Hierarchical Multi-Scale Convolutional Network for Murmurs Detection on PCG Signals

Yujia Xu*, Xinqi Bao*, Hak-Keung Lam, Ernest N. Kamavuako

Department of Engineering, King's College London, London

Abstract

Computer-aided analysis is of great help in improving heart sound classification. PhysioNet Challenge 2022 *provides a platform for researchers to test and compare their proposed classification algorithms. In the Challenge, our team (HearTech) proposed a recording quality assessment method based on frequency density distribution for label correction to prevent the poor-quality recording segments from misleading network optimisation. Besides, a hierarchical multi-scale convolutional neural network (HMS-Net) was designed to conduct both the murmur (T1) and clinical outcome (T2) classification. HMS-Net extracts convolutional features from the spectrograms on multiple scales and fuses them through its hierarchical architecture. The network builds long short-term independencies between multi-scale features and improves the classification performance. Finally, the prediction of a patient is based on the ensembled segment predictions by sliding window. In the five-fold cross-validation by patients, the proposed algorithm performed an average weighted accuracy of* 0.81 *(best* 0.853*) on T1 and an average challenge score of* 9808 *(best* 9242*) on T2. In the Challenge hidden validation set, the proposed algorithm achieved* 0.806 *weighted accuracy on T1 and* 9120 *challenge score on T2, ranking* 1^{st} *and* 4^{th} *out of* 305 *entries, respectively.*

1. Introduction

Early screening is vital in detecting cardiovascular disease (CVD) and necessary action to reduce the risk of worsening heart disease. The initial suspicion often depends on the medical staff to listen to murmurs in the heart sound (recorded as phonocardiogram, PCG) during auscultation. However, due to the limitation of listening ability and clinical experience, auscultation is not always trustworthy [1]. Therefore, a more robust and accurate computer-aided PCG analysis algorithm is greatly needed to improve the situation.

The existing PCG classification methods can be divided into two types: (1) feature-based machine learning (ML) methods and (2) deep learning (DL) based methods. Feature-based ML requires manual extraction of the features, which heavily depends on PCG segmentation and feature settings. This usually causes robustness and portability issues. The inputs can be raw signals or their time-frequency distributions (TFDs) for DL-based approaches. Deep CNNs can extract the spatial features automatically, generally skip the segmentation and require fewer input settings. However, DL approaches require large datasets to improve classification performance. In recent years, large PCG datasets such as [2] and [3] have made DL approaches more competitive.

In the previous study [4], the 2-D TFDs as inputs for PCG classification were proved to outperform the raw signals on deep CNNs. Furthermore, the current mainstream CNNs were designed for the image recognition field with the local attention characteristic [5]. The receptive field of each CNN layer is fixed without considering the long short-term dependencies of the time-domain signal information. Therefore, the primary aim of this study is to design a novel CNN with hierarchical multi-scale architecture to improve the classification performance by fusing multi-scale features. In addition, the low-quality recording segments involving artefacts may mislead the network optimisation. Hence, the second aim of this study is to improve the classification accuracy by designing a quality assessment method to correct the labels for the low-quality segments. The proposed algorithm has been applied in the PhysioNet Challenge to verify the performance.

2. Methodology

2.1. Database and Pre-processing

The database used in this study is the PhysioNet Challenge 2022 publicly released data, containing 3163 PCG recordings from 942 patients (Murmurs: 695 absent, 68 unknown and 179 present; Outcomes: 456 abnormal and 486 normal.). See [4] for more details.

In this study, the recording sampling frequency is downsampled from 4000 to 2000 Hz for faster data loading

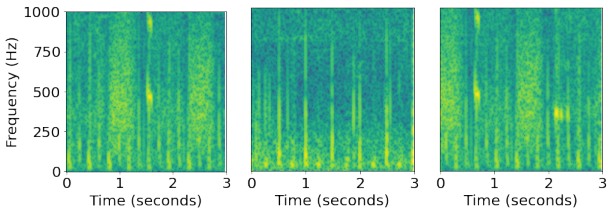
*The authors contributed equally to this work.

speed without information loss of the murmurs (ranging from 20–500 Hz [6]). Afterwards, the signal is normalised by z-score normalisation.

