

# Modified Variable Kernel Length ResNets for Heart Murmur Detection and Clinical Outcome Prediction Using Phonocardiogram Recordings

Vijay Vignesh Venkataramani<sup>1\*</sup>, Akshit Garg<sup>1\*</sup>, U Deva Priyakumar<sup>1</sup>

<sup>1</sup> International Institute of Information Technology, Hyderabad, India

\* Authors contributed equally and are co-first authors.

## Abstract

In this work, we describe an end-to-end deep learning architecture for Heart Murmur Detection from Phonocardiogram(PCG) recordings as part of The George B. Moody PhysioNet Challenge 2022. Our team, “Team IIITH” received a weighted accuracy score of 0.708 (ranked 19<sup>th</sup> out of 40 teams) and Challenge cost of 13264 (ranked 22<sup>nd</sup> out of 39 teams) on the official hidden test set.

In our approach, the PCG recordings are first down-sampled to 1000 Hz before being passed through a Butterworth’s low and high pass filter to remove baseline wanders and high-frequency noise present in the recordings. The PCG recordings are then broken down into 10-second segments and normalized to bring all trainable samples to the same size. To extract embeddings more efficiently, we built a custom 1-dimensional Residual Network (ResNet) where the 10-second inputs are passed through variable-sized kernel ResNets in parallel, before being concatenated and passed through the next ResNet layer to account for different length dependencies across the PCG signal. The output of this custom ResNet is then fed to a 2-layer feed-forward network for final classification. Cross-Entropy Loss with class weights was employed to account for class imbalance. Our approach obtained a 5-fold Cross-Validation weighted accuracy score of 0.71 and challenge cost score of 12067 on the training set.

## 1. Introduction

Unusual audible traces of the heartbeat are called heart murmurs. Although heart murmurs can be harmless (commonly referred to as innocent heart murmurs), in other circumstances, they might be a sign of an underlying cardiac condition [1–3]. The development of machine learning (ML) models for heart murmur identification is crucial because identifying heart murmurs can aid in the early detection of underlying heart disorders. In the PhysioNet/Computing in Cardiology Challenge 2022 [4–6], we were provided with multi-positional phonocardiogram

(PCG) recordings for heart murmur detection. PCGs are the audio recordings of the heartbeats which are recorded with the help of either a normal stethoscope or an electronic stethoscope. This paper will discuss our proposed work employing Residual Networks (ResNets) [7–10] to identify the heart murmurs using the PCG recordings.

## 2. Methods

In this section we will first put forward the data processing techniques we used. We will then discuss the model architecture employed for the challenge. Moving forward we will discuss about the model training and model evaluation techniques we employed to predict the heart murmur and clinical outcome using the patient’s PCG signal. We used a 5 fold cross validation technique to generate train and test set. It was ensured that the 10-second segments for all the folds came from non overlapping patients.

### 2.1. Data Preprocessing

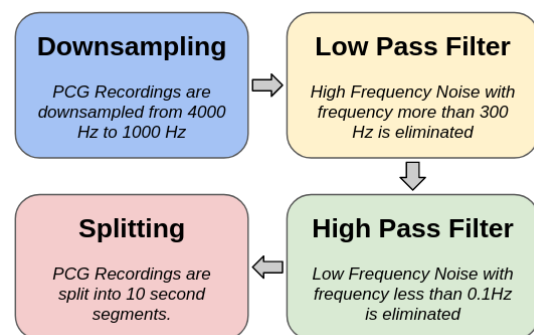


Figure 1. Data Preprocessing Pipeline

The dataset provided in the challenge had PCG Recordings recorded at 4000 Hz. Figure 1 demonstrates the data preprocessing pipeline we utilized for the challenge. We first downsampled the data from 4000 to 1000 Hz as data collected with a high sampling rate might hamper the train-

