Transfer Learning in Heart Sound Classification using Mel Spectrogram

Xin Li[1], G. Andre Ng[1,2], Fernando S. Schlindwein[1,2]

[1]University of Leicester, Leicester, United Kingdom
[2]National Institute for Health Research, Leicester Cardiovascular Biomedical Research Centre,
Glenfield Hospital, Leicester, United Kingdom

Abstract

Congenital heart illnesses impact roughly 1% of newborns, and they are a significant cause of morbidity and mortality in a variety of serious situations, including progressive heart failure. Phonocardiogram (PCG) studies can reveal crucial clinical information about heart malfunction caused by congenital and acquired heart disease. One of the 23th PhysioNet/Computing in Cardiology Challenge 2022 tasks is to develop computer tools for detecting the presence or absence of murmurs from multiple heart sound recordings from multiple auscultation locations.

Mel spectrograms were generated from up to 30 seconds per recording and reshaped at input of pre-trained AlexNet. The last three layers of AlexNet were modified to suit the task as multilabel classification. The database was split into 80% for training and 20% for validation. The database appeared imbalanced, so the class with small number of data entries was oversampled proportionally before training. The prepossessing and classifier were implemented in Matlab R2022a. For subjects with more than one recording, outputs of the classifier for each recoding were then integrated with customised possibility thresholds optimised to the challenge scoring system.

Team Leicester Fox's best entry in the official phase achieved challenge scores of 0.502 for murmur detection and 13825 for outcome prediction. Transfer learning and neural networks approaches showed potential application for murmurs detection using PCG.

1. Introduction

Congenital heart illnesses impact roughly 1% of newborns, and they are a significant cause of morbidity and mortality in a variety of serious situations, including progressive heart failure [1]. Congenital cardiac disorders are predicted to impact about 0.5 million children in East Africa alone [2], with around 0.8% of the births affected [3]. Diagnosis and treatment of congenital and acquired cardiac problems in children is challenging in some developing countries, due to a lack of infrastructure and cardiac experts in wide geographical regions, as well as difficulties in accessing health care. Furthermore, the present COVID-19 epidemic complicates clinical evaluation of patients by delaying critical in-person patient-doctor meetings, which has a detrimental influence on screening and monitoring efforts. A non-invasive examination of the mechanical function of the heart conducted at the point of care can offer early information about congenital and acquired cardiac problems in infants.

Phonocardiogram (PCG), as a non-invasive tool, can reveal crucial clinical information about heart malfunction caused by congenital and acquired heart disease [4]. This is accomplished by detecting aberrant sound waves in the PCG signal, often known as heart murmurs. Murmurs are irregular waves caused by turbulent blood flow in cardiac and vascular tissues. They are linked to particular disorders such as septal abnormalities, ductus arteriosus failure in infants, and faulty cardiac valves. However, abnormities in PCGs are usually detected by experienced clinicians with special training on stethoscopes. There has been relatively little research on the automated identification of pertinent clinical information and diagnosis using PCGs. As the PCG is an audio signal, which is inherently one-dimensional (amplitude over time). Clinical decisions are made on hearing the audio by humans. We propose using the Mel spectrogram as data input in this work, a transformation that reveals the frequency content of the signal across time on a scale that is more suitable to humans, as we perceive frequency logarithmically [5]. In this work, we use transfer learning to develop computer tools based on pre-trained neural networks for detecting the presence or absence of murmurs from multiple heart sound recordings from multiple auscultation locations.

2. Methods

2.1. Database

Database consists of 942 patients from one or more prominent auscultation locations: pulmonary valve (PV), aortic valve (AV), mitral valve (MV), tricuspid valve (TV), and other. For each patient, recordings were unified labelled as subject label (three classes: present, absent, unknown).

2.2. Training Data Labelling

For each subject, all sound wave files from multiple prominent auscultation locations and their corresponding labels were identified. As each subject has a unified label for both clinical outcome and murmur detection. However, for positive class, not all the PCG recordings were necessarily labelled as positive. It is believed that some positive recordings that were labelled negative or neutral may still contain valuable information that were not discoverable for human hearing. Therefore, all PCG files associated with a positive class were labelled positive, regardless of their individual labelling. For negative and neutral classes, all PCG files were labelled as negative or neutral, which should also be aligned with their individual labelling.

