Employing Adversarial Machine

Learning and Computer Audition

for Smartphone-Based Real-Time

Arrhythmia Classification in Heart

Sounds

Aditya Kendre Cumberland Valley HS

ENBM0 3 5 Mechanicsburg, PA, USA

Introduction

Background Constraints

Around 33% of all deaths are the result of Cardiovascular Constraints include the use of a small dataset and noisy

diseases, these diseases often cause arrhythmia1 [1]. recordings of heart sounds auscultated5 at different locations.

Traditional arrhythmia diagnosis requires Electrocardiogram2

(ECG) analysis, this limits the use case to hospitals and Related Works

clinics with specialized equipment. Hence, 13.1% of patients Although heart sound databases do exist, these datasets are

have undiagnosed Atrial Fibrillation3 (AF) [2]. Heart sounds limited by the number of pathologies that are collected, often

or Phonocardiograms4 (PCGs) provide a distinct advantage having to divide the dataset into two categories: normal and

over traditional ECGs, in that it records the acoustic abnormal. Currently, only three major supervised PCG

properties of the heart’s movement. This allows for greater datasets exist PhysioNet Classification of Heart Sound

versatility and ease of use [3]. Challenge dataset, PASCAL Heart Sound Challenge dataset,

and the Littman Heart Sound and Murmur Library.

Hypothesis Additionally, little effort has been done to increase the

If a novel heart sound analysis system is developed to detect development and labeling of comprehensive datasets of PCG

a variety of arrhythmias, then arrhythmias will have a signals that cover the complete range of pathologies.

decreased undiagnosed rate. In diagnosing heart sounds, two major challenges arise:

localization and classification. Localization aims to find the

Engineering Goals position of biomarkers in heart sounds. By doing this, heart

1.Increase the Number of Cardiovascular Pathologies sounds can be segmented into signals containing a single

analyzed in heart sounds- Develop a model to construct

heart sound. Furthermore, classification attempts to

heart sounds from pre-existing datasets that cover the

categorize heart sounds into normal and abnormal groups by

complete range of pathologies that are likely to be

exploiting the information extracted from localization.

encountered in clinical settings.

Conventional heart sound localization and classification

2.Develop an End-to-End System - Create a system that is

able to record, analyze, and predict (end-to-end system) methods involve time and/or frequency and are typically

heart sounds for Cardiovascular modalities without dependent on machine learning algorithms to enhance the

specialized equipment. results. These algorithms typically include artificial neural

3.Real-World Testing - Test the end-to-end system in a real- networks (ANNs), support vector machines (SVMs), self-

world environment to ensure practicality and generality of organizing maps (SOMs), and are limited to the number of

the system. samples and pathologies covered in a given dataset. This leads

1. Irregular heartbeat to a surface-level analysis of the heart sounds.

2. Recording of the electrical signals produced by the heart

3. A type of arrhythmia that produces an irregularly irregular heartbeat

4. Recordings of the sounds created by the physical movement of the heart

5. Listening to sounds created from the heart

Methods

Dataset Environment &

Data Management Dataset

Type

Lengths

Recording Quality

Pathologies Ratios

Although heart sounds are analyzed more often than Classification of

PCG Normal: 3541 (77.1%)

ECG recordings, a better variety of ECG datasets exist. Heart Sound 5-120 Extremely noisy and

& Abnormal: 551 (12.0%)

Recordings - seconds low signal quality

Thus, we proposed using both ECG and heart sound PhysioNet 2016

ECG Noisy: 501 (10.9 %)

(PCG) datasets for arrythmia classification. PCG

Normal: 351 (40.0%)

recordings often are recording in non-ideal Noisy and taken from Murmur: 129 (14.7%)

1-30

environments that are filled with unwanted background PASCAL 2011 PCG

seconds

iStethoscope and Extra: 65 (7.4%)

digital stethoscopes Artifact: 86 (9.8%)

noise and interference. Hence, we preprocess the data Unlabeled: 247 (28.1%)

by denoising, normalizing, standardizing, and Stenosis: 5 (31.3%)

transforming the signal. This allows a model to extract Septal: 1 (6.3%)

Littman Heart Ejection: 3 (18.8%)

meaningful features efficiently, revealing the Sound Clean and taken from Coarctation: 1 (6.3%)

physiological structure of the heart sounds. The data is PCG 2 seconds

& digital stethoscope Prosthetic: 1 (6.3%)

then augmented to enable a significant increase in the Murmur Library Regurgitation: 2 (12.5%)

Pericarditis: 1 (6.3%)

diversity of data available while training a model, this is Gallop: 2 (12.5%)

done without collecting new data. It aims to slightly Normal: 9528 (34.2%)

alter existing data to a point where the model cannot CD: 5486 (19.7%)

PTB-XL ECG 10 seconds n/a MI: 5250 (18.9%)

recognize the augmented data as one it has trained on HYP: 4907 (17.6%)

before, but still maintains important distinguishing ST/TC: 2655 (9.5%)

characteristics. Table 1: Dataset table that shows the dataset type (PCG or ECG), the lengths of the recordings (in

seconds), the environment in which the signals were recorded, the recording quality and the

Cross-validation number of categories in the dataset (Kendre, 2021).

