# Automatic Classification of Periodic Heart Sounds Using Convolutional Neural Network

Jia Xin Low, Keng Wah Choo

***Abstract*—**This paper presents an automatic normal and abnormal heart sound classification model developed based on deep learning algorithm. MITHSDB heart sounds datasets obtained from the 2016 PhysioNet/Computing in Cardiology Challenge database were used in this research with the assumption that the electrocardiograms (ECG) were recorded simultaneously with the heart sounds (phonocardiogram, PCG). The PCG time series are segmented per heart beat, and each sub-segment is converted to form a square intensity matrix, and classified using convolutional neural network (CNN) models. This approach removes the need to provide classification features for the supervised machine learning algorithm. Instead, the features are determined automatically through training, from the time series provided. The result proves that the prediction model is able to provide reasonable and comparable classification accuracy despite simple implementation. This approach can be used for real-time classification of heart sounds in Internet of Medical Things (IoMT), e.g. remote monitoring applications of PCG signal.

***Keywords*—**Convolutional neural network, discrete wavelet transform, deep learning, heart sound classification.

## I. Introduction

THE recent advancement in field of Artificial Intelligence (AI) has made previously impossible tasks become a reality. In particular, the use of Deep Learning (DL) algorithms such as the CNN, Long-Short Term Memory (LSTM), Recursive Neural Network (RNN), has created impactful applications in the biomedical fields. For instance, applications on skin cancer classification [1], cardiac arrhythmia detection [2], lung sound classification [3], and many others have been reported over the last two years due to the emerging of AI algorithms which have outperformed the traditional methods.

In this research, the Deep Learning approach is proposed for early heart disease detection. Heart disease has been the leading cause of death globally. Early detection and intervention of heart disease will help to mitigate the situation. Automatic recording of PCG and ECG signals beyond hospital or clinics, such as at home or care centre, serves as a useful diagnostic tool to detect early sign of heart disease. In this scenario, individual could be referred to medical doctor for further investigation once abnormalities are detected in their PCG or ECG during their daily lives.

Jia Xin Low is with the Biomedical Engineering and Materials Group, School of Engineering, Nanyang Polytechnic; *corresponding author phone: 65500805 (e-mail: low_jia_xin@nyp.edu.sg).

Keng Wah Choo is with the Biomedical Engineering and Materials Group School of Engineering, Nanyang Polytechnic; phone: 65500587 (e-mail: choo_keng_wah@nyp.edu.sg).

## II. Related Work

There are several studies reported on heart sounds classification. Typical challenges faced are the selection of meaningful features and good training and validation datasets. Along with the advancement in deep learning algorithms, findings with improved heart sounds classification accuracy have been published. These approaches usually require a deep learning neural network to classify heart sounds, which works with both clear signals and noisy signals. The features used include Short Time Fourier Transform [3], Mel-frequency cepstral coefficient [4], and a few others time-frequency based features. These existing feature-base approaches require few segments of the heart sounds to be processed at any one time. However, our proposed approach is able to process every single segment of heart sounds per heartbeat. This significantly increases the resolution of the prediction and is able to pin-point exactly which heart beat has produced an erratic heart sounds cycle. In addition, as deep learning algorithms are capable to learn from the datasets provided and identify good features automatically, this research aims to feed the heart sounds time series input as raw as possible into a novel CNN prediction models and push the limits of models for better accuracy.

## III. Methods

The heart sounds datasets that are used in our work belonged to the Massachusetts Institute of Technology heart sounds database (MITHSDB). The ECG and PCG datasets were extracted from the publicly available database, the 2016 PhysioNet/Computing in Cardiology Challenge database [5]. These signals are noisy and are classified into either normal or abnormal. The training datasets are downloaded from the dataset; a total of 180 unique datasets were used in our classification model development. Since our model requires ECG as the pre-condition for automatic heart sound segmentation, only datasets with detectable ECG were used, regardless of the quality of the heart sounds. 80% of the database serves as the training set for CNN prediction model development; while the remaining 20% is used as the testing set for model validation. The performance of the model is then compared using the classification accuracy.

