ML-Based Analysis of Phonocardiogram Signals for Heart Sound Classification

[1]SVH Nagendra*, [2]Vijay K Pandey
[1,2]Department of Mechanical Engineering, Vivekananda Global University, Jaipur, India
[1*]svh.nagendra@vgu.ac.in
[2]vijay_pandey@vgu.ac.in

Abstract. Cardiovascular diseases are one of the major components of the global health budget. Phonocardiography (PCG), the recording of heart sounds, offers a non-invasive method for cardiac assessment, with potential for automated analysis to aid diagnosis. This study compares the effectiveness of five distinct machine learning approaches for classifying PCG recordings as 'normal' or 'abnormal.' Data were collected from 40 participants (ages 10-30) using a PCG machine at four standard auscultation locations. Recordings were classified by five experts. Five models were implemented and evaluated: two Random Forest (RF) classifiers using aggregated Mel-Frequency Cepstral Coefficients (MFCCs) from different cardiac segments, a 1D Convolutional Neural Network (CNN) using sequences of cycle-based features, a Bidirectional Long Short-Term Memory (BiLSTM) network using similar cycle sequences augmented with timing features, and a Transfer Learning approach using DenseNet201 on Mel Spectrogram images. Performance was evaluated using Test Accuracy and ROC AUC. Results showed that the RF models performed poorly, particularly in class discrimination (Test ROC AUC < 0.4). The deep learning models significantly outperformed RF, with the BiLSTM (Test Acc: 0.7453, Test ROC AUC: 0.8046) and DenseNet201 (Test Acc: 0.749, Test ROC AUC: 0.79) achieving the highest test scores. However, the BiLSTM demonstrated better alignment between validation and test performance compared to the DenseNet model, which showed signs of potential overfitting or poor generalization based on lower validation scores. The findings highlight the importance of sequence modeling and appropriate feature representation for PCG classification and suggest the BiLSTM approach offered the most robust performance in this evaluation.

Keywords: Cardiovascular; PCG; Machine Learning; Accuracy: Heart Sound

1 Introduction

Cardiovascular diseases are the leading cause of mortality worldwide, imposing a significant health concern. The economic impact is also substantial, affecting national healthcare systems and individual households, particularly in developing nations. So, early and accurate diagnosis is very crucial for effective management and reducing mortality. The term "auscultation," which is listening to heart sounds using a stethoscope, has been a tested method of cardiovascular physical examination for centuries.

B. Singh et al. (eds.), *Proceedings of the International Conference on Advances in Computing Technology and Artificial Intelligence (COMPUTATIA 2026)*, Atlantis Highlights in Intelligent Systems 18,
https://doi.org/10.2991/978-94-6239-713-2_14

It provides valuable diagnostic clues about the heart's function, including valve operation and blood flow dynamics. However, the proficiency of manual auscultation is highly dependent on the clinician's skill and experience, and studies have indicated that there is a potential decline in these skills among new trainees. Furthermore, the interpretation is subjective, and grading murmur intensity, for instance, can lack consistency. Phonocardiography (PCG), the technique of recording heart sounds graphically, offers a more objective and permanent record of cardiac acoustic events. PCG allows for the detection of subtle sounds and provides a basis for quantitative analysis. The collaboration of digital signal processing and machine learning has opened new avenues for computer-aided analysis of PCG signals. Automated PCG analysis holds the promise of providing accessible, low-cost, and objective tools for screening and diagnosis. It shows that it is potentially complementing the traditional methods and supporting telemedicine applications. Despite advancements, developing robust automated classification systems remains challenging due to signal variability, noise interference, and the need for effective feature extraction and modeling techniques. Several techniques ranging from classical machine learning approaches to deep learning models have been experimented with. In this study, we attempt to add to this growing list by comparatively evaluating the performance of five machine learning models, including Random Forest and 1DCNN, BiLSTM, and Transfer Learning with DenseNet201—applied to the task of classifying PCG recordings as normal or abnormal, based on labels derived from expert auscultation.

2. Literature Review

In the present state of development of computational cardiology, the accessibility and quality of data resources is one of the central themes, which has been emphasized by a number of major initiatives. The source CirCor DigiScope Dataset was a pioneering work with 5282 phonocardiogram (PCG) recordings of 1568 pediatric patients with a total of more than 312 hours. There are two types of importance to this dataset: the scale and the extensive annotations of 215,780 heart sounds, which, in addition to the characteristics of the murmur, such as timing, shape, pitch, quality, location, and grading, are specifically created to take research beyond just basic murmur recognition and into detailed classification with data-intensive machine learning models.