Therefore, from the 942 subjects, we have generated a total of 3163 PCG recordings with the above labelling logic.

2.3. Data processing

As we proposed that heart beats were not annotated, a 30-second duration was chosen to include enough information with sufficient number of heart beats for each record. For PCG records that are longer than 30 seconds, data was truncated, whilst data was padded with zero when the PCG records are shorter. For each record, the time-frequency representation was achieved by generating Mel spectrogram, which is a spectrogram where the frequencies are converted to the Mel scale. The Mel scale provides a linear scale for the human auditory system, and is related to Hertz using the following formula [6] (eq. 2):

$$m = 2595 log_{10}(1 + \frac{f}{700}) \tag{2}$$

Where *f* represents frequency in Hertz and *m* in Mel scale. The Mel scale provides a linear scale for the human auditory system. This was achieved using the Matlab function melSpectrogram with 64 bands within the frequency range from 25 Hz to 2000 Hz. The resulting Mel Spectrogram graph was then scaled and resized as 227 x 227 images. Images were saved as a lossless format (.png) and Matlab datastores were created for quick access.

2.4. Imbalance Data

Out of the 942 subjects, 179 were positive class, 695 were negative and 68 were neutral, with each subject consisting 1-6 recordings.

The database appeared imbalanced, and the classes (positive and neutral) with small number of data entries were oversampled proportionally before training using Matlab implementation python library imbalanced-learn [7]. Let $\mathcal{X}$ be an imbalanced dataset with $\mathcal{X}_{min}$ and $\mathcal{X}_{maj}$ the subset of samples belonging to the minority and majority class, respectively. The balancing ratio of the dataset is defined as (eq. 1):

$$r_{\mathcal{X}} = \frac{|\mathcal{X}_{min}|}{|\mathcal{X}_{maj}|} \tag{1}$$

The balancing process is equivalent to resampling into a new dataset $\mathcal{X}_{res}$ such that $r_{\mathcal{X}} > r_{\mathcal{X}_{res}}$. Data balancing can be performed by oversampling such that new samples are generated in $\mathcal{X}_{min}$ reach the balancing ratio $r_{\mathcal{X}_{res}}$ [7].

2.5. Data Augmentation

For each PCG record, data was augmented by left and right shifting a small instance (**Figure 1**). Also, three levels of Gaussian noises (amplitude with 0.5-, 1- and 1.5-times standard deviation) were added to the signal to model different signal-to-noise ratio scenarios.

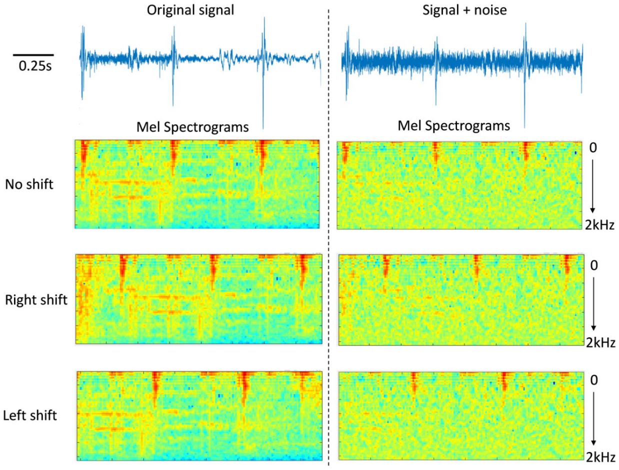

Figure 1. Left: example of an original PCG waveform, and the Mel spectrogram with time shifts; Right: example of the waveform of original PCG with added noise, and their corresponding Mel spectrogram with time shifts.

2.6. Model Architecture

Each Mel Spectrogram graph was converted to a 227 x 227 x3 image with equal weight to RGB channels. This was then fed into a pre-trained AlexNet [8], which includes eight layers with learnable parameters. Relu activation is used in each of the five levels of the model, with the exception of the output layer, which uses max pooling followed by three fully connected layers. The last a few layers were modified to suit the tasks of Murmur detection and clinical diagnostic (**Figure 2**).

