Using artificial intelligence (AI) to turn a mobile smartphone into a stethoscope

香 港 大 學 THE UNIVERSITY OF HONG KONG

D2H Laboratory of Data Discovery for Health

Authorship: Xin Jiayi
Major: Applied Artificial Intelligence

Name: Xin Jiayi
University No.: 3035770525
BASc(AppliedAI)

Research Colloquium for Science UG Students 2021-22

Introduction

Motivation

According to the World Health Organization (2021), cardiovascular diseases (CVDs) are the number one cause of deaths globally. There is a need to develop an accurate, low-cost and convenient early CVD screening tool. The project related to this research is to develop such a web-based application that can identify abnormal heart sounds using the embedded microphones in consumer-grade mobile phones.

Task Description

Automatic heart sound segmentation is to precisely identify the start and the end of each stage of a heart cycle. It is an essential pre-processing step of the screening algorithm since it can separate the signal into periodic cycles as well as detect the presence of extra heart sounds (third and fourth heart sound) and heart murmurs, which helps the classification of abnormal heart sounds. The difficulty of the task is amplified in a mobile-phone heart sound collection setting due to possible ambient noises.

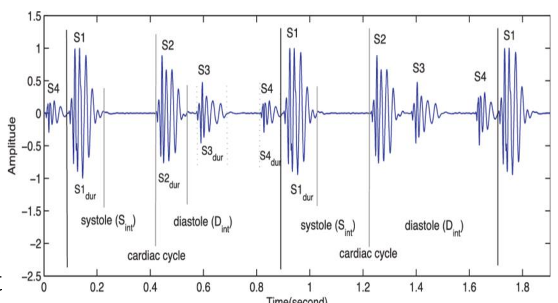

Fig 1: A phonocardiogram with extra heart sounds (Source: Varghees & Ramachandran, 2014)

Related work

Before 2018, a logistic regression hidden semi-Markov model (LR-HSMM) developed by Springer et al. (2015), was widely regarded as the state-of-the-art (SOTA) model. Recently, Messner et al., (2018) used Recurrent Neural Networks (RNNs), which not only achieve better performance but is also capable of detecting arrythmia for which the LR-HSMM failed.

Dataset

PhysioNet 2016 Challenge dataset (Clifford et al., 2016)

- 3153 recordings in total with hand-corrected ground truth labels
- Unbalanced - much more normal heart sound than abnormal one
- Recorded with stethoscopes
- Train-test-validation split:
  - Ensure that the recordings from each subject only appear in only one of training set, test set or validation set

Dataset	Challenge set	# Normal	# MVP	# Benign	# AD	# MPC	# CAD	# MR	# AS	# Pathological	Total
Test	a	116	126	114	13	23	0	0	0	0	392
	b	135	0	0	0	0	29	0	0	0	164
	e	205	0	0	0	0	3	0	0	0	208
	total	456	126	114	13	23	32	0	0	0	764
Validation	b	16	0	0	0	0	5	0	0	0	21
	c	1	0	0	0	0	0	0	2	0	3
	d	2	0	0	0	0	0	0	0	1	3
	e	105	0	0	0	0	13	0	0	0	118
	f	9	0	0	0	0	0	0	0	5	14
	total	133	0	0	0	0	18	0	2	6	159
Train	b	144	0	0	0	0	39	0	0	0	183
	c	6	0	0	0	0	0	12	6	0	24
	d	24	0	0	0	0	0	0	0	25	49
	e	1471	0	0	0	0	130	0	0	0	1601
	f	68	0	0	0	0	0	0	0	26	94
	total	1713	0	0	0	0	169	12	6	51	1951

Fig 2: Table representation of train-test-validation split

In-house Mobile Phone Collected Heart Sound Dataset (collected by Chun Yat Yee)

- Edited audio files (removing very noisy parts at the beginning and/or at the end)
- 142 recordings in total
- Recorded by the built-in microphone of a mobile phone

