Automatic Identification of Heart Abnormalities Using PCG Signals

V. Pravin, Narendran Srinivasan[✉], Palukuri Rohith, U. Vishnu Arvind,
and Devi Vijayan

Department of Electronics and Communication Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore, India
`cb.en.u4ece18233@cb.students.amrita.edu, v_devi@cb.amrita.edu`

Abstract. A phonocardiogram (PCG) signal holds aural information generated by the heart during a cycle. A close examination of the PCG signal can reveal valuable cardiac information thereby allowing detection of abnormalities and diagnosis of heart diseases. An automation-aided analysis of PCG signals can play a vital role in the medical field, especially in remote patient monitoring, apart from being a very efficient approach. In this study, PCG signals are classified under 5 different classes based on the features extracted. The five classes are normal, mitral stenosis, mitral regurgitation, mitral valve prolapse, aortic stenosis (N, MS, MR, MVP, AS). Mel-Frequency Cepstral Coefficients (MFCCs) are extracted from the PCG audio signals and fed into a deep learning based convolutional neural network (CNN). The proposed approach achieves a maximum accuracy of 99.64% which outperforms the existing state-of-the-art approaches.

Keywords: Mel-Frequency Cepstral Coefficients · 2-D Convolutional Neural Network · Cross Validation

1 Introduction

Among all the necessities for the normal functioning of the human body, a healthy heart is the most important. It is the heart that carries out mechanical and electrical activities to ensure blood is pumped to all parts of the body. A problem in the functioning of the heart can therefore be devastating. Cardiovascular diseases (CVDs) are a very common cause of demise for individuals. According to WHO surveys, approximately 33% of all deaths are related to CVDs. An early and accurate detection of abnormalities or diseases can essentially save the lives of countless individuals. Amidst the most popular modalities that exist to monitor the health of a functioning heart are electrocardiogram (ECG), photoplethysmography (PPG) and phonocardiogram [12] (PCG). An ECG signal is a recording of the electrical activity of the heart; a PPG estimates the blood flow rate by employing light based sensors; PCG signals are audio recordings of heart sounds and murmurs present in one cardiac cycle.

A PCG signal is obtained using a machine called phonocardiograph. It uses a high-fidelity microphone to record the sounds and murmurs made by the heart. There are two fundamental heart sounds in every PCG signal - S1 and S2. These are caused by the atrioventricular and semilunar valves during their closure, and are also what we generally associate with the 'lub-dub' sound our hearts make. The interval between S1 and S2 is called systole ('lub') and the vice versa is called diastole ('dub'). A normal PCG signal contains only S1 and S2, however abnormalities cause other sounds or murmurs to arise and can be labeled as S3, S4 and so on.

Traditionally, a doctor analyses the sounds produced by the heart using a stethoscope and tries to identify any abnormality in the rhythm or the sound. This is a very difficult skill that requires years of exposure to gain proficiency at. Also, there are a myriad of limitations to the human ear as it ages that make detection of pathological symptoms quite inaccurate.

In this paper, MFCCs [4, 14] have been employed because of the similarities in properties that PCG signals have with speech signals. 26 such coefficients are extracted from a single frame. After extraction of features, a 2-D convolutional network ensues that classifies each audio signal into one of the five classes mentioned earlier.

The following graphical representation depicted in Figs. 1, 2, 3, 4 and 5 are PCG signals from individuals having N, MR, MS, MVP and AS conditions.
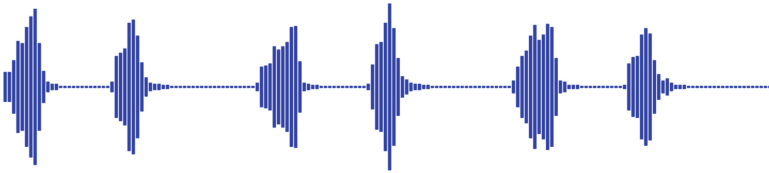
N-Type PCG Signal

Fig. 1. N type PCG signal

MR-Type PCG Signal

Fig. 2. MR type PCG signal

MS-Type PCG Signal

Fig. 3. MS type PCG signal
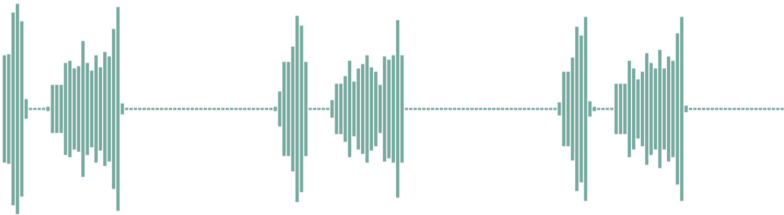
MVP-Type PCG Signal