Cross-validation is used to estimate how accurately a predictive model will

perform in the real-world. In a prediction problem, a model is usually given a

dataset of known data on which training is done (training set). The training

process optimizes the model parameters via backpropagation to make the

model fit the training set as well as possible. Additionally, we need to assess

the best place to stop training to ensure the model does not overfit. Thus,

we need to test the model on an independent set (validation set) to validate

the model’s success. However, this creates a bias towards the samples in the

training and validation set. Hence, we need a third dataset (testing set) that

the model has no bias to evaluate the model. Here, the data was split into Figure 1: Visualizes the ECG and PCG signals recorded

simultaneously, both PCG (bottom) and ECG (top) show evidence of

70% of training, 10% of validation, and 20% of testing. systole and diastole (Behbahani et al).

Methods

Model Development

We propose using Generative Adversarial Networks

(GANs) along with Convolutional Neural Network (CNNs)

and Transformers1 for both heart sound analysis and

heart sound synthesis based on ECGs. GANs pose a

unique advantage over traditional machine learning and

deep learning methods, in that the model learns to mimic

a dataset by creating its own data while classifying the

generated data and the data from the dataset.

Furthermore, Transformers introduce an attention

mechanisms that gives context to all inputs, allowing the

model to enhance important non-linear relationships

within in each signal.

In heart sound analysis, the PCG CNN Encoder is pre-

trained2 to compress the PCG signals into a latent space3,

and then reconstructed into PCGs using the PCG CNN

Decoder. After pre-training, the PCG signal’s latent spaces

are fed into the GAN discriminator (Transformer

Encoder). Here, the Transformer creates a prediction

based on latent space. Additionally, the PCG Generator

tries to mimic PCGs to fool the Transformer into

predicting the generated PCG as normal/abnormal, rather

than as noisy.

In heart sound synthesis, the process of ECG

reconstruction is identical to that of PCG reconstruction.

However, during pre-training the ECG’s latent spaces are Figure 2: The pipeline of the CNN based generator and Transformer based discriminator.

optimized towards the PCG’s latent space. This novel This shows the process of using both PCG and ECG datasets for heart sound (PCG) analysis

concept forces the Generators to produce equivalent and synthesis (Kendre, 2021).

latent spaces for ECG and PCG (given both signals were

recorded simultaneously). After pre-training, the ECG 1. Machine Learning architectures used for feature extraction

2. Initial training that is done to aid the model better performance in a shorter

CNN Encoder is fed ECG data from categorized arrhythmia 3.

amount of time and computational resources

A representation of compressed signals

datasets.

Methods Mean Squared Error:

𝑀𝑆𝐸(𝑥, 𝑦) = {𝑙1 , … , 𝑙𝑁 }⊤ , 𝑙𝑛 = (𝑥𝑛 − 𝑦𝑛 )2

Model Learning

During the training phase, backpropagation is used to optimize

Cross Entropy Loss:

the model's weights and biases. A cost function calculates an exp(𝑥[ 𝑐𝑙𝑎𝑠𝑠 ])

CE(𝑥, 𝑐𝑙𝑎𝑠𝑠) = −log( ) = −𝑥[ 𝑐𝑙𝑎𝑠𝑠 ] + log(෍ exp(𝑥[𝑗]))

error term (loss) based on the model's prediction. Using the ෌𝑗 exp(𝑥[𝑗])

𝑗

error, a gradient is computed with respect to all the parameters

in the model. An optimizer then updates the weights and biases Accuracy:

based on the loss's gradient. The goal of the optimizer is to TP + TN

minimize the cost function's error, such that it correctly classifies ACC =

TP + TN + FP + FN

the heart sounds or correctly synthesizes the heart sounds. In

this study, we used the Adam optimizer in union with Cross- Sensitivity:

TP TP

Entropy Loss and Mean Squared Error (Table 2). The model is TPR = = = 1 − FNR

only trained on the training set; thus, backpropagation only P TP + FN

occurs on the training set. However, while the model is training, Specificity:

the validation set is used to assess the model's ability to TN TN

generalize (this process does not include backpropagation). The TNR = = = 1 − FPR

N TN + FP

model stops learning when the validation set approaches a limit

and no longer learns new information, this is called early Positive Predictive Value:

stopping. TP

PPV = = 1 − FDR

TP + FP

Model Evaluation

Model metrics aid in quantifying the model’s performance, Negative Predictive Value:

TN

allowing us to compare our methods with existing techniques. NPV = = 1 − FOR

The model is evaluated only on the testing set, as to provide a TN + FN

non-biased evaluation. All metrics are calculated based on the F1 Score:

true positive (TP) rate, false positive (FP) rate, true negative (TN) PPV × TPR 2TP

rate, and false negative (FN) rate. Table 2 shows the metrics used F1 = 2 × =

PPV + TPR 2TP + FP + FN

to evaluate the model.

