Phonocardiogram Classification Using 1-Dimensional Inception Time Convolutional Neural Networks

Bjørn-Jostein Singstad[1], Antony M. Gitau[2], Markus Kreutzer Johnsen[3], Johan Ravn[3], Lars Ailo Bongo[4], Henrik Schirmer[5]

[1] Simula Research Laboratory, Department of Computational Physiology, Oslo, Norway
[2] Kenyatta University, Nairobi, Kenya
[3] Medsensio AS, Oslo, Norway
[4] UiT, The Arctic University of Norway, Tromsø, Norway
[5] Akershus University Hospital, Lørenskog, Norway

Abstract

Murmurs are sounds caused by turbulent blood flow that are often the first sign of structural heart disease in patients. These sounds are detected by auscultating the heart using a stethoscope, or more recently by a phonocardiogram (PCG). We aim to identify the presence, absence, or unclear cases of murmurs, as well as predict normal or abnormal clinical outcome from PCG recordings using machine learning.

We trained and tested two 1-dimensional convolutional neural networks (CNN) on a PCG data set from a pediatric population of 1568 *individuals. One model predicted murmurs, while the other model predicted clinical outcomes. Both models were trained to give recording-wise predictions, while the final predictions were given for every patient (patient wise predictions).*

This paper describes our participation in the George B. Moody PhysioNet Challenge 2022. The objective of this challenge was to identify heart murmurs and clinical outcomes from Phonocardiogram recordings. Our team, Simulab, trained a clinical outcome classifier that achieved a challenge cost score of 8720 *(ranked 1st out of 305 submissions) and the murmur classifier achieved a weighted accuracy of* 0.585 *(ranked* 182*nd out of* 305 *submissions) on the validation set.*

1. Introduction

Cardiovascular diseases are one of the major causes of death and represent 32% of all global deaths [1]. Heart sounds provide an important source of information to a clinician in detecting abnormal murmurs which might be a sign of structural heart disease [2, 3]. The most common and cost-effective tool for acquiring heart sounds is the stethoscope [4]. However, studies show that auscultation using a stethoscope is generally poorly performed both by medical students [5] and physicians [6], and many physicians cannot reliably distinguish abnormal from normal heart sounds, especially in children [7].

A phonocardiogram (PCG) is a digital representation of a heart sound and can be recorded by a phonocardiograph. A phonocardiograph is a stethoscope that transmits the sounds to a digital sampling device instead of transmitting them to the clinician's ears like a stethoscope. Figure 1 show an example of a PCG plotted in the time domain, and also the two most prominent peaks in a cardiac cycle which are called the first heart sound (S1) and the second heart sound (S2). S1 originates from the closing of mitral and tricuspid valves after blood flows from atria to ventricles, while S2 is caused by the closing of the aortic and pulmonary valve after blood is ejected from the two ventricles. Heart sounds are usually auscultated at the four different locations on the chest wall which corresponds to the aortic, pulmonary, tricuspid and mitral valves.

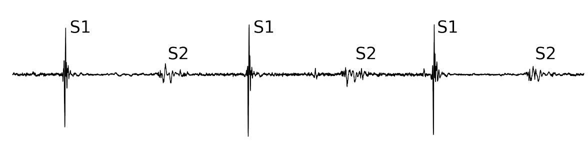

Figure 1. An example of a phonocardiogram showing three cardiac cycles

Back in 1987 Rangayyan, R. M., & Lehner, R. J. stated that: "The heart sound signal has much more information than can be assessed by the human ear or by visual inspection of the signal tracings on paper as currently practiced" [8] and since then, several attempts have been made to process [9], analyse [10] and classify PCG recordings using deep learning methods [11].

This paper describes our approach to the George B. Moody PhysioNet Challenge 2022. The objective of this challenge was to detect murmurs as present, absent or unclear and classify the clinical outcome of a patient as normal or abnormal from PCGs recorded from one or more auscultation locations.

2. Method

We use a supervised machine learning approach, a convolutional neural network (CNN), to detect murmurs and classify the clinical outcome of a patient using a single PCG signal. We implemented the models in Python (3.8.9) using Tensorflow (2.8.2). The code will be open sourced and published on GitHub [1].