Fig. 4. MVP type PCG signal
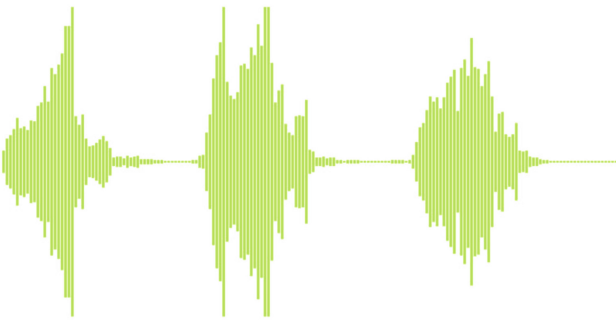
AS-Type PCG Signal

Fig. 5. AS type PCG signal

2 Related Work

Chowdhury et al. [1] employs DWT to decompose the PCG signals into multiple sub-bands having different frequencies. The sub-bands which contain unnecessary noise are dropped. For feature extraction, MFCC and Mel-scaled power spectrograms (Mel- Scale) are used. The latter is then fed through a 5-layered feed-forward DNN model trained by keras. The model has an accuracy, specificity and sensitivity of 97.10%, 94.86% and 99.26% respectively.

K. Poudel et al. [2] encountered a problem of an unbalanced dataset and employed a pre-processing method called SMOTE (Synthetic Minority Over-Sampling Technique) to counter it. Mel-Scale and MFCCs have been used for feature extraction from the PCG signals. They then pass this to a 1-D CNN model that has 4 hidden layers. The layers have been implemented with the ReLu activation function having filters of sizes 128 to 1024, with each increment doubling in size. The PCG signal is then classified in the database. The authors have used Shannon energy envelopes to develop a segmentation technique. The model has an accuracy of 93.20%, specificity of 94.20% and sensitivity of 89.20%.

Alkhodhari et al. [3] have used the combination of CNN and Bi-LSTM for the automatic extraction of features from the PCG signals. The VHD classes namely AS, MR, MVP, MS were preprocessed by MODWT and z-scoring normalization. The model was tested and trained using a 10-fold cross validation with CNN-Bi-LSTM network as well as CNN and Bi-LSTM individually. The model has an Accuracy of 99.32%, specificity of 99.58% and Sensitivity of 98.30%

The work of N. Baghela et al. [4] proposes a machine learning model to automatically diagnose CVDs using PCG signal. The model has a combination of 1-D CNN layers and Dense layers. Extensive preprocessing such as pitch correction, amplitude normalization, etc were done along with augmentation to increase the dataset size. The model was trained and evaluated using 10-fold cross validation, with an accuracy of 98.6%.

Shuvo et al. [5] have employed automatic detection of CVDs under the classes - N, AS, MR, MS and MVP using raw PCG signals. They use a CRNN architecture for this. Their model has representational and sequence residual learning phases. The time invariant features of the PCG signal are extracted using Adaptive Feature Extractor (AFE), Frequency Feature Extractor (FFE) and Pattern Extractor (PE), which are all included under representational learning. The latter includes bidirectional connections, which is used for the extraction of temporal features. Their model achieved 99.6% accuracy in the GitHub dataset and 86.57% in the Physionet dataset.

Li Oh et al. [6] proposed the WaveNet model which consisted of 6 residual blocks. 1000 PCG signals were collected from an open database which consisted of signals from 5 different classes. The signals were resampled at 8 Khz and were then normalized between -1 to 1. The model was cross-validated using 10 folds. It was trained for 3 epochs and the optimization algorithm used was Adam. The learning rate was set to 0.0005. The model has an average accuracy of 97%.

3   Proposed Methodology

2-D CNNs [13] are widely used in image recognition and object detection. For audio signals, 1-D convolutions are preferred as the kernel is only expected to slide across the time axis. In this paper, we extracted Mel Frequency Cepstral Coefficients from the audio signals. MFCCs are represented as 2-D data, with one axis representing the coefficient and the other axis representing time. We extracted 26 such coefficients. The magnitude of the frequency is represented by color. As a result, the MFCCs can be considered as a 2-D image. We have used 2048 samples in a window with a hop length of 512. The proposed methodology is depicted in Fig. 6.
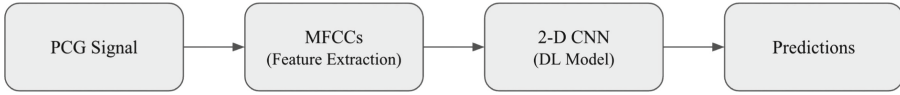
3.1 Block Diagram