### *A. Discrete Wavelet Transform*

Discrete Wavelet Transform (DWT) is a technique which is widely used in signal processing due to its great time and frequency localization ability. It can also be used to extract the local characteristics of the ECG signal. It is defined as:

$$W_\varphi(j_0,k) = \frac{1}{\sqrt{M}}\sum_x f(x)\varphi_{j_0,k}(x) \quad (1)$$

$$W_\psi(\mathrm{j},k) = \frac{1}{\sqrt{M}}\sum_x f(x)\psi_{\mathrm{j},k}(x) \quad (2)$$

for $j \geq j_0$ and the Inverse DWT (IDWT) is defined as:

$$f(x) = \frac{1}{\sqrt{M}}\sum_k W_\varphi(j_0,k)\varphi_{j_0,k}(x) + \frac{1}{\sqrt{M}}\sum_{j=j_0}^{\infty}\sum_k W_\psi(\mathrm{j},k)\psi_{\mathrm{j},k}(x) \quad (3)$$

where $f(x)$, $\varphi_{j_0,k}(x)$, and $\psi_{\mathrm{j},k}(x)$ are the functions of the discrete variable $x = 0,1,2,\ldots,M{-}1$. Normally we let $j_0 = 0$ and select $M$ be a power of 2 (i.e., $M = 2^J$) so that the summations in (1)-(3) are performed over $x = 0,1,2,\ldots,M{-}1$, $j = 0,1,2,\ldots, J{-}1$, and $k = 0,1,2,\ldots,2^j - 1$. The coefficients defined in (1) and (2) are usually called approximation and detail coefficients, respectively.$\varphi_{j_0,k}(x)$ is a member of the set of expansion functions derived from a scaling function $\phi(x)$, by translation and scaling using:

$$\varphi_{\mathrm{j},k}(x) = 2^{j/2}\varphi(2^j x - k) \quad (4)$$

$\psi_{\mathrm{j},k}(x)$ is a member of the set of wavelets derived from a wavelet function $\Psi(x)$, by translation and scaling using:

$$\Psi_{\mathrm{j},k}(x) = 2^{j/2}\psi(2^j x - k) \quad (5)$$

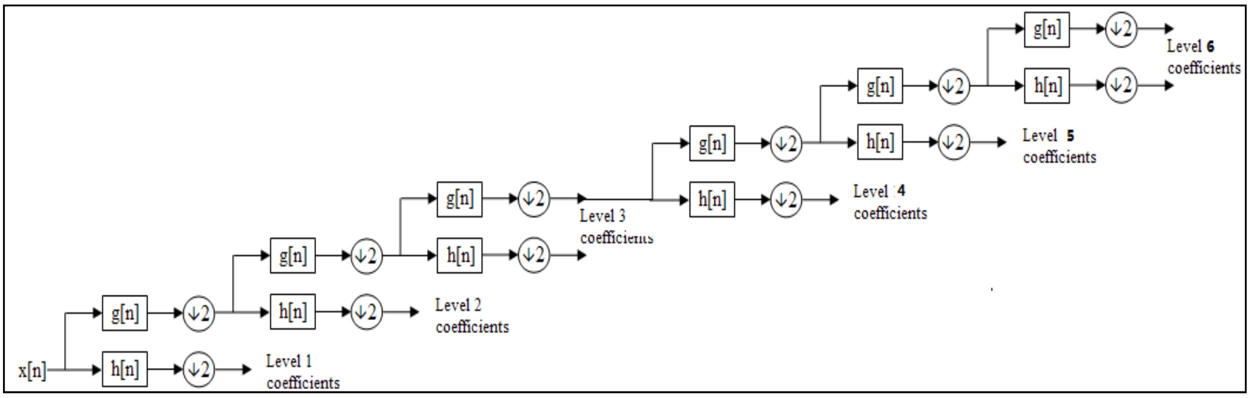
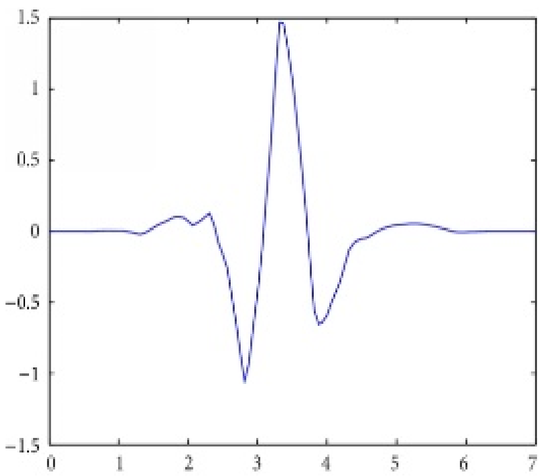
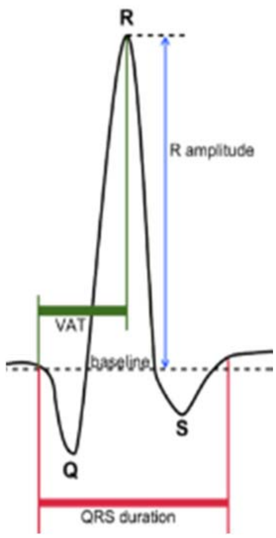