This dataset (version 1.0) was formally released on websites such as PhysioNet, where it would be available to researchers in large numbers. Supplying PCG-focused resources, [1] created the EPHNOGRAM database, which contains open access to both simultaneously measured electrocardiogram (ECG) and PCG signals to allow conducting research regarding the interaction between the electrical and mechanical activity of the heart. Previous initiatives, such as the database presented by [2], were also intended to present standardized and open-source heart sound data with clinical data to support the evaluation of algorithms. Moreover, the DigiScope app technology proposed by [3] proved that big amounts of real-world data, which are collected and annotated in clinical practices, can be achieved. Even challenges such as the PhysioNet/Computing in Cardiology Challenge 2016 and the PASCAL Heart Sound Classification Challenge inspired innovation by offering benchmark datasets and tasks, which unsurprisingly tend to emphasize the role of segmentation and complex machine learning pipelines using feature engineering (e.g., MFCCs, wavelets) and classifiers (e.g., SVMs, ensembles) or deep learning to classify normal or abnormal heart sounds. The necessity of standardized data is also supported by the fact that [4] proposes that age groups should be standardized in pediatric trials, which would enhance consistency and comparability. There has been a major advancement in automated analysis of heart sounds, especially in segmentation and classification using advanced signal analysis and machine learning. Proper segmentation—the definition of S1, systole, S2, and diastole—is essential. The adaptive sojourn time Hidden Semi-Markov Model (HSMM) allowed [5] to enhance the segmentation accuracy, particularly when it comes to variable heart rate. Another high-accuracy segmentation technique introduced by [6] was a combination of logistic regression and HSMMs. Convolutional Neural Networks (CNNs) are a potent instrument of deep learning, and [7] has successfully utilized it to discover features directly on the signal, which can be used to do precise segmentation. In addition to segmentation, AI is being used in diagnostics; [8] created a deep learning system that promises to diagnose PCG-pediatric Congenital Heart Disease (CHD) using PCG signals. Scientists are also addressing individual issues in auscultation: [9] focused on the issue of distinguishing between systolic ejection clicks and split S1 sounds only by ear, whereas [10] have managed to use signal processing to differentiate between the fetal and maternal sounds during pregnancy. [11] came up with computational models that simulated the fetal PCGs, and these models helped in testing the algorithm.

The conceptual basis of such analyses is based heavily on digital audio and discrete-time signal processing, as discussed in introductory textbooks by [12], [13], and [14], and advanced research on immersive audio processing may become a future direction. The performance of biometric systems based on heart sound has also been tested, but in this case, the variability and noise are also a challenge. The practice of the implementation of such systems also presupposes the data exchange protocols' consideration, which is demonstrated by [15], who has discovered that protocol selection greatly determines real-time decision support system performance. The technological endeavors are based on the clinical reality of cardiovascular disease (CVD), a primary cause of mortality and morbidity in the world. The prevalence of CVD (with the main influence of hypertension) and its contributing risk factors, considerable healthcare expenditure, and its economic outcome are also continuously emphasized in statistical updates provided by the American Heart Association [16] and European organizations [17]. Quantifying this heavy economic burden in a variety of settings, such as low- and middle-income nations and the EU, a significant cost of acute events and treatment is often higher than per capita health spending [18], [19], and [20]. One important component of this burden is valvular heart disease (VHD), where such studies as [21] have determined that even non-symptomatic adults have non-negligible prevalence rates that are both age-dependent and hypertension-related. It is of importance to understand certain valve conditions, and [22] examined the assessment issues of degenerative mitral stenosis; [23] presented the clinical entity and findings of pulmonic stenosis; and [24] gave an overview of ventricular septal defect (VSD), with its characteristic murmur. Clinical skills, especially the art of auscultation, are important as they underpin accurate diagnosis, as described in such lengthy textbooks as Heart Disease by Braunwald [25] and in guides such as Clinical Methods [26] and Karnath and Thornton reviews [27]. Nonetheless, there are fears about the decreasing skills in the art of auscultation. Continue, as emphasized by [28] and exhibited in such studies as [29], which identified shortcomings among trainees. This has led to an interest in simulation-based training, which is widely used in UK medical schools but inconsistently [30], and technological aids, such as point-of-care ultrasound (also known as "in sonance"), which [31] recommend as the fifth pillar of physical examination. Telemedicine is becoming increasingly important, too; networks as described by [32] and [33] have been shown to increase access to pediatric cardiology in Brazil, and [34] established the viability of teleoperated robotic interfaces in remote access to auscultation. It is a fundamental competence to differentiate between innocent murmurs, common in children, and pathological ones, and this has been discussed in reviews by [35] and [36] and the pioneering work by [37], which emphasized the importance of careful assessment even of seemingly innocent systolic murmurs. It is important to understand the nature of the murmurs, such as the musical quality of the research of [38], and the physiological provenance of the sounds ([39] review; [40] phono-echocardiographic correlation of S2). Even rare results, such as diastolic murmurs in seemingly healthy children, should be subjected to further inquiry [41]. Subjective traditional grading could be improved with objective methods of grading murmurs that are assessed by [42].

