Multimodal Deep Learning for Synchronous Heart Sounds and Electrocardiogram Classification

Bruno Oliveira[1]
up202108986@edu.fc.up.pt

Miguel Coimbra[1]
mcoimbra@fc.up.pt

Francesco Renna[1]
francesco.renna@fc.up.pt

[1]INESC-TEC and FCUP,
University of Porto, Portugal

Abstract

We propose a multimodal model for the binary classification (normal/abnormal) of synchronous heart sounds and electrocardiogram (ECG). The preliminary results shows that there is an improvement in using both signals for classification instead of heart sounds or ECG alone (*e.g.*, F1-score of 0.86 vs 0.79 and 0.84 respectively), which is useful for the detection of heart abnormalities in low-income countries, with the help of a multimodal stethoscope.

1 Introduction

Cardiovascular diseases (CVD) are the leading cause of dead, accounting for 32% of all deaths worldwide in 2019. More than 75% of global deaths caused by CVDs are from countries with lower- and middle-income levels, since primary healthcare programs, services and dedicated equipment are not readily available [1]. Methods like Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and echocardiography offer high-resolution images and thorough analysis of both the heart's physiological function and its structure [2]. Nonetheless, their usage is limited due to their considerable expense, the need for sophisticated equipment and specialized staff. As a result, these techniques are not commonly employed as initial screening approaches. On the other hand, observing the heart's sounds, or phonocardiogram (PCG), through cardiac auscultation and analysing the electrocardiogram (ECG) remains the prevalent approaches for initial screening. Simultaneously recording and analysing both PCG and ECG signals during routine auscultations provides a rapid assessment of the heart's condition. This approach enhances screening accuracy by leveraging the complementary information offered by these two signals [3].

Sophisticated signal processing and machine learning methods have effectively been applied to automatically identify diseases using both signals. Nonetheless, most studies still prioritize harnessing the PCG and ECG separately [4]. Commonly, the PCG is utilized for initial stethoscope-assisted examinations, while the ECG is reserved for more intricate diagnostic evaluations [5]. This work's objective is to successfully use deep learning approaches to classify simultaneously recorded ECG and PCG in normal/abnormal, with further algorithm application in a multimodal stethoscope to overcome the drawbacks mentioned before, reducing costs and the need of specialized staffs (easier to train/explain the signal than the images).

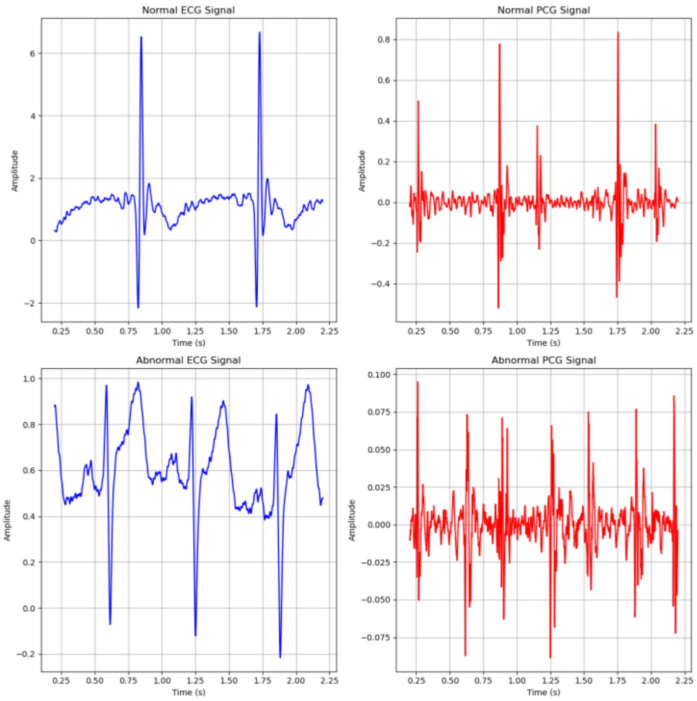

Figure 1: Comparison between synchronous normal ECG and PCG signals with synchronous abnormal ECG and PCG signals. Data taken from Physionet/CinC Challenge 2016 training set "A".