Fig. 1 A 6th level DWT decomposition model used for QRS peaks detection from the ECG signals

In this research, the ECG signal is decomposed up to 6th level (Fig. 1) using the 'sym4' wavelet as it resembles the QRS complex (shown in Fig. 2), which makes it a good choice for QRS detection. All coefficients are then set to zero except the 6th level DWT coefficients, which are used to reconstruct the ECG signal.




(a)




(b)

Fig. 2 Diagram of (a) sym4 wavelet, (b) QRS complex

*B. Preprocessing of Recorded Signal*

In the preprocessing stage, ECG signals were filtered by a bandpass FIR filter between 20 Hz and 200 Hz. High pass FIR filter with 20 Hz cut off frequencies is applied to heart sound datasets. Next, the ECG signals are decomposed and reconstructed into time-frequency representations using DWT technique as described. By taking the square values and applying peak detection algorithm, the QRS peaks are detected from the ECG signals.

Based on the detected QRS peaks location, the periodic heart sounds containing S1 and S2 are segmented and extracted from the respective PCG signals. Fig. 3 shows a typical periodic heart sounds segment extracted from this process. Due to interpersonal and intrapersonal heart rate variability, the total number of sample within each segment varies significantly. A resampling process is applied to normalize the signals in the following section to address this.

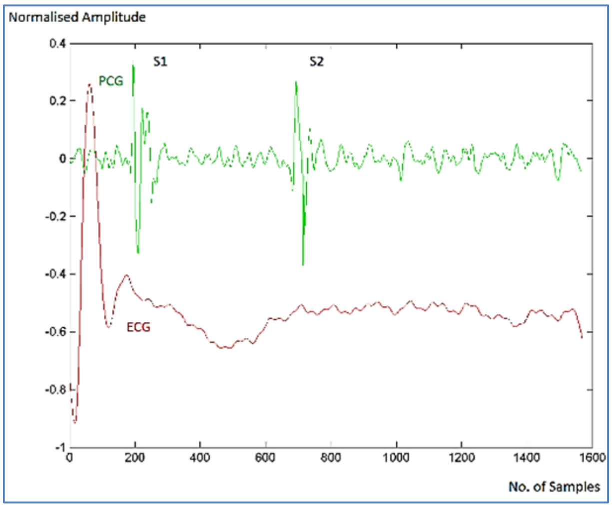



Fig. 3 A heard sound (PCG) segment recorded concurrently with ECG, indicating the S1 and S2 heart sounds within cycle of ECG – periodic heart sounds

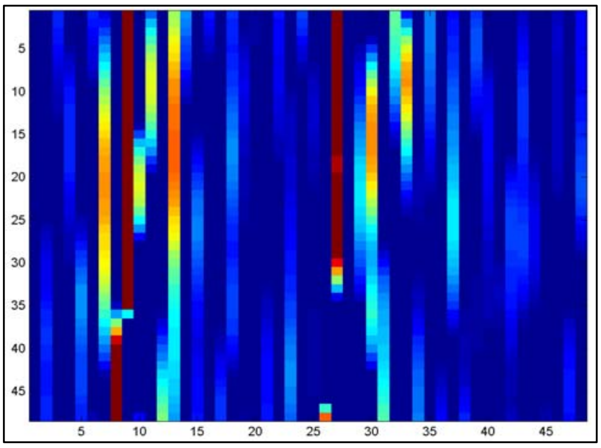


Fig. 4 A typical intensity map constructed from a healthy periodic S1S2 heart sound segment