The pediatric condition involves certain considerations as discussed by [43], such as increased heart rates and innocent murmurs. Although the accuracy of the suspected CHD screening in newborns by means of an auscultation is moderately good, which makes it an effective screening method, confirmation of the condition through an echocardiography is needed [44]. Fundamentals of Phonocardiogram (PCG): Fig. 1 shows the sample classification of S1, S2, systole, and diastole. The cardiac cycle produces distinct sounds primarily associated with valve closures. The first heart sound (S1), often described as "lub," corresponds mainly to the closure of the atrioventricular (mitral and tricuspid) valves at the onset of systole. The second heart sound (S2), "dup," marks the end of systole and is associated with the closure of the semilunar (aortic and pulmonary) valves. The timing and characteristics of S1 and S2, including potential splitting (asynchronous valve closure), provide diagnostic information. Pathological conditions often manifest as additional sounds or alterations in S1 and S2. Heart murmurs, typically longer sounds resulting from turbulent blood flow, are often indicative of valvular defects (stenosis or regurgitation), septal defects, or other structural abnormalities. Murmurs can be systolic or diastolic, and their timing, shape (e.g., crescendo-decrescendo), location, intensity, and quality (e.g., harsh, musical) are key diagnostic features. Innocent murmurs, common especially in children, must be distinguished from pathological ones. Other sounds like ejection clicks or gallops (S3, S4) can also indicate specific conditions. It is important to know the physiological causes of these sounds in order to know how to interpret them. PCG Technology and Datasets: The acquisition of heart sounds and subsequent analysis have been made possible by the change from acoustic stethoscopes to digital recording systems (digital stethoscopes and PCG devices). Recording techniques and equipment characteristics (transducers, filters, and amplifiers) are to be standardized to be reliably used in clinical applications. Public datasets have made a great contribution to the study of automated PCG analysis. The most prominent are the datasets of the PhysioNet/Computing in Cardiology Challenge (e.g., 2016, 2022), where they give the labels of normal and abnormal, as well as detect the murmur. CirCor DigiScope provides a big library of pediatric PCGs in their recordings with copious annotations. The EPHNOGRAM data bank offers recorded ECG and PCG concurrently. There are also special data, including fetal PCG data.

2.1 PCG Signal Processing

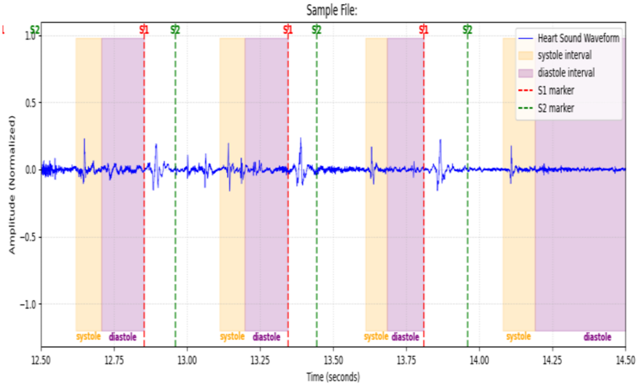


Fig. 1. Sample classification of S1, S2, systole, and diastole

Analyzing raw PCG signals typically involves several preprocessing steps. Noise reduction is often necessary, employing techniques like filtering or wavelet de-noising. Segmentation, the process of identifying the boundaries of S1, systole, S2, and diastole, is a critical step for many analysis methods. Various algorithms based on signal characteristics, Hidden Markov Models (HMMs), or deep learning have been developed for segmentation. Feature extraction aims to derive quantitative descriptors from the PCG signal that capture discriminative information. Common features include:

Time-domain features, which identify durations of S1, S2, systole, and diastole; amplitude characteristics; and zero-crossing rate (ZCR). The frequency-domain features are derived using spectral analysis techniques like the Fast Fourier Transform (FFT) or Short-Time Fourier Transform (STFT) to derive spectral centroid, bandwidth, and rolloff. Mel-Frequency Cepstral Coefficients (MFCCs) are widely used in speech processing. The time-frequency features are derived in the form of wavelet transforms or spectrograms/Mel spectrograms, which provide details of variation in frequency content over time.