2.1. Data

The data set used in this work consists of 5272 PCGs from a pediatric population of 1568 individuals [12, 13]. 3163 PCGs from 942 individuals were used for training. The remaining 2109 PCGs from 149 and 477 patients, are only available to the organizers of the challenge, and were used for validation and testing respectively. Each patient could have one or more PCG recordings taken from a location close to the aortic valve, pulmonary valve, tricuspid valve, mitral valve or in some cases unknown.

Each patient was labeled with a clinical outcome (abnormal/normal) and murmur (present/unknown/absent), annotated by a clinical expert [4]. In cases of present murmur, the location of the recorded murmur was given in the training set.

2.2. Pre-processing

2.2.1. Signal processing

The PCG signals in the training data were recorded with a sampling frequency of 4000Hz. We downsampled all signals to 100Hz. In addition, we zero padded all signals in the training data such that all signals were of a length equal to the longest signal (6451 samples long). 6451 samples were used as the threshold also for the validation and test data. Signals with length $l < 6451$ were given a zero-padded tail of length $6451 - l$ and signals longer than 6451 were truncated.

2.2.2. Label processing

The data set was relabeled from patient-wise labeling to recording-wise labeling. This was done by labeling all PCG recordings from a patient with the same clinical outcome as the original overall label. The recording-wise relabeling of murmurs, however, is shown in Algorithm 1.

Algorithm 1 : Patient to recording wise murmur labels

Input: p = patient, r = PCG record,
t = total population, l = label
Output: r_l = recording wise labels
for n in t **do**
  if $p_{n_l} = Absent$ **then**
    all r_l **in** $p_n = Absent$
  else if $p_{n_l} = Unknown$ **then**
    all r_l **in** $p_n = Unknown$
  else if $p_{n_l} = Present$ **then**
    for m **in** p_{n_r} **do**
      $r_{l_m} = Present$
    end for
  end if
end for

2.3. Models

Two classification models were trained; one model to classify murmurs and the other to classify outcomes. Both models were 1 dimensional CNNs with an Inception Time architecture [14]. The murmur model was a multi-class classifier, set to classify whether a murmur was present/absent/unknown in the heart sound recording. The outcome model was a binary classifier used to classify whether the patient would have a normal or abnormal outcome. The murmur classifier was trained using weighted categorical cross entropy, while the clinical outcome classifier was trained using weighted binary cross entropy. The weights in both models were determined to be inversely proportional to the prevalence of the classes.

2.4. Post-processing

The recording-wise predictions from the model were finally converted back to patient-wise predictions. The murmur conversion is shown in Algorithm 2 and the clinical outcome conversion is shown in Algorithm 3.

2.5. Model selection (local development)

To estimate the performance of the models and find the optimal model architectures and hyper-parameters we did local development on the training set before submitting the final model to the organizers of the challenge. The training data were divided into local training and validation sets using 5-fold CV with stratification on patient level, and new models were trained and validated each successive round.

[1] This link will be valid after the challenge is finished: https://github.com/Bsingstad/Heart-murmur-detection-2022-private