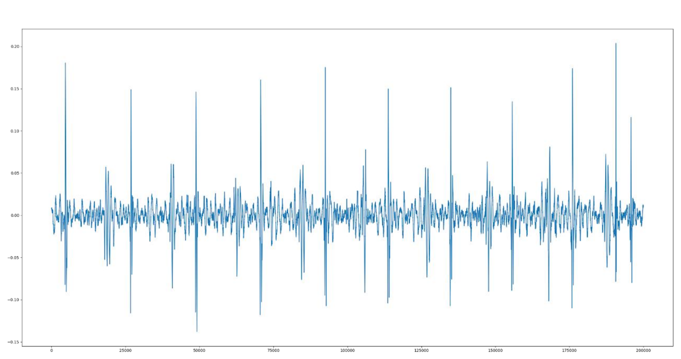

Fig 3a: Phonocardiogram of heart sounds recorded by stethoscope

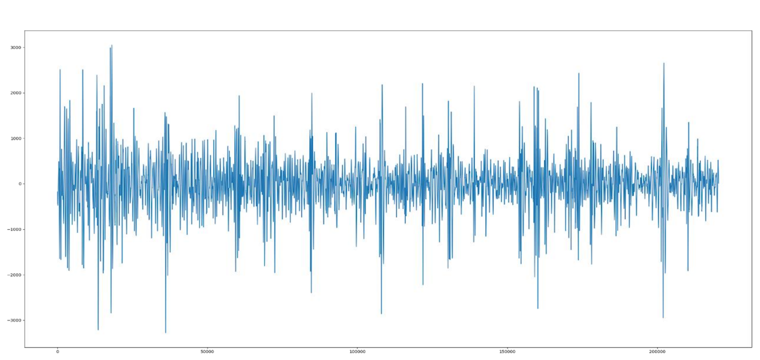

Fig 3b: Phonocardiogram of heart sounds recorded by mobile phone

Overview

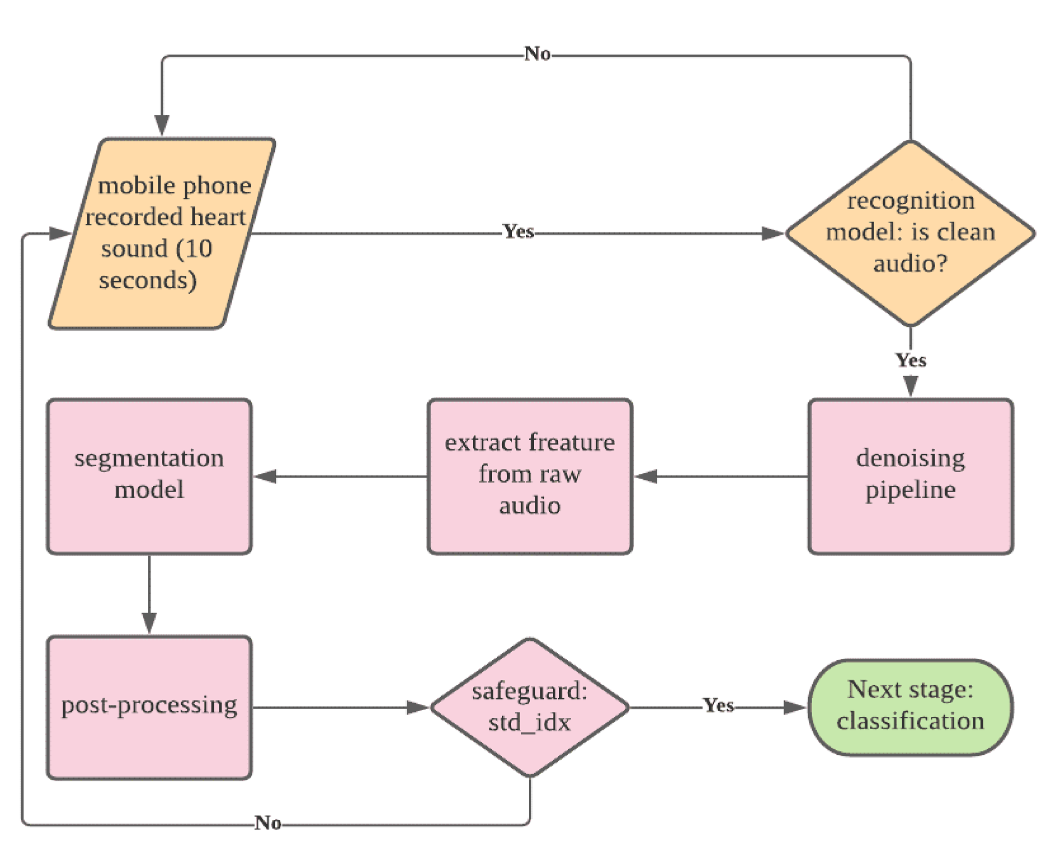

Fig 4: Overview of the CVD screening framework

Demonstration

- This website is developed by Mr. Shichao Ma with the heart sound segmentation algorithm integrated.
- This Hearty Helper program can
  - Automatically detect heart sound
  - Distinguish ambient noise from clean audio
  - Perform automatic heart sound segmentation
  - Offer real-time heartbeat rate
  - Generate a phonocardiogram

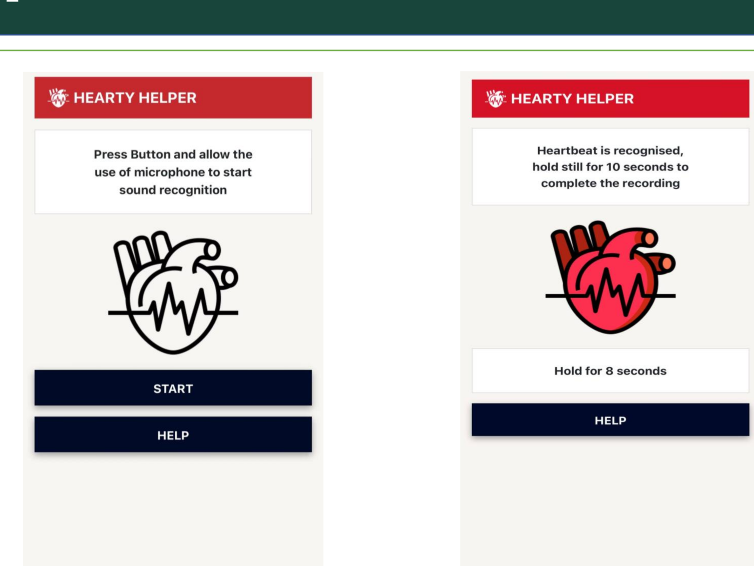