2.2 Machine Learning in PCG Classification

Table 1. Code number and name

Code Number	Model Name
Code 1	RF
Code 2	RF (HyperTuned)

Code 3	1D CNN
Code 4	BiLSTM
Code 5	DenseNet201

Table 1 shows the 5 different models implemented for training and testing purposes. The machine learning (ML) and deep learning (DL) techniques are increasingly used for automated PCG classification tasks, such as distinguishing normal from abnormal sounds or identifying specific pathologies. Random Forest is an ensemble learning method that builds multiple decision trees and aggregates their predictions. It has been applied to PCG classification, often using handcrafted features. RF is known for its robustness and ability to provide feature importance measures, though it relies on effective feature engineering. Studies have used RF with features derived from wavelet decomposition, spectral analysis, and sparse representations. Convolutional Neural Networks (CNNs), particularly 1D CNNs for sequential data or 2D CNNs for image-like representations (spectrograms), have shown strong performance in PCG analysis. They can automatically learn hierarchical features from the input. CNNs have been used with raw or filtered PCG signals, MFCC sequences, or time-frequency representations like Mel spectrograms as input. They have been applied to both segmentation and classification tasks. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) variants, are well-suited for modeling sequential data like PCG signals due to their ability to capture temporal dependencies. They have been used for segmentation and classification. Often, RNNs are combined with CNNs (CNN-LSTM architectures) to leverage CNNs' feature extraction capabilities and RNNs' sequence modeling strengths. Features like MFCCs or wavelet features are commonly used as input to LSTM/BiLSTM models.

Transfer learning (TL) techniques adapt models pre-trained on large datasets (often ImageNet for image-based models) to the PCG classification task. This typically involves converting PCG segments into spectrograms or Mel spectrograms and treating them as images. Pre-trained CNN architectures like AlexNet, ResNet, DenseNet, MobileNet, and Inception, or even models pre-trained on large audio datasets, are fine-tuned for heart sound classification. TL can be particularly beneficial when labeled PCG data is limited.

3. Methodology

The aim is to compare the performance of different machine learning paradigms on a specific PCG classification task using a consistent dataset and expert-derived labels.

3.1 PCG Signal Processing

Phonocardiogram recordings were collected from 40 participants aged between 10 and 30 years. A PCG recording device was used to capture heart sounds at four standard auscultation locations on the chest: aortic valve (AV), pulmonic valve (PV), tricuspid valve (TV), and mitral valve (MV). The recordings were saved as WAV audio files using the PCG instrument, as shown in Figure 2.

Fig. 2. PCG machine used for data acquisition

3.2 PCG Signal Processing

The recordings: Each WAV file recording was independently listened to by three clinical experts. Based on their auscultation findings, each expert classified the recording as either 'normal' or 'abnormal.' For the purpose of this study, a consensus or majority vote among the experts was used to assign the final binary label (Normal/Abnormal) to each recording, forming the ground truth for model training and evaluation.

3.3 Preprocessing

Common preprocessing steps were applied across most implementations (Codes 1-4) before feature extraction. This involved using a bandpass filter on the raw audio signal to remove frequencies not of interest to heart sounds and remove noise, and then peak normalization to normalize the magnitude scale. Code 5, which makes use of spectrogram images, had its own processing pipeline: audio resampling, padding or truncation to constant length, El spectrogram conversion, 224*224 pixel resizing, RGB conversion, and DenseNet-specific input processing. Code 5 data was also divided into train/validation directory structures that can be used in image classification structures.
Different feature sets were extracted for the various models: In Code 1 the Mel-Frequency Cepstral Coefficients (MFCCs) were extracted. Specifically, the mean, standard deviation, delta (first derivative), and delta-delta (second derivative) were calculated for 30 MFCCs. These features were extracted only from S1 and S2 segments, identified using external TSV files. Features from S1 and S2 were concatenated into a single vector per recording. For Code 2, the same set of 30 MFCCs and their derivatives were extracted, but from all four major cardiac segments: S1, Systole, S2, and Diastole. Features were calculated for each segment type and then concatenated into one vector per recording. Zero vectors were used for missing/short segments. In case of Code 3, the features were extracted per segment (S1, Sys, S2, Dia) within identified heart cycles. The feature set included 13 MFCCs, Zero-Crossing Rate (ZCR), Spectral Centroid, Spectral Bandwidth, Spectral Rolloff, and Segment Duration. Features from segments within a cycle were concatenated to form a cycle vector.

