

# Detecting abnormal heart sound using mobile phones and on-device IConNet

Linh Vu  
linh.vu@monash.edu  
Monash University  
Malaysia

Thu Tran  
ndttran.2019@phdcs.smu.edu.sg  
Singapore Management University  
Singapore

## ABSTRACT

Given the global prevalence of cardiovascular diseases, there is a pressing need for easily accessible early screening methods. Typically, this requires medical practitioners to investigate heart auscultations for irregular sounds, followed by echocardiography and electrocardiography tests. To democratize early diagnosis, we present a user-friendly solution for abnormal heart sound detection, utilizing mobile phones and a lightweight neural network optimized for on-device inference. Unlike previous approaches reliant on specialized stethoscopes, our method directly analyzes audio recordings, facilitated by a novel architecture known as IConNet. IConNet, an Interpretable Convolutional Neural Network, harnesses insights from audio signal processing, enhancing efficiency and providing transparency in neural pattern extraction from raw waveform signals. This is a significant step towards trustworthy AI in healthcare, aiding in remote health monitoring efforts.

## CCS CONCEPTS

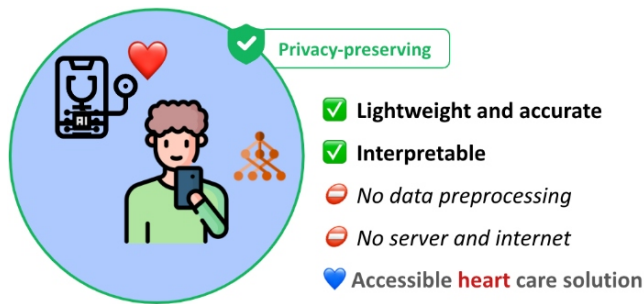
• **Applied computing** → *Bioinformatics*.

## KEYWORDS

abnormal heart sound detection, AI in healthcare, signal processing

## ACM Reference Format:

Linh Vu and Thu Tran. 2024. Detecting abnormal heart sound using mobile phones and on-device IConNet. In *N2Women '24: Network Networking Women Professional Development, June 03, 2024, Tokyo, Japan*. ACM, New York, NY, USA, 3 pages.



**Figure 1: System illustration: abnormal heart sound detection using mobile phones**

*N2Women '24, June 03, 2024, Tokyo, Japan*

© 2024 Association for Computing Machinery.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *N2Women '24: Network Networking Women Professional Development, June 03, 2024, Tokyo, Japan*.

## 1 INTRODUCTION

The cardiovascular disease screening process detects abnormalities such as heart murmur, which is an irregular sound audible during the heartbeat cycle through a stethoscope. Detection of a heart murmur suggests underlying cardiac issues, prompting further evaluation through echocardiography and electrocardiography tests to pinpoint the specific heart disease. To enhance the accessibility of early diagnosis, we introduce a novel system for detecting abnormal heart sounds using mobile phones and an on-device neural network. Our system does not require extra equipment, a server, or a specific data preprocessing pipeline, which is an advantage compared to existing works.

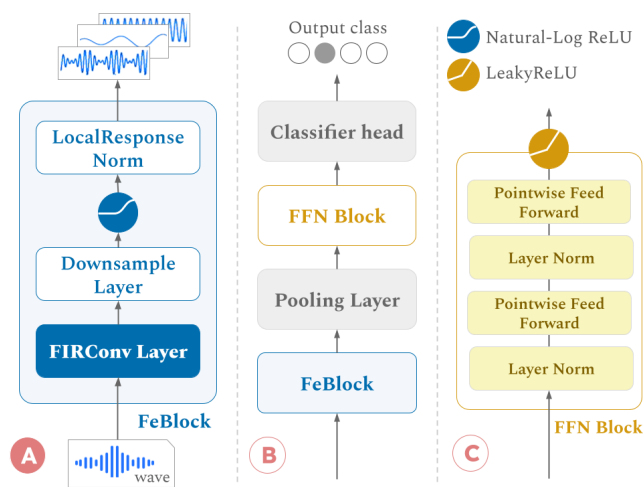
## 2 BACKGROUND

As mobile devices become more popular and have more computational power, various systems have been developed to identify heart murmurs using mobile phones and machine learning. Ashrafuzza-man et al. [1] introduced a novel system utilizing smartphone technology, including the camera and a mobile stethoscope, to estimate heart rate and detect heart attacks and other related diseases by employing Fuzzy Logic, a component of Data Mining. Thiyagaraja et al. [6] developed a mobile app that read input from a customized stethoscope to record the heartbeat sound. Their system uses both discrete and continuous wavelet transforms to downsample the audio, then extract Mel-Frequency Cepstral Coefficients (MFCC) to train a Hidden Markov Model for classification. However, these approaches have some drawbacks, such as the need for preprocessing algorithms and specific equipment. It is uncertain whether these methods would suit real-world applications where the recording environment and audio quality can vary.

As deep learning has become more prevalent with many advanced techniques to help model the data in its raw form, recent studies have been developing different deep learning models to tackle this problem. The main challenge of this approach is that long-recorded audio often includes mixed signals such as heart sound, breath, and background noise. Therefore, a complex preprocessing pipeline is mandatory. For example, Li et al.'s pipeline in [3] involves a 2000Hz downsampling, a 5th-order Butterworth low-pass filtering of the 0-400Hz band and the signal pre-emphasis algorithm, before extracting MFCC and its derivatives to apply CNN or LSTM models or a combination of both. While achieving high accuracy, these models are complex and not mobile-friendly. In [5], Talab et al. proposed a system that uses mobile phone recording and a deep neural network running on the server side. Sending heart sound recordings over the network to be analyzed on the server raises privacy and efficiency concerns.