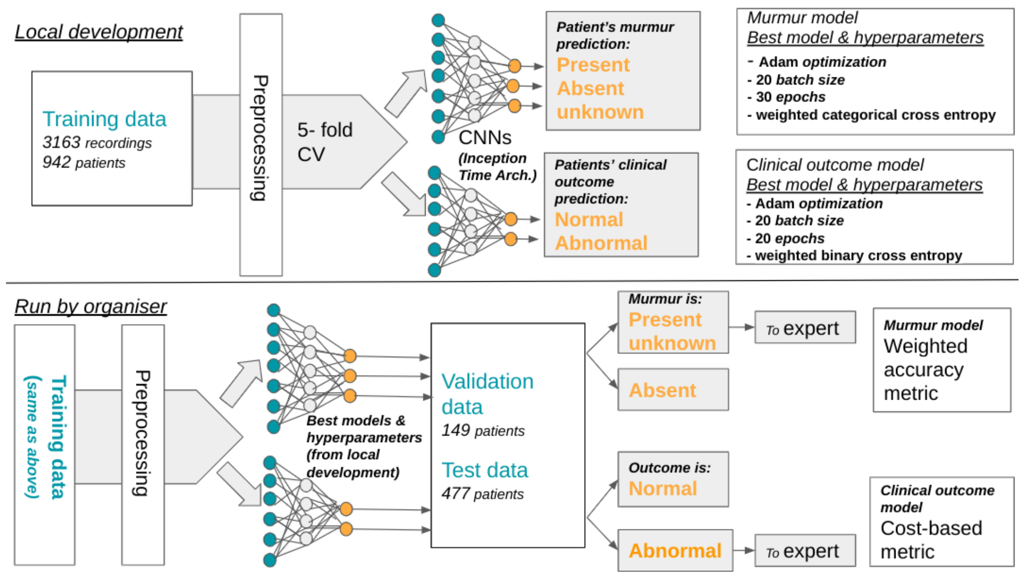

Figure 2. Detailed overview of the local development of the model and the submitted code run by the organizers.

Algorithm 2 : Murmur algorithm

Input: p = patient, r = PCG record,
t = total population, l = label
Output: p_{n_l} = patient wise labels
for n in t **do**
  if any r_l **in** p_n = Absent **then**
    p_{n_l} = Absent
  else if any r_l **in** p_n = Present **then**
    p_{n_l} = Present
  else if any r_l **in** p_n = Unknown **then**
    p_{n_l} = Unknown
  end if
end for

Algorithm 3 : Outcome algorithm

Input: p = patient, r = PCG record,
t = total population, l = label
Output: p_{n_l} = patient wise labels
for n in t **do**
  if any r_l **in** p_n = Abnormal **then**
    p_{n_l} = Abnormal
  else if any r_l **in** p_n = Normal **then**
    p_{n_l} = Normal
  end if
end for

2.6. Submitted model

The best models and hyper-parameters found during local development were used to train the final models by submitting our code to the organizers using a Docker image. The models were trained on the whole training set and then applied to the hidden validation set.

3. Results

The murmur classifier was trained for 30 epochs while the clinical outcome classifier was trained for 20 epochs. Both models were trained using a batch size of 20, and an Adam optimizer starting at a learning rate $= 0.001$.

Table 1 shows the cross-validated results on the training data set in terms of weighted accuracy, challenge cost, accuracy and F-measure, while the results on the validation set is only given in terms of weighted accuracy for the murmur model and challenge cost for the clinical outcome model.

Model	Metric	Training	Validation	Test
Murmur	Weighted accuracy	0.497 ± 0.083	0.585	-
	Challenge metric	13158 ± 1283	-	-
	Accuracy	0.446 ± 0.070	-	-
	F measure	0.403 ± 0.055	-	-
Outcome	Weighted accuracy	0.713 ± 0.042	-	-
	Challenge metric	12315 ± 903	8720	-
	Accuracy	0.51 ± 0.047	-	-
	F measure	0.465 ± 0.061	-	-

Table 1. Scores obtained by the murmur and clinical outcome classifier on the training set (5-fold cross-validation) and the hidden validation and test set.

4. Discussion and conclusion

The results achieved on the hidden validation set were significantly better than the CV results on the training set. The clinical outcome classifier even outperformed all other

George B Moody Challenge 2022 competitors' classifiers with a better challenge cost score on the hidden validation set. However, the surprisingly good results may be achieved by coincidence, but we got 5 scores within the 16 best submissions on the clinical outcome ranking with a mean challenge cost score of 9079 ± 254. The difference in performance on the training set compared to the validation set may be due to a different class distribution in the two data sets, but this is just speculation, since the validation set is hidden from the participants of the challenge.

Pre-training of the models were also tested using the 2016 PhysioNet Challenge data [15, 16]. We tried different approaches to continue the training of the pre-trained models, like freezing all layers except the last one, training for few/many epochs and high/low learning rate. However, there were no significant improvements during CV on the training set and the performance on the validation set was lower compared to no pre-training.