In Code 4, it uses the same segment features as Code 3 (13 MFCCs, ZCR, etc.) per segment within cycles. Additionally, explicit timing features were calculated for each cycle: systole duration, diastole duration, S1-S2 interval, cycle duration, and systole/diastole ratio. All segment and timing features were concatenated into one vector per cycle. Finally, in Code 5, the input feature was the Mel spectrogram image itself (224x224x3), generated from fixed-duration audio chunks.

3.4 Model Architectures and Training

The five distinct models were implemented and evaluated: In Code 1 & 2, standard Random Forest classifiers were used, taking the aggregated feature vectors (from S1/S2 or all segments) as input. Code 3, the 1D convolutional neural network, was designed to process sequences of cycle-based feature vectors. Sequences were padded/truncated to a fixed maximum length (Max_Cycle) to ensure consistent input dimensions for the CNN.

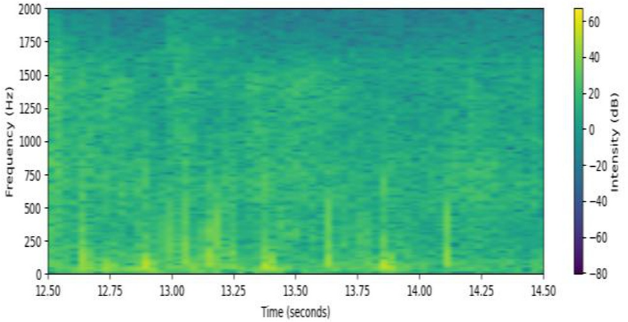


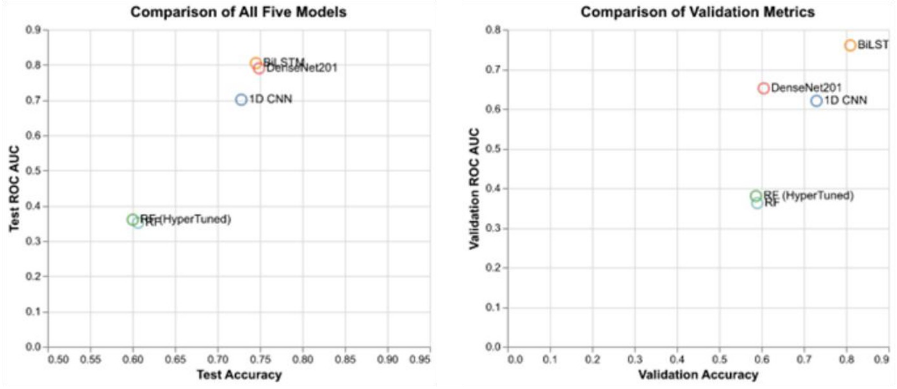
Fig. 3. Sample spectrogram extracted from the data

The Bidirectional Long Short-Term Memory (Code 4) network was implemented, also taking sequences of cycle-based feature vectors (augmented with timing features) as input, with padding/truncation applied. In Code 5, a pre-trained DenseNet201 model (trained on ImageNet) was adapted using transfer learning. The original classification layer was removed, and new dense layers were added as a classification head. The process involved an initial phase of training only the new head with the base model's weights frozen, followed by a fine-tuning phase where some later layers of the DenseNet base were unfrozen and trained with a low learning rate. Models were trained using appropriate splits of the expert-labeled dataset into training and validation sets. Performance was ultimately evaluated on a held-out test set.

4. Evaluation Metrics

The primary metrics used to evaluate and compare the models were test accuracy and test area under the receiver operating characteristic curve (ROC AUC). Validation accuracy and validation ROC AUC were also monitored during training, where available, to assess generalization and guide model selection or identify potential overfitting.

5. Results



(a) (b)

Fig. 4. (a) Test Accuracy vs. Test ROC/AUC for all models (b) Validation Accuracy vs validation ROC AUC for all models

The performance of the five implemented models on the held-out test set, along with relevant validation metrics, is summarized in Table 2.

Table 2. Performance Summary of Implemented Models

Model Name	Test Accuracy	Test ROC AUC	Val Accuracy	Val ROC AUC
RF	0.6065	0.3514	0.59	0.362
RF (HT)	0.6003	0.3602	0.587	0.38
1D CNN	0.7281	0.7009	0.73 (epoch 29)	0.62 epoch 9
BiLSTM	0.7453	0.8046	0.81 epoch (2)	0.76 epoch (1, 5, 8)
DenseNet201	0.749	0.79	0.6053	0.6516

The Random Forest models (Codes 1 and 2) had test accuracies of about 60% but had very low test ROC AUCs (0.3514 and 0.3602), which means that they performed worse than random chance predicting between classes. The 1D CNN model (Code 3) was much better with a test accuracy of 72.81 and a fair test ROC AUC of 0.7009. The BiLSTM model (Code 4) further enhanced the performance since it attained a test accuracy of 74.53% and the maximum test ROC AUC of 0.8046. Code 5 was the DenseNet201 transfer learning model that achieved the best test accuracy of 74.9% and a good test ROC AUC of 0.79. Code 5, however, had a significant difference in test and validation performance (Val Acc: 0.6053, Val ROC AUC: 0.6516).