Fig. 6. Block diagram

3.2 Architecture and Training

Table 1. Model architecture

Layer	#Filters, Size	#Neurons	Strides, Padding	Activation
2-D Convolution	32, (3,3)		1, 'Same'	ReLu
2-D Convolution	32, (3,3)		1, 'Same'	ReLu
2-D Max Pooling	-, (2,2)		(2,2), None	-
2-D Convolution	64, (3,3)		1, 'Same'	ReLu
2-D Max Pooling	-, (2,2)		(2,2), None	-
2-D Convolution	128, (3,3)		1, 'Same'	ReLu
2-D Max Pooling	-, (2,2)		(2,2), None	-
2-D Convolution	64, (3,3)		1, 'Same'	ReLu
Dense		512	-	ReLu
Dense		256	-	ReLu
Dense		5	-	Softmax

- Input layer: 32 filters of dimensions 3 × 3 with stride size set to 1 and padding set to 'same' and activation function set to relu, resulting in an output dimension of (26, 44, 32) (Table 1).
- Hidden Layer 1: 32 filters of dimensions 3 × 3, stride size set to 1, padding set to 'same' and activation function set to relu.
- Hidden Layer 2: 64 filters of dimensions 3 × 3 with stride size set to 1, padding set to 'same' and activation function set to relu.

- Hidden Layer 3: 128 filters of dimensions 3×3 stride size set to 1, padding set to 'same' and activation function set to relu.
- Hidden Layer 4: 64 filters of dimensions 3×3 with stride size set to 1, padding set to 'same' and activation function set to relu. The output is flattened.
- Hidden Layer 5: Dense layer comprising 512 units and activation function as relu.
- Hidden Layer 6: Dense layer comprising 256 units and activation function as relu.
- Output Layer: Dense layer comprising 5 units and activation function as softmax.

The model was trained for 15 epochs on a Tesla K80 GPU. The loss function used was categorical cross entropy, with Adam being the choice of the optimizer with a learning rate of 0.001.

4   Results and Discussion

The dataset (link included) used in this study includes a total of 1000 PCG signals from patients (inclusive of both sexes and all age groups) with normal and 4 different valvular heart diseases (MS, MR, MVP, AR). The 1000 signals are divided into the 5 classes of 200 signals each. The duration of each signal is fixed at 2 s. To evaluate the performance metrics of the model, cross validation with fold size 10 has been used.

Table 2 shows the results of the cross validation with accuracy as the parameter.

Table 2. Training and validation accurary of each fold

Fold	Training Accuracy	Validation Accuracy
1	100%	100 %
2	100 %	99.49 %
3	100 %	98.97 %
4	100 %	100 %
5	100 %	100 %
6	100 %	98.46 %
7	100 %	100 %
8	100 %	99.49 %
9	100 %	100 %
10	100 %	100 %
	Average	**99.64 %**

Table 3. Parameter values of each fold

Fold	Parameters	AS	MR	MS	MVP	N
1	Precision	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00
	F1-Score	1.00	1.00	1.00	1.00	1.00
2	Precision	1.00	0.97	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	0.97	1.00
	F1-Score	1.00	0.99	1.00	0.99	1.00
3	Precision	0.98	1.00	1.00	0.97	1.00
	Recall	1.00	0.97	1.00	0.97	1.00
	F1-Score	0.99	0.99	1.00	0.97	1.00
4	Precision	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00
	F1-Score	1.00	1.00	1.00	1.00	1.00
5	Precision	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00
	F1-Score	1.00	1.00	1.00	1.00	1.00
6	Precision	0.98	1.00	1.00	0.97	0.98
	Recall	1.00	0.95	1.00	0.97	1.00
	F1-Score	0.99	0.97	1.00	0.97	0.99
7	Precision	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00
	F1-Score	1.00	1.00	1.00	1.00	1.00
8	Precision	1.00	1.00	1.00	0.97	1.00
	Recall	1.00	1.00	0.97	1.00	1.00
	F1-Score	1.00	1.00	0.99	0.99	1.00
9	Precision	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00
	F1-Score	1.00	1.00	1.00	1.00	1.00
10	Precision	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00
	F1-Score	1.00	1.00	1.00	1.00	1.00
Average	Precision	0.99	0.99	1.00	0.99	1.00
	Recall	1.00	0.99	0.99	0.99	0.99
	F1-Score	0.99	0.99	0.99	0.99	1.00

The lowest validation accuracy was 98.46% and the highest validation accuracy was 100%. The mean validation accuracy across all the folds was 99.64%.