2.5. Model Training

Training and classification were implemented in Matlab environment using a single GPU. Models were trained on 80 % of data as a training set, and the accuracy was evaluated on 20 % of data as a validation set. Both neural networks were trained for 30 epochs with mini-batch size of 64 samples, where each epoch was shuffled. The neural networks apply the Adam optimization method with learning rate set to 0.0001. Lost function of sparse categorical cross-entropy (eq. 3) was used with accuracy as evaluation metric. The cross-entropy function was the objective function to be optimised during the model training process as follows:

$$L(X,r) = -\frac{1}{m}\sum_{i=1}^{m} log_p(R = r_i|X) \tag{3}$$

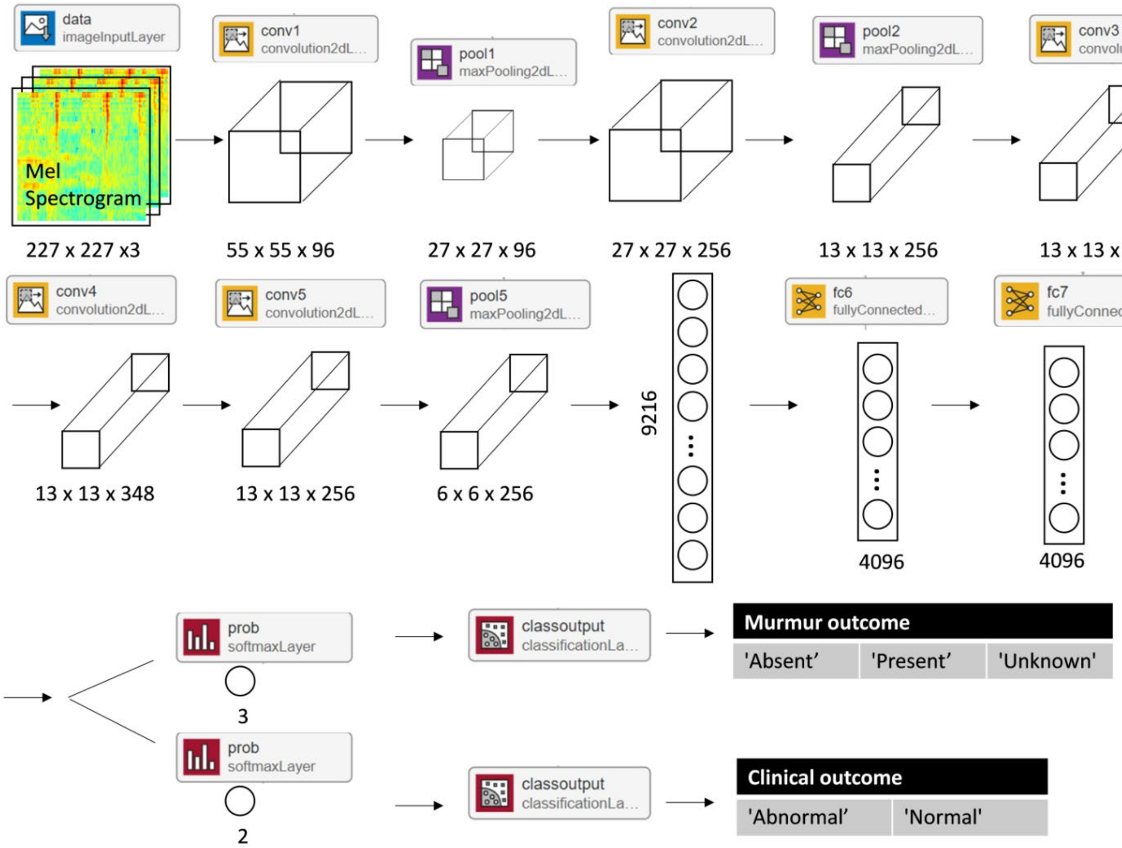

Figure 2. Graphical diagram illustrating the architecture of the modified pre-trained AlexNet.

3. Results

Our"Leicester Fox" team successfully ran all 5 entries in the unofficial phase. Our best entry for the unofficial phase of the PhysioNet/CinC 2022 competition received a Challenge Score of 539.591 on full data, ranking at 13th out of 166.