Both murmur and the clinical outcome classifiers were trained using single PCG recordings, and the auscultation location was not taken into consideration. However, in the preliminary phase of the challenge, we also tested multi-channel PCG classifiers, but they were outperformed by the single-channel classifiers. This observation taken into account in addition to the performance of our classifier compared to the other challenge participants' classifiers on the validation set, supports the hypothesis that a CNN can detect abnormalities from PCG recordings without knowing the auscultation location. This finding might have implications for the further development of CNN-based PCG classifiers. However, further studies are needed to provide an in-depth explanation of how these CNNs interpret the PCGs. A greater focus on the explainability of these models could produce interesting findings that could be of clinical relevance.

References

[1] WHO Cardiovascular diseases (CVDs). URL https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed 21.08.2022.

[2] Tavel ME. Cardiac Auscultation. Circulation March 2006;113(9):1255–1259, DOI: https://doi.org/10.1161/CIRCULATIONAHA.105.591149.

[3] Montinari MR, Minelli S. The first 200 years of cardiac auscultation and future perspectives. Journal of Multidisciplinary Healthcare March 2019;12:183–189, DOI: https://doi.org/10.2147/JMDH.S193904.

[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. preprint, Health Informatics, August 2022. DOI: https://doi.org/10.1101/2022.08.11.22278688.

[5] Vukanovic-Criley JM, Criley S, Warde CM, Boker JR, Guevara-Matheus L, Churchill WH, et al. Competency in cardiac examination skills in medical students, trainees, physicians, and faculty: a multicenter study. Archives of Internal Medicine March 2006;166(6):610–616, DOI: https://doi.org/10.1001/archinte.166.6.610.

[6] Roy D, Sargeant J, Gray J, Hoyt B, Allen M, Fleming M. Helping family physicians improve their cardiac auscultation skills with an interactive CD-ROM. The Journal of Continuing Education in the Health Professions 2002;22(3):152–159, DOI: https://doi.org/10.1002/chp.1340220304.

[7] Barrett MJ, Mackie AS, Finley JP. Cardiac Auscultation in the Modern Era: Premature Requiem or Phoenix Rising? Cardiology in Review October 2017;25(5):205–210, DOI: https://doi.org/10.1097/CRD.0000000000000145.

[8] Rangayyan RM, Lehner RJ. Phonocardiogram signal analysis: a review. Critical Reviews in Biomedical Engineering 1987;15(3):211–236.

[9] Emmanuel BS. A review of signal processing techniques for heart sound analysis in clinical diagnosis. Journal of Medical Engineering Technology August 2012;36(6):303–307, DOI: https://doi.org/10.3109/03091902.2012.684831.

[10] Debbal SM, Bereksi-Reguig F. Computerized heart sounds analysis. Computers in Biology and Medicine February 2008;38(2):263–280, DOI: https://doi.org/10.1016/j.compbiomed.2007.09.006.

[11] Chen W, Sun Q, Chen X, Xie G, Wu H, Xu C. Deep Learning Methods for Heart Sounds Classification: A Systematic Review. Entropy Basel Switzerland May 2021;23(6):667, DOI: https://doi.org/10.3390/e23060667.

[12] Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation June 2000;101(23), DOI: https://doi.org/10.1161/01.CIR.101.23.e215.

[13] 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 June 2022;26(6):2524–2535, DOI: https://doi.org/10.1109/JBHI.2021.3137048.

[14] Ismail Fawaz et al. H. InceptionTime: Finding AlexNet for time series classification. Data Mining and Knowledge Discovery November 2020;34(6):1936–1962, DOI: https://doi.org/10.1007/s10618-020-00710-y.

[15] Clifford G, Liu C, Springer D, Moody B, Li Q, Abad R, et al. Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016. September 2016; URL http://www.cinc.org/archives/2016/pdf/179-154.pdf.

[16] Clifford GD, Liu C, Moody B, Millet J, Schmidt S, Li Q, et al. Recent advances in heart sound analysis. Physiological Measurement August 2017;38(8):E10, DOI: https://doi.org/10.1088/1361-6579/aa7ec8.

Address for correspondence:

Bjørn-Jostein Singstad
Kristian Augustus Gate 23, 0164 Oslo, Norway
bjornjs@simula.no