2 Methodology

The primary focus of this study revolves around the simultaneous analysis of PCG and ECG signals, comparing the preliminary models of ECG and PCG alone with the joint multimodal architecture that arises from the combination of both preliminary models. As of now, the datasets publicly available for both signal joint analyses continue to be extremely scarce. In 2016 Physionet released one of them, a dataset centred on CVDs, in which the training set "A" encompasses both synchronously ECG and PCG [6]. The dataset comprises a total of 405 signals, all with sample frequency of 2000 Hz, consisting of 113 normal signals (approximately 28%) and 292 abnormal signals (approximately 72%), thus the dataset is extremely unbalanced. Figure 1 shows a data example from this dataset.

2.1 Preprocessing

Before feeding the networks with ECG an PCG signals, preprocessing was required. Each PCG signal was first divided into non-overlapping 2 seconds segments. Afterwards, since the dominant frequency range of the PCG signal is concentrated below 300 Hz, a Buterworth bandpass filter of order 4 was implemented, with a high pass filter with a cutoff frequency of 25 Hz and a low pass filter with a cutoff frequency of 400 Hz, followed by a spike elimination procedure to remove undesired noise [7]. Spectral features of PCG were extracted using static, delta and delta-delta Mel-frequency cepstral coefficients (MFCCs). This cepstral transformation enables the recognition of cyclic patterns in the audio and the differentiation of signal components convoluted in the frequency domain [8]. A window length and a hop length of 128 and 64 milliseconds, respectively, was employed.

Regarding the ECG signal, 15 ECGs from the dataset have missing data points, so linear interpolation was applied to solve this issue. Subsequently, a Butterworth bandpass filter of order 4 was employed, with a low pass filter et at a cutoff frequency of 20 Hz since the relevant information within the ECG signal is predominantly concentrated at frequencies below 20 Hz [9]. Following that, the signal was down sampled to 500 Hz and then normalized between 0 and 1 to improve neural network training efficiency. Lastly, the signal was partitioned into non-overlapping 2 seconds segments.

2.2 Models

For the scope of this work, 3 neural networks where considered: a 2D-CNN for PCG, a 1D-CNN for ECG, and a hybrid neural network, encompassing both the two previously mentioned architectures. In what concerns the model for PCG, it consists of two 2D convolution layers, each of them followed by a max pooling layer (dimension reduction) and a dropout layer to avoid overfitting. There is also a batch normalization after the first convolution layer. It ends with a dense layer, another batch normalization layer, and the output layer. Other network' characteristics are described in Table 1.

The ECG model is comprised of three 1D convolution layers, all of them succeeded by a batch normalization layer and a max pooling layer. Two dense layers and the output layer wraps up the model' layout. Table 2 outlines other network' characteristics. Moreover, the hybrid neural network takes as input both the ECG and PCG segment. Each of them is processed by a neural network as the ones described before (2D-CNN for PCG and a 1D-CNN for ECG). Then the output from the convolution layers (after the flatten layer before the final dense layers) of each network are concatenated in a single feature vector which is then followed by two dense layers and an output layer. The parameters are the same as the networks described above, such as the use of ReLU as the activation function.

Table 1: PCG network parameters.

Convolution kernel 1	*5X5, stride 1X1*
Convolution kernel 2	*3X3 stride 1X1*
Max pooling kernel	*2X2 stride 1X1*
Optimizer	*Adam*
Learning rate	*$1X10^{-3}$*
Batch size	*32*
Max number of epochs	*300*
Loss function	*Binary Cross entropy*
Activation function	*ReLU*

Table 2: ECG network parameters.

Convolution kernel	*6*
Max pooling kernel 1	*3, stride 2*
Max pooling kernel 2	*2, stride 2*
Optimizer	*Adam*
Learning rate	*$1X10^{-3}$*
Batch size	*32*
Max number of epochs	*300*
Loss function	*Binary Cross Entropy*
Activation function	*ReLU*

2.3 Training and evaluation

The models' performance assessment involved employing a 5-fold cross validation approach. The validation set was created by randomly splitting the training data within each fold into 20% for validation and 80% for training. The data partitioning was structured in order to prevent recordings from the same patient from being present in two distinct folds. Additionally, stratification was applied to ensure an even distribution of classes across the various folds. Considering the class imbalance on this dataset, it was incorporated a weight adjustment into the loss function to increase the model's focus on the underrepresented class (normal), applying the following formula:

$$\text{Weight Positives/Negatives} = (1 \div \sum\text{Positives/Negatives}) \times (\sum\text{Total instances} \div 2)$$

Since each patient is represented by a set of 2 seconds instances, the strategy involved training the model on individual segments and then obtaining a patient level label by averaging.

For the evaluation, the accuracy, sensitivity, specificity, precision and F1-score were computed for each fold and then averaged for all folds for each of the three proposed models.