Table 3 shows the performance of the model for each class on metrics such as precision, recall and F1-scores for all 10 folds. The following parameters are calculated as follows:

$$Precision = \frac{TP}{FP + TP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1\text{-}Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

Table 3 shows the parameter values for all the folds while Table 4 compares the model presented in this paper with other models.

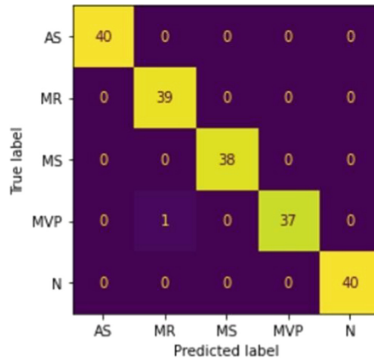The following figures present the confusion matrices for folds that do not have a validation accuracy of 100%.

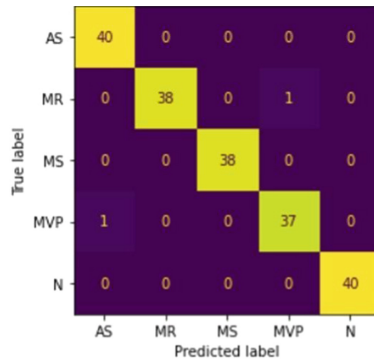**Fig. 7.** Confusion matrix for Fold 2
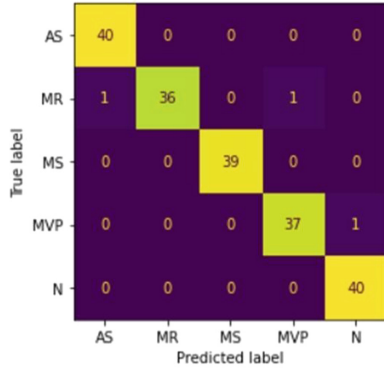
Fig. 8. Confusion matrix for Fold 3
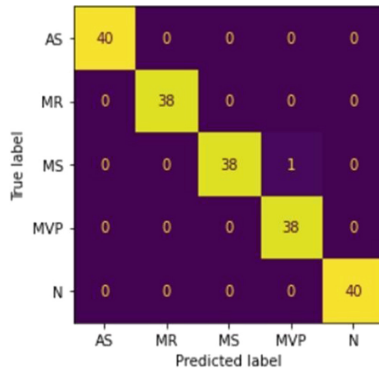
Fig. 9. Confusion matrix for Fold 6

Fig. 10. Confusion matrix for Fold 8

From Figs. 7, 8, 9 and 10 it is evident that the misclassifications have occurred at certain instances.

- In the confusion matrix for fold 2, as shown in Fig. 7, 1 signal attributed to MVP has been misclassified as MR, resulting in an overall accuracy of 99.49%.
- For fold 3, 1 MVP signal has been misclassified as AS and 1 MR signal has been misclassified as MVP, lowering the overall accuracy to 98.97%.
- Fold 6 shown in Fig. 9 has the most number of misclassifications and hence the least overall accuracy of 98.46%. 2 MR signals have been incorrectly classified as AS and MVP respectively. In addition to this, 1 MS signal has been misclassified as N.
- In fold 8, 1 MS signal has been classified as MVP thereby resulting in an overall accuracy of 99.49%.

MR is incorrectly classified three times, while MVP and MS signals are misclassified twice.

Even though MFCCs are not traditional two-dimensional images, the 2-D CNN model was able to perform surprisingly well. It matches and even surpasses the performance of 1-D CNN and LSTM [15] in some cases.

Table 4. Study comparison

Model	Features	Accuracy
MD Chowdhury et al. [1]	MFCC, Mel-Scale	97.10%
M. Chowdhury et al. [2]	MFCC	93.20%
Li Oh et al. [6]	Deep Learned	97.00%
Yaseen et al. [9]	MFCC, Wavelet	87.2%
P.Lubaib et al. [10]	MFCC, envelope detection	99.01%
Ghosh et al. [11]	Chirplet transform	98.33%
Proposed Study	**MFCC, 2-D CNN**	**99.64%**

5   Conclusion

Manual detection of heart abnormalities is a challenging and time-consuming task that requires specific expertise. This study proposes a computer aided diagnosis (CAD) system using 2-D CNN for classification of cardiovascular diseases. 2-D CNNs are uncommon in the audio domain, but continue to gain traction. The proposed method achieves an average 10-Fold cross validation accuracy of 99.64%, which surpasses many other state of the art models in this dataset. This model does not require extensive pre-processing and is relatively light-weight. The overall accuracy of the model may be further improved by performing data augmentation.