## C. Transformation to Intensity Map

The preprocessed signal, which is a one-dimensional time series, is normalized by re-sampling into a standard 2304-element vector. The objective of this process is to remove variation between heart sounds segments, either between individuals, or within the same PCG recording. Each time-series segment was then transformed into a 48x48 square matrix, forming a new two-dimensional intensity map, as shown in Fig. 4. Distinctive feature is then observable through the peak values contributed by typical healthy S1 and S2 signals. However, when systole murmur is present in between S1 and S2, which is classified as abnormal heart signal, the distinct feature described earlier vanishes. More peaks can be observed in between S1 and S2, and the intensity map of such signal is shown in Fig. 5. It serves as a recognizable feature that could be fed into our DL prediction model, for normal and abnormal heart sounds classification. It is understood that through training, the model will be able to learn, analyze, classify different types of abnormalities, and differentiate them from the normal ones.

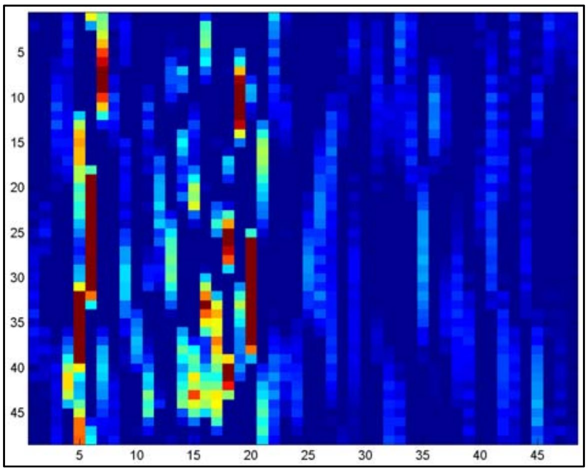


Fig. 5 An intensity map constructed from a periodic S1S2 heart sound segment with systole murmur

## D. CNN Classification

CNN classifier is known to have the capability to automatically adjust its coefficients to improve classification of the input data. The transformation of the heart sound time-series into intensity map fits nicely into a typical architecture of CNN model. This type of neural network is typically trained using supervised learning. The number of hidden layer depends largely on the availability of large datasets. Having deeper layers of network would increase the number of weighs and biases to be updated, as well as the training time. Another disadvantage of deeper network is the tendency for the model to be biased towards the training data, making it challenging for generalized datasets. In our work, two and three convolutional layers CNN models are chosen, with feature maps of varying dimensions, ranging from 32 to 64 and 256 respectively in each convolutional layer.

For our CNN architecture, it is first started with a convolutional layer of 11x11 as kernel size, producing a hidden layer of (48x48x32) feature maps. After the first convolutional layer, a 2x2 pooling is imposed to summarize the information extracted from the feature maps into a hidden layer of (24x24) matrix. Then, a second convolutional layer of the 3x3 kernel size is applied to produce a hidden layer of

(24x24x64) feature maps. Another 2x2 pooling is imposed after the second convolutional layer, to summarize the information extracted from these feature maps into a hidden layer of (12x12) matrix. In the last step, a densely connected layer of dimension 1024 is built, before a softmax activation function is applied to generate the output data. A drop-out feature is introduced to avoid overfitting [6] of input dataset by the CNN model. This forms the first basic 2 convolutional layers CNN model (referred to as Model 1) for our supervised learning algorithm. The proposed models are built using Python Tensorflow API [7]. Our models are trained using stochastic gradient descent based on Adam optimizer [8]. The architecture is shown in Fig. 6, generated using the TensorBoard library provided by Tensorflow.