Fig 5a: (left) User interface of the website; (right) User interface when heart sound is detected

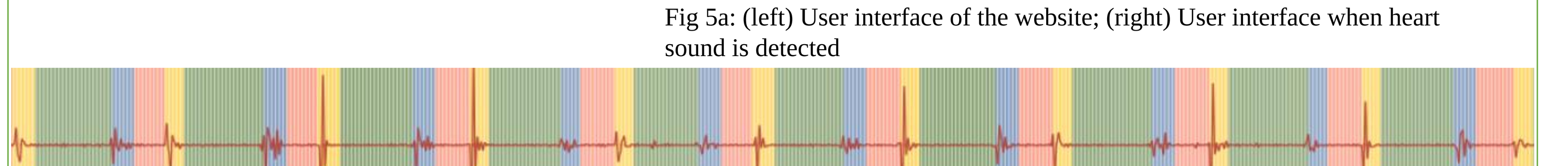

Fig 5b: Phonocardiogram generated by the website

Acknowledgement

This research is completed under the Laidlaw research scholarship. I would like to express sincere gratitude to Dr. Ho who constantly offers guidance, encouragement and support, Mr. Shichao Ma, my mentor during the research, who spares his time to discuss with me and answers my questions, and all the lab mates who kindly assist and offer feedback to this research.

Framework

Denoising Pipeline (Springer et al., 2016)

- The audio firstly passes through a bandpass filter of 50 to 600 Hz.
- A spike removal algorithm is performed to remove the rubbing noise.