3 Results and Discussion

In what concerns the obtained results (Figure 2), it is observed that the multimodal model (ECG + PCG) has overall better results, showcasing a better accuracy and F1-score (overall mean of 0.86 for the multimodal network vs 0.79 for the PCG network and 0.84 for ECG network). This indicates that the proposed model has a better balance between the false positives and the false negatives, which is important in the medical field. It is interesting to note that the ECG model has the best sensitivity (true positive rate), which means that it is better at correctly identifying individuals who have the anomaly (e.g., a cardiac abnormality) as positive cases. On the other hand, the PCG model has the best specificity score (true negative rates), implying a superior performance at correctly identifying individuals who do not have the medical condition or anomaly. In other words, it is less likely to classify individuals without the condition as positive cases, reducing the chances of false alarms or unnecessary interventions. In addiction the PCG model does also have a slightly better precision.

Regarding the intra variability of each model, it is perceived that the PCG model has more outliers and variability when compared to the other 2 models. This could mean that the model performance is less stable, suggesting that there are cases that the model struggles to handle. One solution to this problem could be investigating the cases where the model struggles or fail to perform and see if the same cases are correctly classified in the other two models.

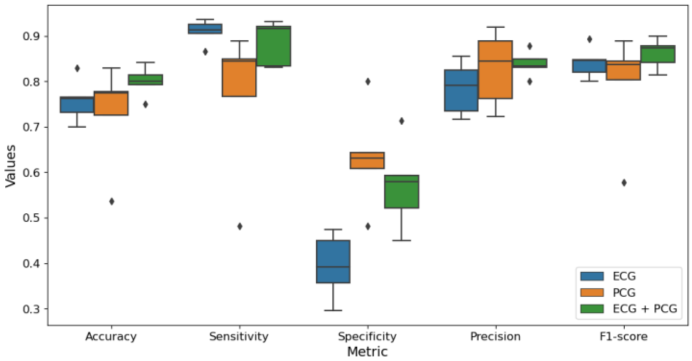

Figure 2: Boxplots illustrating the metrics calculated for each of the five folds during cross-validation, across the three distinct models.

4 Conclusions

This work shows that the ECG + PCG model deals better with the binary classification task, improving on using ECG or PCG alone for initial heart condition' screening.

Future work could pass by improving the model's performance, using explainable AI techniques (XAI) to see why the models fail to classify some instances.

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020. This work is also co-financed by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project HfPT, with reference 41.

5 References

[1] 'Cardiovascular diseases (CVDs)'. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed Sep. 05, 2023).

[2] Q. Counseller and Y. Aboelkassem, 'Recent technologies in cardiac imaging', *Front. Med. Technol.*, vol. 4, p. 984492, Jan. 2023, doi: 10.3389/fmedt.2022.984492.

[3] X. Bao, Y. Deng, N. Gall, and E. Kamavuako, 'Analysis of ECG and PCG Time Delay around Auscultation Sites', Jan. 2020, pp. 206–213. doi: 10.5220/0008942602060213.

[4] P. Li, Y. Hu, and Z.-P. Liu, 'Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods', *Biomed. Signal Process. Control*, vol. 66, p. 102474, Apr. 2021, doi: 10.1016/j.bspc.2021.102474.

[5] R. Hettiarachchi *et al.*, 'A Novel Transfer Learning-Based Approach for Screening Pre-Existing Heart Diseases Using Synchronized ECG Signals and Heart Sounds', in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2021, pp. 1–5. doi: 10.1109/ISCAS51556.2021.9401093.

[6] G. D. Clifford *et al.*, 'Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016', in *2016 Computing in Cardiology Conference (CinC)*, Sep. 2016, pp. 609–612.

[7] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, 'Segmentation of heart sound recordings by a duration-dependent hidden Markov model', *Physiol. Meas.*, vol. 31, no. 4, p. 513, Mar. 2010, doi: 10.1088/0967-3334/31/4/004.

[8] K. Rao and M. k e, *Speech Recognition Using Articulatory and Excitation Source Features*. 2017. doi: 10.1007/978-3-319-49220-9.

[9] J. Li, L. Ke, Q. Du, X. Ding, and X. Chen, 'Research on the Classification of ECG and PCG Signals Based on BiLSTM-GoogLeNet-DS', *Appl. Sci.*, vol. 12, no. 22, Art. no. 22, Jan. 2022, doi: 10.3390/app122211762.