### 3 SYSTEM DESIGN

We propose using IConNet to identify abnormal heart sounds from mobile phone recordings. The IConNet model is an end-to-end lightweight neural network architecture that eliminates the need for heart sound segmentation, low-pass filtering, and MFCC-based feature extraction. The IConNet consists of two front-end blocks with 128 and 32 kernels, respectively, a max-pooling layer and a 2-layer feed-forward network (FFN) classifier with 256 nodes on each layer. The total number of parameters is 154180, of which 45568 parameters come from the front-end blocks. This is much smaller compared to MobileNet and MobileNet v2 models [4] having 4.2 million and 3.4 million parameters, respectively. The IConNet model size is 493.3 kB without further optimization, such as parameter quantization, satisfying on-device resource constraints.



**Figure 2: The IConNet architecture for end-to-end audio classification: A- the front-end block containing the FIRConv layer; B- the proposed general architecture for end-to-end audio classification; C- the proposed classifier for abnormal heart sound detection.**

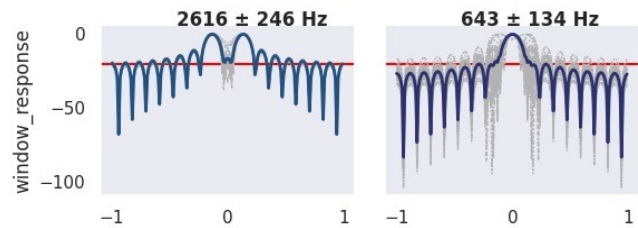
### 4 EVALUATION

We employ the widely-used PhysioNet/CinC Challenge dataset<sup>1</sup> for heart sound classification evaluation. This dataset comprises 2575 *normal* and 665 *abnormal* samples. To validate the effectiveness of the IConNet in identifying relevant features, we resample the waveform from 2000 Hz to 16000 Hz and conduct 4-fold cross-validation, reporting UA and F1 metrics. Deng et al.[2] serve as our baseline, utilizing a preprocessing pipeline with a bandpass filter, MFCC features, and a CRNN model consisting of three 2D CNN layers and two LSTM layers. We also include the MFCC performance for comparison. Preprocessing for other models involves waveform trimming and downsampling, excluding the IConNet model.

<sup>1</sup><https://physionet.org/content/challenge-2016/1.0.0/>

**Table 1: Abnormal heart sound detection result on the PhysioNet dataset**

Model	UA	F1
MFCC + FFN	82.98	88.68
MFCC deltas + CRNN [2]	85.67	90.60
IConNet	<b>87.48</b>	<b>92.05</b>



**Figure 3: Frequency response of filters from different bands. Each sub-figure portrays window filters of a particular frequency range, in which the average response of all corresponding filters (grey color) is presented in vibrant colored lines. The red line at -20dB represents the threshold at which noise is perceived as not noticeable.**

### 5 DISCUSSION

Based on the results presented in Table 1, it is clear that the baseline model [2] performed better than the MFCC + FFN model, thanks to its preprocessing steps that included band-pass filtering and the use of MFCC deltas. The baseline model achieved a 90.06% F1 score. However, our proposed architecture surpassed both models with an F1 score of 92.05%, which is 2% higher than the baseline model. While this result still does not yet outperform the state-of-the-art Resnet result reported by Li et al. in [3], it has successfully demonstrated the effectiveness of our proposed method in classifying heart sound data.

Furthermore, the visualization of the front-end filters in Figure 3 confirms that it allocates band-pass filters that actively change the window shapes to extract essential information in the range of  $643 \pm 134$  Hz. The windows have learned to transform into band-stop filter shapes for the high-frequency range above 2000 Hz, which only contain meaningless artifacts from the resampling step. Understanding the features utilized by the backbone neural network model, which influences its decisions, is vital for ensuring reliable outcomes, particularly in health applications.

### 6 CONCLUSION

In conclusion, this paper presents the initial phase in the development of a mobile-based heart health monitoring system. Future enhancements will enable users to track a broader range of symptoms and receive tailored recommendations for medical check-ups. Additionally, the integration of data from wearable trackers will further streamline heart health management, making it more accessible and convenient. By fostering greater awareness and proactive engagement with heart health, this system aims to empower individuals to take charge of their well-being.

## REFERENCES

- [1] Md Ashrafuzzaman, Md Mazaharul Huq, and et al. 2013. Heart attack detection using smart phone. *International journal of technology enhancements and emerging engineering research* 1, 3 (2013), 23–27.
- [2] Muqing Deng, Tingting Meng, and et al. 2020. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. *Neural Networks* 130 (2020), 22–32.
- [3] Feng Li, Zheng Zhang, Lingling Wang, and Wei Liu. 2022. Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning. *Frontiers in Physiology* 13 (2022), 1084420.
- [4] Mark Sandler, Andrew Howard, and et al. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proc. CCVPR*. IEEE, 4510–4520.
- [5] Elhoussine Talab, Omar Mohamed, and et al. 2019. Detecting heart anomalies using mobile phones and machine learning. In *International Conference on Bioinformatics and Bioengineering*. IEEE, 428–432.
- [6] Shanti R Thiyagaraja, Ram Dantu, and et al. 2018. A novel heart-mobile interface for detection and classification of heart sounds. *Biomedical Signal Processing and Control* 45 (2018), 313–324.