ing of the Convolutional Neural Networks (CNNs) [11, 12]. We then used Butterworth [13, 14] low pass filter set at a frequency cutoff of 300Hz to ensure that high-frequency noise with more than 300Hz is eliminated. This was then followed by a high pass filter where the cutoff frequency was set to 0.1Hz to ensure low-frequency noise and baseline wander from the PCG recordings could be eliminated. The PCG recording of each patient was then divided into equal-length, 10-second segments, with the intention of using each segment as a single data point. We utilised a stride of 2.5 seconds to generate segments. Thus when a PCG recording was divided into 10-second segments, the subsequent segments had an overlap of 7.5 seconds while the remaining 2.5 seconds were distinct. We were able to preserve consistent length PCG across all data points thanks to signal splitting, which also helped us grow the dataset size. For each segment of PCG, the labels were kept same as the original PCG from which the segment was derived. After segmentation we were able to derive 17890 10-second segments from 942 patients. For heart murmur prediction task we ended up with 3342 10-second segments where the heart murmur was present while 13696 segments were generated where the heart murmur was not present. There were 849 segments where it was not certain whether the heart murmur was present or not. For the clinical outcome prediction task we had 9205 segments where the clinical outcome was usual while there were 8685 segments where the clinical outcome was unusual. Finally we split the dataset according to a 5 fold cross-validation setup, where it was made sure that no patients overlapped between folds, to ensure unbiased testing results.

## 2.2. Model Architecture

Figure 2 gives the complete model architecture employed for the challenge. For both the tasks of detecting cardiac murmurs and the prediction of clinical outcomes, we used the same model architecture. We obtained an embedding for the PCG signal using 1D Residual Networks (ResNets). A fully connected neural network was then given the PCG embedding to make the final prediction. We employed different kernel length ResNets because the big kernel length CNNs aid in extracting information from a large portion of the signal at once, while the smaller kernel CNNs can extract the signal attributes with a narrower context. To obtain the final embedding, the outputs from each ResNet with a different kernel size are concatenated, which should be able to account for different length dependencies across the PCG signal. We experimented with two distinct concatenation methods. For concatenation along the length (concatenated length), we concatenated the outputs from all the ResNet blocks along the length; as a result, the concatenated signal had length greater than the individual unconcatenated signals but the number of filters in the con-

catenated signal are same as that of the incoming unconcatenated signals. For the second approach (concatenated filters) we concatenated signals along the filters; hence the number of filters in the concatenated signal increased although the length of concatenated signal remained same to that of the incoming unconcatenated signals. Multiple layers of ResNet blocks followed by concatenation operation were used to get the final embedding.

## 2.3. Model Training

After preprocessing we were able to generate multiple PCG segments with respect to a single patient. Each of this PCG segment was used as a separate datapoint for training. For the two tasks of heart murmur detection and clinical outcome prediction two separate models were trained, although the model architecture used for both the task was the same. For the task of heart murmur prediction initially there were 3 classes, present (3342 10-second segments), absent (13696 10-second segments) and uncertain (849 10-second segments). As the number of 10-second segments in uncertain class were very low, we decided to ignore the PCG signals with label as uncertain, thus the task changed from multi class classification to binary class classification. As the number of 10-second segments for present class were much less compared to the absent class, we used weighted Cross Entropy loss to eliminate the class imbalance. For the task of clinical outcome prediction both classes had similar representation hence both the classes were given equal weightage. Adam optimiser was used for both the tasks, where the learning rate was set at 0.0001. The model was trained for 30 epochs on the training set.

## 2.4. Model Evaluation

Model evaluation strategy needs to differ from model training as in evaluation we need to provide a single label for a patient. Given that the model can only operate on 10-second segments, we used the same data preprocessing techniques described above. To obtain the final diagnosis for a patient, the labels for each of its 10-second parts are put together. When at least 25% of the segments yielded positive results for the heart murmur, we concluded that the patient's PCG had a murmur. We settled on a low threshold of 25% since it is significantly more detrimental to anticipate a false negative than a false positive. For the heart murmur prediction task we ignored the uncertain class therefore all the predictions for the test set would lie in the two classes where either the heart murmur is present or absent. For the task of clinical outcome prediction when at least 60% of the segments yielded unusual for the clinical outcome, we predicted that the patient's PCG will have an unusual clinical outcome.