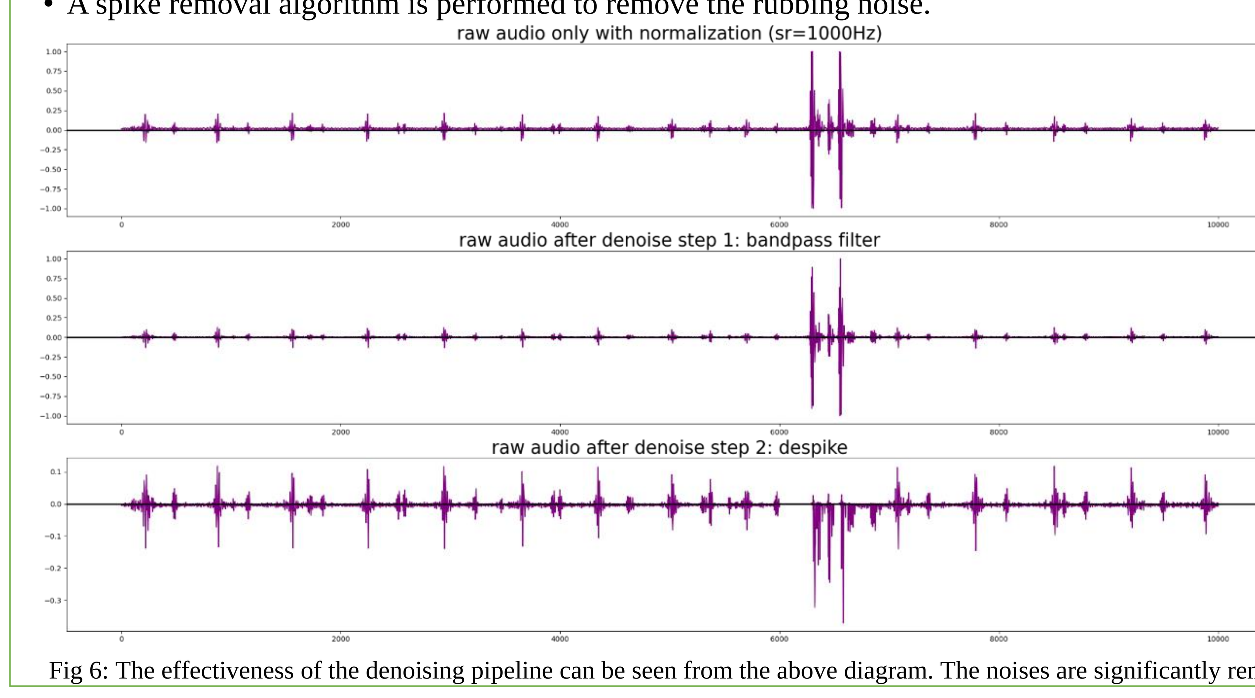

Fig 6: The effectiveness of the denoising pipeline can be seen from the above diagram. The noises are significantly removed.

Feature extraction

- We perform Short-time Fourier Transform (STFT) on the 10-second heart sound audio to obtain 41 log magnitude spectrogram features.
- We transform the spectrum to zero mean and unit variance.
- We also normalize all the features to the range of 0 to 1.

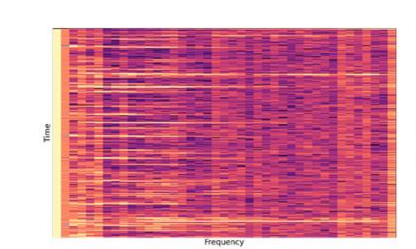

Fig 7: Periodic features can be spotted from the spectrogram.

Segmentation model

- Bidirectional Gated Recurrent Unit (BiGRU)
- 2 Bidirectional layers
- Softmax output layer
- Produce a label for each 20 milli-second frame
- Experiment on 200, 150, 100, 50, and 25 units
- Dropout 0.1

Evaluation metrics adopted from Messner et al., (2018)