6. Discussion

The findings are useful in the assessment of the suitability of various machine learning approaches to the classification of PCG signals according to the expert labels in the given case. The most impressive aspect is the obvious insufficiency of the models of Random Forest (Codes 1 and 2) that were based on the aggregated MFCC features. It was found that with seemingly fair accuracy (around 60 percent), their ROC AUC scores were extremely low (less than 0.4), which shows that they have failed to learn a meaningful separation between normal and abnormal classes. This is a strong indication that the process of aggregation, which eliminates the temporal order of the cardiac events, obliterates important diagnostic information that is inherent in the dynamics of the PCG signal. The mere addition of the features of more segments (Code 2 vs. Code 1) did not help resolve this problem, which supports the notion that the temporal structure is primary. On the contrary, all three deep learning methods (Codes 3, 4, and 5), i.e., those that used information on time or frequency, showed significantly higher performance. The 1D CNN shift to sequences of cycle-based features (Code 3) provided a substantial improvement in accuracy, as well as, more critically, ROC AUC. This brings out the advantage of maintaining the sequential order of the heart cycles and the model learning patterns across the cycles despite its comparatively simple CNN architecture and a more detailed set of engineered features.

The two sequence-based deep-learning models compared, i.e., BiLSTM (Code 4), performed better than the 1D CNN (Code 3), especially on ROC AUC (0.8046 vs. 0.7009). This implies that the architecture of the BiLSTM, which is particular to identifying long-range temporal dependencies in both forward and backward directions and possibly even explicitly timing information (duration, ratio), led to an advantage in the ability to capture the complexity of dynamics distinguishing normal and abnormal heart sounds within this dataset. The validation performance of the BiLSTM, which peaked at the beginning of the training, seemed fairly consistent with the test performance, specifically, ROC AUC (~0.76 vs. 0.8046). The highest test accuracy and a competitive test ROC AUC were obtained with the transfer learning method based on DenseNet201 in spectrograms (Code 5). It shows how effective it can be to utilize strong pre-trained image models in the analysis of time-frequency representations of audio signals. Nevertheless, the large difference in performance between the test set and the validation set (e.g., Test Acc 0.749 and Val Acc 0.6053, respectively) begs serious questions concerning the generalization of the model in this particular implementation. The explanation of this gap could be due to overfitting of the fine-tuning stage as the model was learning spurious correlations with the training/validation data or possibly because of variation between the data distributions in the validation and test splits. Even though the test scores are high, the validation performance is poor, which, without additional research and, possibly, alternative fine-tuning approaches or data augmentation, makes the strength of this model doubtful.

Altogether, the findings are a solid case in support of the methodologies that retain and model the time structure of PCG signals. The use of sequence-based models such as CNNs and LSTMs over cycle-level features was much better than aggregation-based models. Of the deep learning models that have been experimented with, the BiLSTM

model (Code 4) appeared to be potentially the most trustworthy, with high performance over the test set and relatively stable validation metrics, indicating superior generalization over the transfer learning model (Code 5) in this particular experimental design.
There are some limitations to this study. The sample size (40 patients) is quite limited, and the age group (10-30 years old) might not be representative of the general population, in which CVDs are more common. The subjectivity of the expert labels, though clinically significant, can be subjective, but this can be addressed through the use of more than one expert. It would be a good idea to conduct further studies based on more extensive and more varied data and possibly using objective ground truth (e.g., the results of echocardiography).

7. Conclusion

Random forest models do not provide enough aggregation in the case of a random forest. The aggregated MFCC features had low performance, especially in the classification of classes (Test ROC AUC < 0.4), which shows that the removal of the temporal information is harmful. The BiLSTM models that are typically described as sequence modeling are better than 1D CNNs. The deep learning models that took sequences of features based on heart cycles were much more effective as compared to the aggregation-based models, which demonstrates the significance of temporal context. The BiLSTM model with both segment properties and explicit timing information after each cycle showed high test accuracy and the highest test ROC AUC, indicating that it was most appropriate to describe temporal variations in PCG signals.
The BiLSTM approach (Code 4) seemed to have the most promising and potentially reliable methodology in the entire range of those evaluated in this study based on the concerted performance on the test and consistency of validation. Moreover, transfer learning is a promising solution. The approach with the best test accuracy was the DenseNet201 transfer learning applied to spectrograms, although it had a large difference in performance between test and validation, making it questionable whether it was robust and generalized to new samples in this context. Overall, to classify PCGs using expert labels in the context of this study, sequence-based deep learning models, namely the BiLSTM with the use of appropriate temporal and spectral characteristics, have shown to be more effective and reliable than the traditional aggregation algorithms and the intended transfer learning.