2.2. Quality Assessment Method

The murmur labels given in the database correspond to each auscultation location recording of the patients, which means a recording is 'absent', 'unknown', or 'present'. In [7], the PCG duration effect has been proved minor on CNN performance. Thus, in this study, the recordings are cropped into 3s segments as CNN inputs, considering both the shortest signal length 5s and CNN receptive fields.

However, the PCG recordings contain many low-quality segments caused by ambient noise, artefacts, body friction, etc., which will mislead CNN optimisation. Hence, a quality assessment for the murmur label correction of segments is needed. Since the frequency of normal heart sound is between $20-200$ Hz, the energy of most murmurs is much less than that of heart sound [6]. The selected assessment criteria is the ratio of spectral density between $20-200$ Hz to full band ($0-1000$ Hz), named quality ratio.



(a) Ratio 0.28 (b) Ratio 0.52 (c) Ratio 0.26

Figure 1: Spectrogram of a segment with a quality ratio of (a) 0.28 from signal labelled 'absent'. (b) 0.52 from the same signal. (c) 0.26 from signal labelled 'unknown'.

Fig. 1a and 1b are spectrograms of two segments from one 'absent' recording. Fig. 1c is of a segment from an 'unknown' recording. There are visible differences between 1a and 1b, especially in the higher frequency bands. On the contrary, these high-frequency noises in 1a are similar to those in 1c. After manual frequency analysis on recordings, the label correction strategy is: if the quality ratio is larger than 30%, this segment murmur label follows the recording label. Otherwise, it will be relabelled as 'unknown'. It should be noted this label correction is only for murmur labels but keep outcome labels unchanged.

2.3. Model Interpretation

The CNN inputs are the multi-scale spectrograms of 3s segments. Three scales ($\times 1.0$, $\times 0.5$, $\times 0.25$) are selected. The parameters for the spectrograms are given in Table 1. The multi-scale spectrograms provide CNNs with time-frequency features in different resolutions and reduce the spatial information loss in single spectrogram.

Table 1: Parameter settings for multi-scale spectrogram.

Scale	Nfft	Window length	Hop length
$\times 1.0$	446	200	27
$\times 0.5$	222	100	54
$\times 0.25$	110	50	108

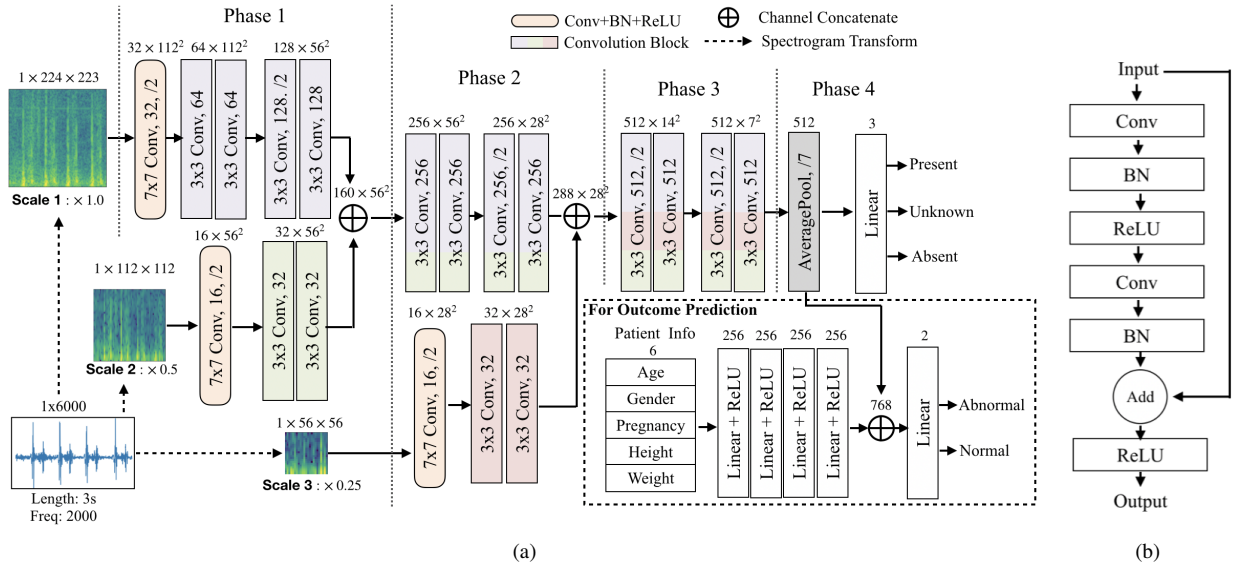
Inspired by [8,9], in this study, a hierarchical multi-scale convolutional neural network (HMS-Net) is proposed to improve the PCG classification performance by building long short-term dependencies between multi-scale inputs with its hierarchical architecture. Fig. 2a illustrates its overall structure. The fundamental element in HMS-Net, convolutional block, refers to ResNet [10], with its structure diagram shown in Fig. 2b. Three-scale spectrograms of a segment input HMS-Net and output the 3-class murmur prediction. For outcomes prediction, it combines with patient information and outputs the binary result.