The main limitation of the proposed work is the lack of multi-class PCG datasets. While there are multiple datasets for binary PCG signal datasets, it is not the case for non-binary datasets.

References

1. Chowdhury, T.H., Poudel, K.N., Hu, Y.: Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals. IEEE Access **8**, 160882–160890 (2020). https://doi.org/10.1109/ACCESS.2020.3020806
2. Chowdhury, M., Poudel, K., Hu, Y.: Detecting abnormal PCG signals and extracting cardiac information employing deep learning and the shannon energy envelope. IEEE Signal Process. Med. Biol. Symp. **2020**, 1–4 (2020). https://doi.org/10.1109/SPMB50085.2020.9353624
3. Alkhodari, M., Fraiwan, L.: Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. Comput. Methods Programs Biomed. **200**, 105940 (2021). https://doi.org/10.1016/j.cmpb.2021.105940
4. Baghel, N., Dutta, M.K., Burget, R.: Automatic diagnosis of multiple cardiac diseases from PCG signals using convolutional neural network. Comput. Methods Programs Biomed. **197**, 105750 (2020). https://doi.org/10.1016/j.cmpb.2020.105750. Epub 2020 Sep 10 PMID: 32932128

5. Shuvo, S.B., Ali, S.N., Swapnil, S.I., Al-Rakhami, M.S., Gumaei, A.: CardioXNet: a novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings. IEEE Access **9**, 36955–36967 (2021). https://doi.org/10.1109/ACCESS.2021.3063129

6. Oh, S.L., et al.: Classification of heart sound signals using a novel deep WaveNet model. Comput. Methods Programs Biomed. **196**, 105604 (2020). https://doi.org/10.1016/j.cmpb.2020.105604. Epub 2020 Jun 12 PMID: 32593061

7. Ismail, S., Siddiqi, I., Akram, U.: Localization and classification of heart beats in phonocardiography signals—a comprehensive review. EURASIP J. Adv. Sig. Process. **2018**(1), 1–27 (2018). https://doi.org/10.1186/s13634-018-0545-9

8. Yang, T.-C., Hsieh, H.: Classification of acoustic physiological signals based on deep learning neural networks with augmented features. In: 2016 Computing in Cardiology Conference (CinC), pp. 569–572 (2016)

9. Yaseen, Son, G.-Y., Kwon, S.: Classification of Heart Sound Signal Using Multiple Features. Appl. Sci. **8**, 2344 (2018). https://doi.org/10.3390/app8122344

10. Lubaib, P., Ahammed Muneer, K.V.: the heart defect analysis based on PCG signals using pattern recognition techniques. Procedia Technol. **24**, 1024–1031, ISSN 2212–0173. https://doi.org/10.1016/j.protcy.2016.05.225

11. Ghosh, S.K., Ponnalagu, R.N., Tripathy, R.K., Acharya, U.R.: Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals. Comput. Biol. Med. **118**, 103632 (2020). https://doi.org/10.1016/j.compbiomed.2020.103632. Epub 2020 Jan 30 PMID: 32174311

12. Kesav, R.S., Bhanu Prakash, M., Kumar, K., Sowmya, V., Soman, K.P.: Performance improvement in deep learning architecture for phonocardiogram signal classification using spectrogram. In: Singh, M., Tyagi, V., Gupta, P.K., Flusser, J., Ören, T., Sonawane, V.R. (eds.) ICACDS 2021. CCIS, vol. 1440, pp. 538–549. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-81462-5_48

13. Kishore, S.L.S., Sidhartha, A.V., Reddy, P.S., Rahul, C.M., Vijaya, D.: Detection and diagnosis of Covid-19 from chest X-ray images. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 459–465 (2021). https://doi.org/10.1109/ICACCS51430.2021.9441862

14. Supriya, P., Jayabarathi, R., Jeyanth, C., Ba, Y., Sarvesh, A., Shurfudeen, M.: Preliminary Investigation for Tamil cine music deployment for mood music recommender system. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 1111–1115 (2020). https://doi.org/10.1109/ICACCS48705.2020.9074249

15. Sujadevi, V.G., Soman, K.P., Vinayakumar, R., Sankar, A.U.P.: Deep models for phonocardiography (PCG) classification. In: 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT), pp. 211–216 (2017). 10.1109/ INTEL-CCT.2017.8324047