Acknowledgements: We thank the Swavalamban Chair for support in the research.

References

1. A. Kazemnejad, P. Gordany, and R. Sameni, "EPHNOGRAM: A simultaneous electrocardiogram and phonocardiogram database," *PhysioNet*, Jun. 2021, doi: 10.13026/tjtq-5911.
2. C. Liu, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, 2016.

3. D. Pereira, "DigiScope—unobtrusive collection and annotating of auscultations in real hospital environments," in *Proc. IEEE Conf. Eng. Med. Biol. Soc.*, 2011, pp. 1193–1196.
4. K. Williams, "Standard 6: Age groups for pediatric trials," *Pediatrics*, vol. 129, no. 3, pp. S153–160, 2012.
5. J. Oliveira, F. Renna, and M. T. Coimbra, "Adaptive sojourn time HSMM for heart sound segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 642–649, Mar. 2019.
6. D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-HSMM-based heart sound segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 822–832, Apr. 2016.
7. F. Renna, J. H. Oliveira, and M. T. Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2435–2445, Nov. 2019.
8. B. Xiao, "Follow the sound of children's heart: A deep learning-based computer-aided pediatric CHDs diagnosis system," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1994–2004, Mar. 2020.
9. N. M. Hoeting, "Systolic ejection click versus split first heart sound: Are our ears deceiving us," *Congenital Heart Dis.*, vol. 12, no. 4, pp. 417–420, 2017.
10. D. Riknagel *et al.*, "Separation and characterization of maternal cardiac and vascular sounds in the third trimester of pregnancy," *Int. J. Gynecol. Obstet.*, vol. 137, no. 3, pp. 253–259, Jun. 2017.
11. M. Cesarelli, M. Ruffo, M. Romano, and P. Bifulco, "Simulation of foetal phonocardiographic recordings for testing of FHR extraction algorithms," *Comput. Methods Programs Biomed.*, vol. 107, no. 3, pp. 513–523, 2012.
12. U. Zölzer, *Digital Audio Signal Processing*. Hoboken, NJ, USA: Wiley, 2008.
13. J. Watkinson, *The Art of Digital Audio*, 3rd ed. Waltham, MA, USA: Focal Press, 2001.
14. A. Oppenheim, R. Schafer, and J. Buck, *Discrete-Time Signal Processing*. Hoboken, NJ, USA: Prentice Hall, 1999.
15. P. Gomes, S. Faria, and M. Coimbra, "The effect of data exchange protocols on decision support systems for heart sounds," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 5384–5387.
16. E. J. Benjamin *et al.*, "Heart disease and stroke statistics 2019 update: A report from the American Heart Association," *Circulation*, vol. 139, no. 10, pp. e56–e528, Mar. 2019.
17. E. Wilkins *et al.*, *European Cardiovascular Disease Statistics*. Bruxelles, Belgium: European Heart Network, 2017.
18. A. Gheorghe *et al.*, "The economic burden of cardiovascular disease and hypertension in low- and middle-income countries: A systematic review," *BMC Public Health*, vol. 18, p. 975, 2018.
19. A. Murphy *et al.*, "The household economic burden of non-communicable diseases in 18 countries," *BMJ Glob. Health*, vol. 5, no. 2, p. e002040, 2020.