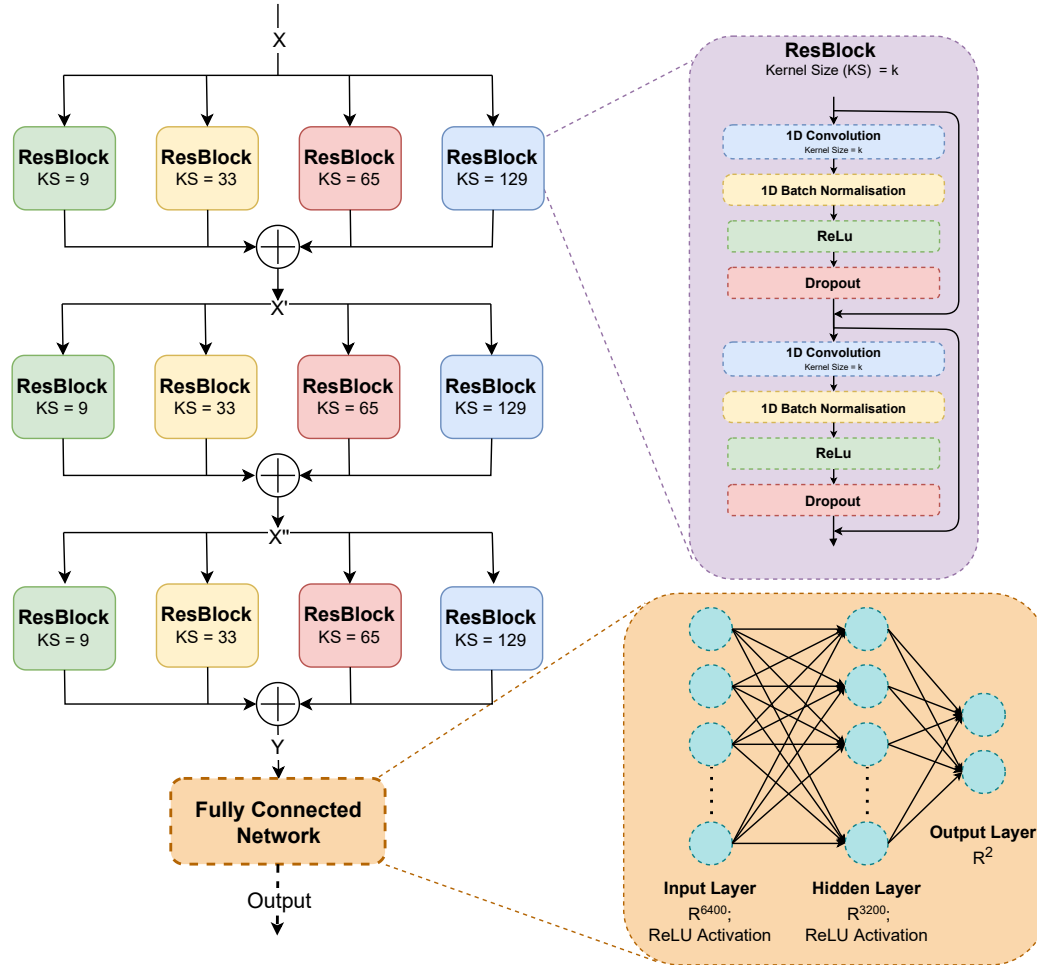


Figure 2. Model Architecture (*KS*: Kernel Size,  $\oplus$ : Concatenation Operation)

### 3. Results

The models were tested with a five fold cross validation setup on our local system for both kinds of architectures (concatenated length/concatenated filters). Refer to Table 1 and 2 for the 5 fold cross validation results on the local environment. The concatenated length setup obtained a better score for both the heart murmur detection and outcome prediction tasks. Hence, for the final submitted model, both the tasks are handled by the concatenated length architecture. Our final models were tested in the Official Phase and obtained a weighted accuracy of 0.708 and challenge cost of 13264, for the murmur and outcome tasks, fetching the rank of 19 and 22 out of the selected 40 teams on the hidden test set.

### 4. Discussion and Conclusions

To determine the best approach for the problem at hand, we tested a variety of models and machine learning meth-

Training	Validation	Test	Ranking
0.71 $\pm$ 0.04	0.678	0.708	19/40

Table 1. Weighted accuracy metric scores (official Challenge score) for our final selected entry (team "Team.IIITH") for the murmur detection task, including the ranking of our team on the hidden validation set. We used 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set.

Training	Validation	Test	Ranking
12067 $\pm$ 2653	11266	13264	22/40

Table 2. Cost metric scores (official Challenge score) for our final selected entry (team "Team.IIITH") for the clinical outcome identification task, including the ranking of our team on the hidden validation set. We used 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the official hidden test set.

ods for the challenge.

