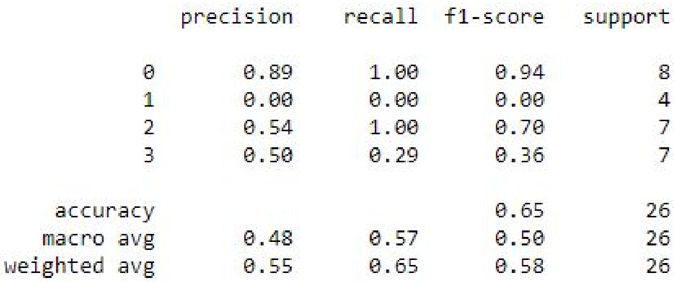


Fig.12: Architecture of Skip-connection CNN

Below are the metrics for the architecture tried in Fig.10.


```
              precision    recall  f1-score   support

           0       0.89      1.00      0.94         8
           1       0.00      0.00      0.00         4
           2       0.54      1.00      0.70         7
           3       0.50      0.29      0.36         7

    accuracy                           0.65        26
   macro avg       0.48      0.57      0.50        26
weighted avg       0.55      0.65      0.58        26
```

TABLE I
THE ARCHITECTURE DETAILS FOR THE PCG SIGNAL CLASSIFICATION COLLECTED FROM VARIOUS SOURCES WITH DIFFERENT CARDIAC ABNORMALITY

Architecture Details		CNN	Skip CNN
No. of neurons in the stacked layers		64	64
Zero padding (Border mode)		yes	yes
Filter size		3	3
Pool length		2	2
No. of neurons in the dense layers		128	128
Output Layer	Dataset 1	2	2
	Dataset 2	4	4
	Dataset 3	3	3

I. EXPERIMENTAL RESULTS

A. Dataset Description

PCG Data collected from Clinical and Non- clinical Environment (Database 1): This data source is a part of physionet challenge 2016. It is collected in either clinical or non-clinical environments from several healthy and unhealthy persons around the world. The details of the dataset are explained below in Table II. Table II summarizes the number of samples considered from each class to train and test the model. From Table II, it is observed that there are a total of 2575 normal signals and 665 abnormal signals. The ratio of 80 - 20 split is followed for training and testing the model. The dataset has two classes, normal and abnormal.

PCG data collected using ipro phone app and Digi scope (Dataset 2 & 3): The data source is a part of an associated event AISTATS 2012, sponsored event of PASCAL. Two types of dataset are available. Dataset 2 is collected using the I pro phone app using iStethoscope and the other dataset 3 is collected in a clinical environment using digital stethoscope. Table 1V describes the summary of the dataset available in the AISTATS 2012 challenge and the train test split considered for the proposed method. Dataset 2 has 4 classes. They are normal, murmur, extrasystole and artifact. Dataset 3 has 3 classes namely normal, murmur and extrasystole.

TABLE II
THE SUMMARY OF PCG DATABASE AVAILABLE 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

Result Analysis:

TABLE IV
HYPERPARAMETERS CONSIDERED IN PCG SIGNAL CLASSIFICATION

HYPERPARAMETERS	CNN	Skip CNN
Batch Size	32	32
Learning rate	0.1	0.1
No.of hidden layers	8	8
Optimizer	Adam	Adam
No.of epochs	50	50

The deep learning architectures used are CNN and Skip Residual CNN. In [2], for the dataset collected from clinical and non-clinical environments, CNN gives the promising results when trained with the raw signals. The above mentioned architectures are trained and tested with all the three databases (dataset 1, dataset 2 and dataset 3). The performance of the model is evaluated using standard metrics: precision, recall and F1- Score. The results are displayed in Table V. Here it is seen that there is an increase in classification performance of dataset 2 obtained using the spectrogram method when compared with the existing method.

It is observed from our work, that there is no increase in the performance when the Skip Residual CNN architecture is trained using the existing method, when compared with the CNN output for all the three databases.

In case of database 1 ,the work done has been exactly replicated to the previous existing classification performance 82%.

In case of database 2, the proposed work improved the existing classification performance from 82% to 85%. The performance of the work done for PCG classification is compared against the existing database for all the three databases. The corresponding results based on the standard performance metrics are tabulated in Table V.

In case of dataset 3, we have achieved better results than previously mentioned performance as we have not used fft but normalized the signal and did 1D convolutions.

TABLE V
PERFORMANCE COMPARISON OF PCG CLASSIFICATION FOR THE PROPOSED WORK AGAINST THE EXISTING WORK

DATA	WORK	METRICS	CNN	Skip CNN
1	Proposed	Accuracy	0.82	0.44
		Precision	0.66	0.19
		Recall	0.82	0.44
		F1-score	0.73	0.27
	Existing	Accuracy	0.82	-
		Precision	0.83	-
		Recall	0.82	-
		F1-score	0.83	-
2	Proposed	Accuracy	0.85	0.65
		Precision	0.85	0.55
		Recall	0.85	0.65
		F1-score	0.84	0.58
	Existing	Accuracy	0.80	-
		Precision	0.85	-
		Recall	0.79	-
		F1-score	0.82	-
3	Proposed	Accuracy	0.83	0.33
		Precision	0.81	0.11
		Recall	0.83	0.33
		F1-score	0.80	0.16
	Existing	Accuracy	0.75	-
		Precision	0.81	-
		Recall	0.76	-
		F1-score	0.79	-

IV. CONCLUSION

In this work we found that data preprocessing is an important task for training. As we can see dataset 2 has given a very good output as compared to dataset 1 and 3. And class imbalance problems can lead to overfitting of data of the majority class. Therefore we used downsampling.

As we can see that CNN has given better output as compared to the popular belief that Skip Residual CNN gave a better output. We presume that Skip Residual CNN architecture does not deal with the problem of vanishing gradient .

As we have noticed the dataset 3 is for feature extraction we used fixed signal length and normalised the signals and performed 1D convolutions with them which gave better classification results than that of fft results.

In [8], a previous work done on the usage of skip connections in Biomedical Image Segmentation, only long skip connections are used to skip the features from the containing path to recover spatial information lost during the process, and the review of gradient flow confirms that for a deep layer only its beneficial to have both long and short skip connections.

REFERENCES

[1] G. Amit, N. Gavriely, and N. Intrator, "Cluster analysis and classification of heart sounds," *Biomedical Signal Processing and Control*, vol. 4, no. 1, pp. 26–36, 2009.

[2] I. Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, "Support vectors machine-based identification of heart valve diseases using heart sounds," *Computer methods and programs in biomedicine*, vol. 95, no. 1, pp. 47–61, 2009.

[3] Performance Improvement of Deep Learning Architectures for Phonocardiogram Signal Classification using Fast Fourier Transform, Gopika.P*, Sowmya.V, Gopalakrishnan.E.A, Soman.K.P

[4] V. Sujadevi, K. Soman, R. Vinayakumar, and A. P. Sankar, "Anomaly detection in phonocardiograms employing deep learning," in Computational Intelligence in Data Mining. Springer, 2019, pp. 525–534.

[5] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk,"Segmentation of heart sound recordings by a duration-dependent hidden markov model," Physiological measurement, vol. 31, no. 4, p. 513, 2010.

[6] https://www.princeton.edu/~cuff/ele201/files/spectrogram.pdf

[7] https://convert.ing-now.com/audio-spectrogram-creator/

[8] Drozdzal, Michal & Vorontsov, Eugene & Chartrand, Gabriel & Kadoury, Samuel & Pal, Chris. (2016). The Importance of Skip Connections in Biomedical Image Segmentation. 10.1007/978-3-319-46976-8_19.