- True positives (TP)
- False positives (FP)
- False negatives (FN)
- Substitutions (S)
- Insertions (I)
- Deletions (D)
- Reference states (N)

$$P_+ = \frac{TP}{TP + FP}$$

$$Se = \frac{TP}{TP + FN}$$

$$F_1 = 2 \cdot \frac{P_+ \cdot Se}{P_+ + Se}$$

$$ER = \frac{S + I + D}{N}$$

Plain BiGRU compare with HSMM

Model	P+ (%)	Se (%)	F1 (%)	ER
HSMM	**91.79**	**91.54**	**91.59**	**0.1468**
F_BiGRU_200_10	93.76	94.18	93.93	0.1208
F_BiGRU_150_10	93.18	93.89	93.48	0.1302
F_BiGRU_100_10	92.60	93.15	92.82	0.1439
F_BiGRU_50_10	91.98	93.09	92.45	0.1519
F_BiGRU_25_10	92.50	93.35	92.86	0.1423
F_BiGRU_200_8	93.46	93.78	93.59	0.1262
F_BiGRU_150_8	**94.56**	**94.62**	**94.57**	**0.1068**
F_BiGRU_100_8	94.26	94.29	94.26	0.1129
F_BiGRU_50_8	92.80	93.65	93.18	0.1361
F_BiGRU_25_8	92.88	93.66	93.22	0.1350
F_BiGRU_200_3	94.45	94.36	94.39	0.1101
F_BiGRU_150_3	94.36	94.38	94.36	0.1109
F_BiGRU_100_3	94.12	93.97	94.03	0.1173
F_BiGRU_50_3	93.13	93.36	93.23	0.1343
F_BiGRU_25_3	90.98	91.87	91.38	0.1716

Fig 8a: The best model from empirical experiment is the one with 150 units per layer and is trained with 8 second window-size. It outperforms the baseline HSMM model by 3% in F1 score.

Best Model (BiGRN_150_8) & HSMM comparison

Type	# 10-sec SW	Model	P+ (%)	Se (%)	F1 (%)	ER
Normal	5257	BiGRN_150_8	96.58	96.47	96.51	0.0685
		HSMM	94.50	94.13	94.24	0.0959
MVP	2908	BiGRN_150_8	89.72	90.04	89.85	0.1998
		HSMM	85.72	85.74	85.62	0.2593
Benign	2579	BiGRN_150_8	96.59	96.51	96.54	0.0684
		HSMM	93.90	93.66	93.71	0.1129
AD	268	BiGRN_150_8	97.45	97.39	97.41	0.0516
		HSMM	94.89	94.97	94.92	0.0937
MPC	548	BiGRN_150_8	89.90	90.97	90.36	0.1893
		HSMM	86.55	85.97	86.15	0.2238
Overall	11560	BiGRN_150_8	94.56	94.62	94.57	0.1068
		HSMM	91.79	91.54	91.59	0.1468

Fig 8b: The best model not only obtain higher overall F1 score, but also outperform the baseline in every category of heart sound. Normal – Normal heart sound. MVP - Mitral valve prolapse. Benign - Benign heart sound. AD - Aortic disease. MPC - Miscellaneous pathological conditions.

Post-processing

- We found that sometimes the segmentation model will yield unrealistic labels
- We follow Renna et al., (2019) by forcing unrealistic labels to be the next stage, i.e. maintaining the S1-systole-S2-diastole sequence

Fig 9(right): We can observe a slight improvement in F1 score of all the models. PP – post-processing.

Model	PP	P+(%)	Se (%)	F1 (%)	ER
F_BiGRU_200_8	Yes	93.95	93.45	93.68	0.1214
F_BiGRU_200_8	No	93.46	93.78	93.59	0.1262
F_BiGRU_150_8	Yes	95.00	94.31	94.64	0.1035
F_BiGRU_150_8	No	94.56	94.62	94.57	0.1068
F_BiGRU_100_8	Yes	94.64	94.00	94.30	0.1101
F_BiGRU_100_8	No	94.26	94.29	94.26	0.1129
F_BiGRU_50_8	Yes	93.36	93.32	93.32	0.1306
F_BiGRU_50_8	No	92.80	93.65	93.18	0.1361
F_BiGRU_25_8	Yes	93.47	93.21	93.32	0.1299
F_BiGRU_25_8	No	92.88	93.66	93.22	0.1350

Post-assessment safeguard

- Calculate the standard deviation of each heart cycle period of the segmentation and use it as a Post-assessment safeguard to reject bad segmentation results caused by noise contamination.
- Mean of standard deviation index of the ground truth of 11560 10-second window is approximately 0.0247.
- With experiments on in-house mobile phone audios, we find that a reasonable threshold should be set between 0.02 and 0.025.