We believed that a transformer-based model architecture could be useful for the assigned task. To create embedding with regard to the PCG signal, we attempted to train a transformer model [15]. This embedding was then sent to a feed-forward network for final prediction. For all of the datapoints, the transformer-based model was unable to accurately predict the presence of a heart murmur. Given that transformers require large amount of data for accurate training, we feel that the data provided for the challenge was insufficient. We utilised Mel-frequency cepstral coefficients (MFCCs) [16] after a transformer. We calculated the MFCCs for each PCG segment and fed the results into a feed-forward network. While the performance of the MFCC technique was better than that of the transformers, it lagged well behind the suggested variable kernel length method.

Among all the approaches we tested, the suggested variable kernel length ResNets fared the best. With the proposed method we were able to fetch 28<sup>th</sup> rank for the heart murmur detection task while we got a rank of 43 for clinical outcome prediction. We think there are a number of ways to further enhance the suggested model. To start, instead of using a straightforward weighted cross entropy loss, a custom loss that can more closely approximate the challenge score should be developed. This might aid in better model training and should improve the final model metrics. Secondly, the provided dataset was very small for training a very complex model. We believe that leveraging pretrained networks that have been trained on other very big datasets can enhance the model's performance. In order to build a model on a much larger dataset and use it to both detect heart murmurs and predict clinical outcomes, unsupervised learning techniques can also be utilised.

## Acknowledgments

We thank IHub-Data, IIIT Hyderabad for support.

## References

- [1] Ahmad MS, Mir J, Ullah MO, Shahid MLUR, Syed MA. An efficient heart murmur recognition and cardiovascular disorders classification system. *Australasian Physical Engineering Sciences in Medicine* 2019;42(3):733–743.
- [2] Doshi AR. Innocent heart murmur. *Cureus* 2018;10(12).
- [3] Sun R, Liu M, Lu L, Zheng Y, Zhang P. Congenital heart disease: causes, diagnosis, symptoms, and treatments. *Cell Biochemistry and Biophysics* 2015;72(3):857–860.
- [4] Reyna MA, Kiarashi Y, Elola A, Oliveira J, Renna F, Gu A, et al. Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. *medRxiv* 2022;URL <https://doi.org/10.1101/2022.08.11.22278688>.
- [5] Oliveira J, Renna F, Costa PD, Nogueira M, Oliveira C, Ferreira C, et al. The CirCor DigiScope dataset: from murmur detection to murmur classification. *IEEE Journal of Biomedical and Health Informatics* 2021;26(6):2524–2535.
- [6] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 2000;101(23):e215–e220.
- [7] Sadegh Ebrahimi M, Karkeh Abadi H. Study of residual networks for image recognition. *arXiv e prints* 2018;arXiv–1805.
- [8] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*. PMLR, 2015; 448–456.
- [9] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*. 2015; 1026–1034.
- [10] Ribeiro AH, Gedon D, Teixeira DM, Ribeiro MH, Ribeiro ALP, Schön TB, et al. Automatic 12-lead ecg classification using a convolutional network ensemble. In *2020 Computing in Cardiology*. IEEE, 2020; 1–4.
- [11] Fukushima K, Miyake S, Ito T. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems Man and Cybernetics* 1983;(5):826–834.
- [12] Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*. IEEE, 2017; 1–6.
- [13] Daud S, Sudirman R. Butterworth bandpass and stationary wavelet transform filter comparison for electroencephalography signal. In *2015 6th International Conference on Intelligent Systems, Modelling and Simulation*. IEEE, 2015; 123–126.
- [14] Gomez V. The use of butterworth filters for trend and cycle estimation in economic time series. *Journal of Business Economic Statistics* 2001;19(3):365–373.
- [15] Gong Y, Lai CI, Chung YA, Glass J. Ssast: Self-supervised audio spectrogram transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36. 2022; 10699–10709.
- [16] Tiwari V. MFCC and its applications in speaker recognition. *International Journal on Emerging Technologies* 2010; 1(1):19–22.

Address for correspondence:

U Deva Priyakumar  
Center for Computational Natural Sciences and Bioinformatics,  
International Institute of Information Technology, Hyderabad,  
Gachibowli, 500032, Hyderabad, India  
deva@iiit.ac.in