In HMS-Net, the convolutional features of the multi-scale spectrograms are extracted at different depths. A larger scale requires deeper layers; thus, HMS-Net has four phases containing layers with incremental depths for extracting features from different scales. For example, in Phase 1, two sub-networks are employed to convolve the features from Scale 1 and Scale 2. The 2-scale features are then concatenated in channel dimension and passed to the next phase. Phase 4 summarises the multi-scale convolutional features by global average pooling and classifies the segment with a linear layer. Overall, HMS-Net extracts features from multiple scales separately at the beginning and fuse these features with its hierarchical design.

Regarding the outcome classifier, patient information, including age, gender (one-hot), pregnant status, height, and weight, is added as extra information to distinguish patients with abnormal clinical outcomes. As shown in Fig. 2a, 256 patient features are extracted from these information via a 4-layer multi-layer perceptron (MLP). The final outcome prediction is obtained from both the convolutional features and the patient features.

2.4. Training Settings

The optimiser is Adamw and the max training epoch is set to 100. The initial learning rate is 10^{-3}. When the training loss has stopped decreasing for five epochs, the learning rate is multiplied by 0.1. The loss function is cross-entropy with 0.1 label smoothing. In each batch, 128 multi-scale spectrograms are fed to HMS-Net.



(a) (b)

Figure 2: (a) Overall structure of the HMS-Net. The in-box text denotes the layer parameters, e.g., '3 × 3 Conv, 128, /2' represents a convolution layer with 128 3 × 3 kernels and stride 2. The text above boxes denotes the output size. The colours in convolution blocks indicate the information from certain scales. (b) Structure of a residual convolution block.

2.5. Murmur Classification

Since the HMS-Net is designed to classify the segments, for recording classification, a sliding window with 3s width and 1s step is applied to classify the whole recording continuously. For a frame slided by multiple windows, its label is calculated by the averaged distribution probabilities of the passed windows. The prediction for the recording is the serial labels per second. If the predicted 'unknown' accounts for over 80% of the serial labels, the recording is catalogued as 'unknown'. Otherwise, the recording is classified with the majority of serial labels (exclude 'unknown').

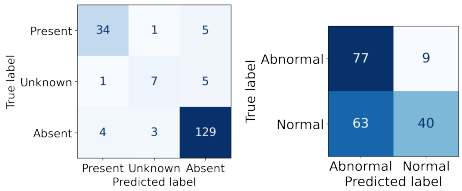
For patient classification, if one location recording is classified as 'present', the patient is labelled 'present'. In terms of 'absent' and 'unknown', the patients are classified by the majority of location recording labels. When there is the same number of 'absent' and 'unknown' recordings of the patient, 'absent' has the priority. All the mentioned thresholds are chosen based on local testing.

2.6. Outcome Classification

Our outcome prediction strategy is similar to murmur prediction but does not involve 'unknown' issue. The serial labels ('abnormal' or 'normal') per second are obtained by sliding window as well. If over $1/3$ frames are predicted as 'abnormal', the recording is predicted as 'abnormal'. When a patient has at least one predicted 'abnormal' recording, our strategy diagnoses the patient as 'abnormal'. Otherwise, the patient is diagnosed as 'normal'.

3. Results

We used five-fold cross-validation by patients to fairly evaluate our methods. See [11] for the scoring metrics of murmur weighted accuracy and outcome challenge score.



(a) (b)

Figure 3: Confusion matrices of (a) murmur classification (b) outcome classification.