In the official phase, we have modified our unofficial model to accommodate the new task of outcome detection. Figure 3 demonstrates the training process of both murmur model and the outcome model respectively. In total, we have submitted three entries (**Table 1**) with different configurations. Our best entry in the official phase achieved challenge scores of 0.502 for murmur detection and 13825 for outcome prediction.

Table 1. Challenge Scores for official entries

Entries	Changes	Score murmur	Score outcome
1	oversample + time shift	0.478	13825
2	no oversample + time shift	0.502	15162
3	oversample(murmur) + time shift + added noise	0.367	18899

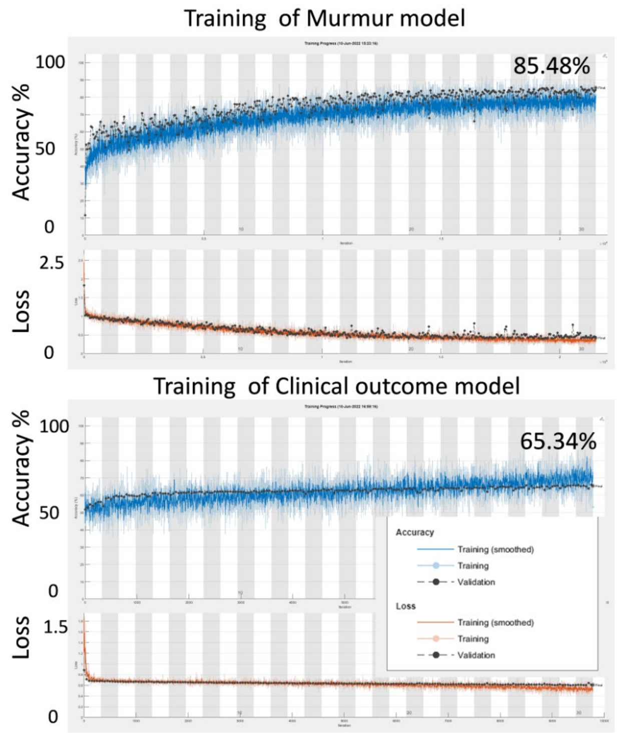

Figure 3. Training Loss and Accuracy per epoch on training and validation set. Top: Murmur model; Bottom; clinical outcome model.

4. Discussion and Conclusions

Our first entry was using oversampling mentioned in Section 2.4, and time shift for data augmentation. This entry achieved the best outcome score (the lower the better). However, without oversample in entry 2, the murmur score achieved the best result, which may be due to a similar class distribution in unseen testing dataset withhold by the organizers. Data augmentation by adding additional noise at different levels worsened the final score, which may suggest unreal noise modelling. Transfer learning and neural networks approaches showed potential application for murmurs detection using PCG. Future work is required to improve the model accuracies.

Acknowledgments

This work was funded by the British Heart Foundation (PG/18/33/33780, Grant AA/18/3/34220), Medical Research Council UK (MR/S037306/1) and the National Institute for Health Research, Leicester Cardiovascular Biomedical Research Centre.

References

[1] Burstein DS, Shamszad P, Dai D, Almond CS, Price JF, Lin KY, et al. Significant mortality, morbidity and resource utilization associated with advanced heart failure in congenital heart disease in children and young adults. Am Heart J. 2019;209:9-19.
[2] Jivanji SGM, Lubega S, Reel B, Qureshi SA. Congenital Heart Disease in East Africa. Front Pediatr. 2019;7:250.
[3] Zuhlke L, Mirabel M, Marijon E. Congenital heart disease and rheumatic heart disease in Africa: recent advances and current priorities. Heart. 2013;99:1554-61.
[4] Singh J, Anand RS. Computer aided analysis of phonocardiogram. J Med Eng Technol. 2007;31:319-23.
[5] Stevens SS, Volkmann J, Newman EB. A Scale for the Measurement of the Psychological Magnitude Pitch. The Journal of the Acoustical Society of America. 1937;8:185-90.
[6] O'Shaughnessy D. Speech communication: human and machine: Addison-Wesley; 1987.
[7] Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res. 2017;18:559–63.
[8] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ, editors. Advances in Neural Information Processing Systems: Curran Associates, Inc.; 2012.

Address for correspondence:

Xin Li
University of Leicester, Leicester, UK
xin.li@leicester.ac.uk