20. J. Leal, R. Luengo-Fernández, A. Gray, S. Petersen, and M. Rayner, "Economic burden of cardiovascular diseases in the enlarged European union," *Eur. Heart J.*, vol. 27, no. 13, pp. 1610–1619, 2006.
21. M. S. Kim *et al.*, "Frequency and clinical associating factors of valvular heart disease in asymptomatic Korean adults," *Sci. Rep.*, vol. 9, no. 1, p. 16741, Nov. 2019.
22. A. A. Oktay *et al.*, "Echocardiographic assessment of degenerative mitral stenosis: A diagnostic challenge of an emerging cardiac disease," *Curr. Problems Cardiol.*, vol. 42, no. 3, pp. 71–100, Mar. 2017.
23. H. A. Snellen, H. Hartman, T. N. Buis-Liem, E. H. Kole, and J. Rohmer, "Pulmonic stenosis," *Circulation*, vol. 38, no. 1s5, pp. V-93–V-101, Jul. 1968.
24. L. E. Ampie and S. El-Amin, "Ventricular Septal Defect," in *Congenital Heart Disease - Molecular Genetics, Principles of Diagnosis and Treatment*, R. Laxer, Ed. London, U.K.: Springer, 2015, pp. 289–297.
25. P. Libby, R. Bonow, D. Mann, and D. Zipes, *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 8th ed. Amsterdam, The Netherlands: Elsevier Science, 2007.
26. H. K. Walker, W. D. Hall, and J. W. Hurst, *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed. Boston, MA, USA: Butterworths, 1990.
27. B. Karnath and W. Thornton, "Auscultation of the heart," *Hosp. Physician*, vol. 38, no. 9, pp. 39–43, 2002.
28. M. Chizner, "Cardiac auscultation: Rediscovering the lost art," *Curr. Problems Cardiol.*, vol. 33, pp. 326–408, 2008.
29. S. Mangione, "Cardiac auscultatory skills of physicians-in-training: A comparison of three English-speaking countries," *Amer. J. Med.*, vol. 110, no. 3, pp. 210–216, 2001.
30. S. J. Owen and K. Wong, "Cardiac auscultation via simulation: A survey of the approach of UK medical schools," *BMC Res. Notes*, vol. 8, p. 427, Sep. 2015.
31. J. Narula, Y. Chandrashekhar, and E. Braunwald, "Time to add a fifth pillar to bedside physical examination: Inspection, palpation, percussion, auscultation, and insonation," *JAMA Cardiology*, vol. 3, no. 4, pp. 346–350, Apr. 2018.
32. S. D. S. Mattos *et al.*, "A telemedicine network for remote paediatric cardiology services in north-east Brazil," *Bull. World Health Org.*, vol. 93, pp. 881–887, 2015.
33. S. D. S. Mattos *et al.*, "Impact of a telemedicine network on neonatal mortality in a state in northeast Brazil," *Popul. Health Manage.*, vol. 21, no. 6, p. 517, 2018.
34. S. Falleni, A. Filippeschi, E. Ruffaldi, and C. A. Avizzano, "Teleoperated multimodal robotic interface for telemedicine: A case study on remote auscultation," in *Proc. 26th IEEE Int. Symp. Robot Hum. Interactive Commun.*, 2017, pp. 476–482.
35. A. R. Doshi, "Innocent heart murmur," *Cureus*, vol. 10, no. 12, p. e3689, Dec. 2018.

36. A. Rosenthal, "How to distinguish between innocent and pathologic murmurs in childhood," *Pediatr. Clin. North Amer.*, vol. 31, no. 6, pp. 1229–1240, 1984.
37. A. Freeman and S. Levine, "The clinical significance of the systolic murmur. A study of 1000 consecutive ‘non-cardiac’ cases," *Ann. Int. Med.*, vol. 6, pp. 1371–1385, 1933.
38. P. L. Stein, H. N. Sabbah, and J. B. Lakier, "Origin and clinical relevance of musical murmurs," *Int. J. Cardiol.*, vol. 4, no. 1, pp. 103–112, Aug. 1983
39. . E. Koetting, "The physiological origins of heart sounds and murmurs: The unique interactive guide to cardiac diagnosis," *J. Invest. Surg.*, vol. 10, no. 6, pp. 397–399, Nov. 1997.
40. Prakash, "Second heart sound: A phono-echocardiographic correlation in 20 cardiac patients," *J. Amer. Geriatrics Soc.*, vol. 26, no. 8, pp. 372–374, 1978.
41. J. Liebman and S. Sood, "Diastolic murmurs in apparently normal children," *Circulation*, vol. 38, no. 4, pp. 755–762, Oct. 1968.
42. R. Keren, M. Tereschuk, and X. Luan, "Evaluation of a novel method for grading heart murmur intensity," *Arch. Pediatrics Adolesc. Med.*, vol. 159, no. 4, pp. 329–334, 2005.
43. C. Ratti *et al.*, "Cardiac auscultation in children," *Recenti Progressi in Medicina*, vol. 105, no. 11, pp. 469–472, Nov. 2014.
44. N. D. Brunetti *et al.*, "Diagnostic accuracy of heart murmur in newborns with suspected congenital heart disease," *J. Cardiovasc. Med.*, vol. 16, no. 8, pp. 556–561, Aug. 2015.