Our method achieved an average murmur classification accuracy of 91.37% (best 92.85%) on segments in the five-fold cross-validation. It performed 83.78% averaged murmur classification accuracy on patients and 0.81 averaged weighted score. Fig. 3a shows the confusion matrix of the best fold on patient classification. The overall accuracy was 89.94%, respectively, on 'present' was 85.0%, 'unknown' was 53.84% and 'absent' was 94.85%. The weighted murmur accuracy was 0.853. Regarding patient outcome classification, our method achieved averaged 56.83% accuracy (best 62.96%), 84.33% averaged sen-

sitivity (best 90.59%), and 9808 averaged outcome (best 9242). The outcome confusion matrix for highest outcome score is shown in Fig. 3b. In the blind validation set, the algorithm achieved 0.806 murmur weighted accuracy and 9120 outcome challenge score.

4. Discussion and Conclusion

The following discussions focus on our experiences during the challenge and the vision for future work.

Label Correction The low-quality ('unknown') segments or recordings caused by artefacts exist considerably and are often fused with heart sounds and murmurs. When training the CNN model on segment inputs, these low-quality parts will greatly mislead the model. It is necessary to identify them while often being neglected. Therefore, a quality assessment method by spectral density was proposed to alleviate the label inconsistency problem. With the assessment method, the accuracy on segments increased by approximately $5-6\%$. Though, in the current method design, it did not involve too many criteria considering the data loading speed for the DL methods. There is huge room to extend the assessment criteria by the PCG time-domain or frequency-domain features to achieve better label correction and improve the classification accuracy. In future work, removing the unknown segments in data pre-processing or alleviating its effect in the testing interface should be studied.

HMS-Net HMS-Net holds the advantage of combining the features from multi-scale spectrograms to improve the classification performance. However, work could still be done on determining the optimal network depth and width, parameter optimisation, etc., to make the network more efficient. Furthermore, the low-quality segments issue made it hard to objectively evaluate its segment classification performance when many label inconsistencies occurred. This is also why the confusion matrix on segments was not provided in the results. Despite this, in local tests, HMS-Net performed approximately 1% better than ResNet34.

Outcome Prediction The clinical outcomes diagnosed by cardiologists are based on multiple assessments. Only PCG with basic patient information is far from enough to reliably and accurately identify the outcome. More patient diagnostic information like echocardiogram can be served as extra inputs to provide CNNs with more valuable information. Besides, the outcome result is quite sensitive to hyper-parameter settings.

Overall, this study proposed a hierarchical multi-scale convolutional neural network with a signal quality assessment method to classify PCG. In the PhysioNet Challenge 2022, it performed outstandingly with 0.806 murmur weighted accuracy and 9120 outcome challenge score. The proposed method may be inspiring and significant in future PCG classification design.

Acknowledgement

This study was supported in part by the King's-China Scholarship Council PhD Scholarship programme The work used the Cirrus UK National Tier-2 HPC Service at EPCC (`http://www.cirrus.ac.uk`) funded by the University of Edinburgh and EPSRC (EP/P020267/1).

References

[1] Kumar K, Thompson WR. Evaluation of cardiac auscultation skills in pediatric residents. Clinical pediatrics 2013; 52(1):66–73.

[2] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, et al. An open access database for the evaluation of heart sound algorithms. Physiological measurement 2016; 37(12):2181.

[3] 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.

[4] Bao X, Xu Y, Lam HK, Trabelsi M, Chihi I, Sidhom L, et al. Time-frequency distributions of heart sound signals: A comparative study using convolutional neural networks. arXiv preprint arXiv220803128 2022;.

[5] Wang X, Girshick R, Gupta A, He K. Non-local neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2018; 7794–7803.

[6] Choi S, Jiang Z. Cardiac sound murmurs classification with autoregressive spectral analysis and multi-support vector machine technique. Computers in biology and medicine 2010;40(1):8–20.

[7] Bao X, Xu Y, Kamavuako EN. The effect of signal duration on the classification of heart sounds: A deep learning approach. Sensors 2022;22(6):2261.

[8] Tao A, Sapra K, Catanzaro B. Hierarchical multi-scale attention for semantic segmentation. arXiv preprint arXiv200510821 2020;.

[9] Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. In International conference on information processing in medical imaging. Springer, 2015; 588–599.

[10] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016; 770–778.

[11] 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;.

Address for correspondence:

Ernest N. Kamavuako
Department of Engineering, King's College London, Strand, London, WC2R 2LS, United Kingdom
ernest.kamavuako@kcl.ac.uk