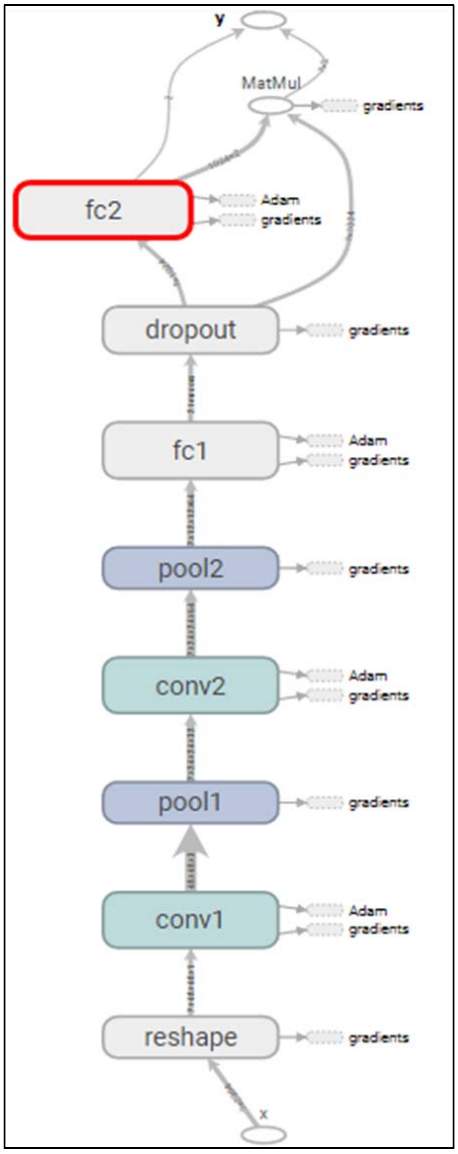



Fig. 6 The architecture of the heart sound classification using the intensity map

To evaluate the CNN performance further, a second model is created (referred to as Model 2), where two more hidden layers are added; one convolutional and one max-pooling layer to evaluate the effect of additional hidden layers on the prediction accuracy. To evaluate the model performance, the kernel and feature maps dimensions are also experimented. It is known that adding more layers would increase computing resources and time required to complete classification, which would affect the practicality of such system to be put on Internet-of-Medical Things, such as running the model from a smart phone, or an embedded wearable medical device.

*E. Converting Segment-Level Classification into Record-Level Classification*

A simple scheme used to combine periodic heart sound segment-level classifications is deployed into recording level classification. It is done by computing the number of abnormal and normal periodic heart sounds within a single recording. Then, the ratio of abnormal/normal number is computed. When the ratio is larger than 0.5 (as threshold), the PCG recording is considered to be abnormal, otherwise it is normal. The process is depicted in Fig. 7. The recordings with ratio lesser than 0.5 can be further analyzed as these are the potential candidates for early stage heart problem detection.

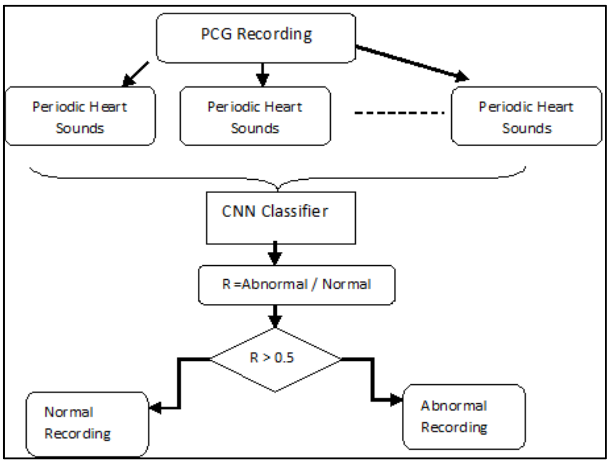



Fig. 7 Simple majority voting scheme for PCG classification

*F. Experimental Design*

The entire training dataset "a" was downloaded for the research. The first 230 datasets (60% of the dataset) with reasonable ECG signals quality are processed, ensuring that heart sounds could be extracted between two ECG pulses using the wavelet-based segmentation algorithm. Among these datasets, 66 are marked as normal and another 164 are marked as abnormal. Each of these datasets has a different recording duration, and hence the number of periodic heart sounds per recording varies. After careful consideration, 1918 normal and 1812 abnormal periodic heart sounds segments are extracted from PCG signals. 80% of these extracted segments are used as input to the CNN classifier for training, and the remaining 20% are used for validation. This cross-validation step was repeated 4 rounds to reduce variability and over fitting. After the model is fine-tuned and finalized, it is used to test the remaining 40% of the dataset (a total of 180), to prove that it could be generalized to classify independent signals.

During the testing phase, ECG signals from these 180 datasets go through the same DWT segmentation process and the periodic heart sounds segments are fed into the CNN classifier. 1437 normal and 3555 abnormal periodic heart sounds segments have been extracted to form the testing datasets. The percentage of the periodic heart sound segments classified as abnormal versus the total number of heart sound segment within each dataset is recorded. If high percentage (>50%) of the heart sound segments is classified as abnormal, this dataset is marked as abnormal, and vice versa. Finally, the accuracy is computed for the PCG recording classification outcomes.