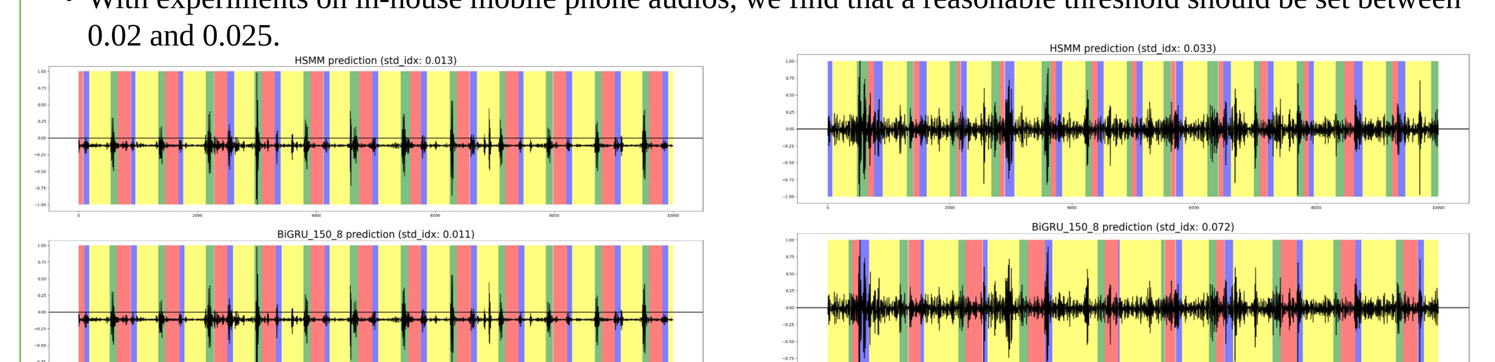

Fig 10: (left) A clean and well-segmented audio with its standard deviation index; (right) A contaminated audio with its standard deviation index. Standard deviation index is empirically useful in identifying low quality heart sound segmentation.

Conclusion and future work

Contribution

- We experiment on utilizing Recurrent Neural Networks to perform heart sound segmentation on a 10-second sound window and the best model outperforms the HSMM baseline by 3% in F1 score on a separate test set.
- We propose a novel workflow to facilitate the heart sound segmentation model in the real-life scenario, including denoising, post-processing and post-assessment.
- We further test the best RNN model on an in-house mobile-phone collected heart sound dataset with the aforementioned workflow and find it achieves satisfactory performance.

In the future

- We hope to collect more labeled mobile-phone collected heart sound data.
- We need to consider the unbalance in the training data.
- We want to experiment with more models and different architectures.
- We can further tune the standard deviation thresholds.
- We will proceed to heart sound classification.

Reference

Clifford, G. D., Liu, C., Moody, B., Springer, D., Silva, I., Li, Q., & Mark, R. G. (2016, September). Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016. In *2016 Computing in cardiology conference (CinC)* (pp. 609-612). IEEE.

Messner, E., Zohrer, M., & Pernkopf, F. (2018). Heart Sound Segmentation—An Event Detection Approach Using Deep Recurrent Neural Networks. *IEEE Transactions on Biomedical Engineering*, 65(9), 1964–1974. https://doi.org/10.1109/tbme.2018.2843258

Renna, F., Oliveira, J., & Coimbra, M. T. (2019). Deep Convolutional Neural Networks for Heart Sound Segmentation. *IEEE Journal of Biomedical and Health Informatics*, *23*(6), 2435–2445. https://doi.org/10.1109/jbhi.2019.2894222

Springer, D., Tarassenko, L., & Clifford, G. (2015). Logistic Regression-HSMM-based Heart Sound Segmentation. *IEEE Transactions on Biomedical Engineering*, 1. https://doi.org/10.1109/tbme.2015.2475278

Varghees, V. N., & Ramachandran, K. (2014). A novel heart sound activity detection framework for automated heart sound analysis. Biomedical Signal Processing and Control, 13, 174–188. https://doi.org/10.1016/j.bspc.2014.05.002

World Health Organization. (2021, June 11). *Cardiovascular diseases (CVDs)*. Cardiovascular Diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)