## IV. Results

There are a number of hyper-parameters that have been fine-tuned in our research while identifying the best prediction model and the two key parameters are the dimension of feature maps and the convolutional filter kernel size. Table I shows that the kernel dimension of the first convolutional layer is very critical for the algorithm to pick up the necessary features in order to differentiate between the normal segments from the abnormal ones. Our conclusion is that a size of 11x11 work best for the model.

TABLE I
Effect of Kernel Dimension on CNN Models

| Kernel | 2x2 | 5x2 | 10x10 | 11x11 | 12x12 |
|---|---|---|---|---|---|
| Accuracy (Model 1) | 0.60 | 0.66 | 0.79 | 0.82 | 0.66 |
| Accuracy (Model 2) | 0.69 | 0.71 | 0.82 | 0.86 | 0.71 |

Besides, the prediction model is evaluated with varying drop-out values from 0.5 to 1.0. As our data size is not large, it did not show any significant change in prediction as the value varies. As a result, the drop-out value is kept at 1.0. The first model built with two convolutional layers performs very well in the training and validation phase, achieving accuracy as high as 0.82, and is able to generalize in the testing phase, achieving an accuracy of 0.71 (shown in Table II). As predicted, its performance will be affected when the heart sound quality is poor. This is likely due to the limited size of the training data.

Our second model with three convolutional layers is able to produce better classification accuracy of 0.86, and is able to generalize in the testing phase, achieving an accuracy of 0.75, which out-performs the model 1 as anticipated.

TABLE II
Accuracy of Model 1 & 2 at Training/Testing Phase

| Optimal Kernel (11x11) | Training & Validation Phase | Generalize Testing Phase |
|---|---|---|
| Accuracy (Model 1) | 0.82 | 0.71 |
| Accuracy (Model 2) | 0.86 | 0.75 |

## V.Conclusion

Two Convolutional Neural Network (CNN) models that are capable of performing automatic abnormal heart sounds classification at the resolution of every heartbeat have been developed and evaluated. This approach uses DWT for PCG segmentation and 2 & 3 Convolutional layers CNN taking in intensity maps as its inputs. Our approach allows one to analyze every single heart sounds segments between two heartbeats, allowing the system to pick up potential abnormal heart sounds from a stream of normal heart sounds. This feature is particularly important for early stage heart disease diagnose. Although a very promising accuracy is achieved, there are much more work to be done to fine tune the models. Several hyper-parameters such as the kernel size, number of feature maps and the normal / abnormality thresholds can be adjusted to achieve higher classification accuracy. The performance of the ECG segmentation also highly depends on the recorded signal quality. The automatic classification would be affected if the segmentation of the periodic heart sounds fails. Alternative methods will be explored to extract the periodic heart sounds even when ECG signal quality is poor. Our future work might also involve exploring other advanced neural networks, such as the LSTM and RNN, not limited to normal/abnormal heart sounds classification, but also various murmur types detection within the heart sounds.

It is concluded that CNN models can be applied to heart sounds classifications down to each heart beat resolution in the case where ECG and PCG are acquired concurrently. This ability allows future development of an automatics diagnostic tool, either as a dedicated device or wearables, where early sign of heart disease can be detected for early intervention.

## Acknowledgement

The authors would like to thank NYP for providing the resources to support the conduct of this research.

## References

[1] Andre Esteva1, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” in Nature 542, 115–118 (02 February 2017).
[2] Pranav Rajpurkar, Awni Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Ng, “Cardiologist-Level Arrhythmia Detection With Convolutional Neural Networks,” in https://arxiv.org/abs/1707.01836.
[3] Qiyu Chen, Weibin Zhang, Xiang Tian y, Xiaoxue Zhang, Shaoqiong Chen and Wenkang Lei, “Automatic Heart and Lung Sounds Classification using Convolutional Neural Networks,” in http://ieeexplore.ieee.org/abstract/document/7820741/.
[4] Jonathan Rubin, Rui Abreu, Anurag Ganguli, Saigopal Nelaturi, Ion Matei, Kumar Sricharan, “Recognizing Abnormal Heart Sounds Using Deep Learning,” in https://arxiv.org/abs/1707.04642.
[5] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AE, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD, “An open access database for the evaluation of heart sound algorithms,” Physiological Measurement 2016;37(9).
[6] Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 2014;15(1):1929–1958.
[7] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster,

Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[8] Kingma D, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv14126980 2014.