Time Complexity:

Real-World Testing & Model Deployment 1

𝑛

The model's viability is crucial for ensuring the model's success in O(n) = ෍ 𝑚( 𝑖𝑛𝑝𝑢𝑡 ~𝑛)𝑡

𝑛

the real world. Thus, we need to conduct trials with the end-to- 𝑖=0

end system to ensure we can deploy and integrate the model Table 2: Metrics shows in the table are used to evaluate the model on the

with ease. testing set to ensure generalization and comparability (Kendre, 2021).

Results

Distribution Table Model Metrics on Datasets

The proposed method introduces significant features Dataset Accuracy F1 Score Sensitivity Specificity PPV NPV

and architectures that aid in abnormality detection.

Using such techniques, the classification metrics PASCAL 73.7±.02 75.0±.04 75.0±.05 72.2±.04 75.0±.02 72.2±.04

showed extremely promising results as shown in Table

3. Specifically, the model achieved 73.7% accuracy on PhysioNet

89.5±.03 64.4±.05 48.0±.06 85.7±.02 96.5±.04 85.7±.03

2016

the PASCAL dataset and 89.5% accuracy on the

PhysioNet dataset. Hence, the model proved to be very

efficient for classification of normal and murmur-filled Synthesized

93.0±.03 94.3±.03 99.5±.03 90.3±.03 99.3±.03 91.8±.03

PTB-XL (AF)

heart sounds. Furthermore, we also propose creating

heart sounds from ECGs for additional arrhythmia-

specific training. Training the model on the synthesized Combined 95.0±.03 94.6±.04 94.3±.04 99.5±.03 90.3±.03 99.3±.02

heart sounds provided a mean accuracy of 93.0%, ECG

to PCG conversion is a feasible approach. Additionally, Table 3: The table displays the metrics collected on the testing set for all classes in the

training the model on all 3 datasets indicated excellent dataset (Kendre, 2021).

results, predicting 95.0% of abnormalities correctly.

ROC/AUC

The receiver operating characteristic curve (ROC) aids in understanding the

model’s ability to predict the correct heart sound class. The curve plots the

true positive rate against the false positive rate at various prediction

thresholds. Based on the ROC, we choose the optimal threshold for all

classes. However, the optimal threshold depend on a subjective trade-off

between the true and false positive rates. Here, we choose to optimize for

increased true positive rate, as we want to ensure all potential subjects with a

disease are sent for further examination. Additionally, the graph illustrates

the area under curve (AUC), this analysis provides an aggregate measure of

the model’s performance over all classification thresholds. All AUC scores are

above the 0.5 threshold; this suggests that the model’s ability to distinguish

between classes is high.

Figure 3: The ROC curve shows the TPR and FPR for different

thresholds for the model’s prediction and the AUC for each

class, which shows the model’s efficiency (Kendre, 2021).

Results

Confusion Matrix

The confusion matrix in Figure 4 illustrates the performance for

each class of the proposed method. Specifically, we evaluated the

model's success on the grounds of accuracy, specificity, and

sensitivity of the classification. We calculated the average true

positives, false positives, false negative and false positive for all

testing sets. Using these floored values, we normalized each label

along the y-axis. The matrix reveals that the most common

misunderstanding occurs between Gallops and Normal rhythms.

This is expected as Gallops are heart sounds that contain an extra

sound like S3 or S4, which are often lacking in amplitude. Overall,

we conclude that the average accuracy of abnormal heartbeat

detection is ~95% with a misclassification rate of just ~5%. Thus, the

model is extremely accurate in detecting abnormalities in heart

sounds and displays capabilities to classify abnormal heart sounds

Figure 4: Matrix of accuracy between categories (pathologies) classified by

into arrhythmia and abnormality types. the model (predicted label) and the true categories (true label) (Kendre,

2021).

t-SNE Visualization

Dataset visualization is critical in understanding the dataset's complexity

and model's effectiveness. Here, we use t-distributed stochastic neighbor

embedding (t-SNE), a statistical method for visualizing multidimensional

data with less computational expense. The method is presented with the

raw prediction values for each input of the validation set and maps the

corresponding predictions into a 2-dimensional space (x, y). Tracked over

the best epoch, the visualization allows us to view the differentiation

between heart sounds from the model’s prediction. The visualization

highlights clear clustering within the dataset, which suggests the model is

stable. Though, it is evident that there is overlapping between abnormal

and normal signals in the y=0.4-0.6 range. Assuming these signals as

ground truth, this implies that additional feature engineering is required to

adequately classify heart sounds.

Figure 5: t-SNE visualization of testing dataset after model

training (Kendre, 2021).

Results Discriminator Time Complexity

Time Complexity

Model complexity is used to gauge and evaluate the efficacy of a model against

an increase in data (n). We mainly focus on the discriminator’s time complexity

as it is most relevant to the problem at hand (space complexity is O(1)).

Depending on model deployment and integration, the complexity can vary. For

example, GPUs have parallel processing capabilities, which allow them to

process multiple signals at once, effectively decreasing the discriminator

complexity to O(1). For this reason, we use the worst-case scenario (a CPU), for

analysis of the proposed method's time complexity. The discriminator’s time

complexity is directly and linearly correlated to the input size, suggesting the

complexity is O(n). This means that the discriminator’s efficiency enables real-

Figure 6: Time complexity of the PCG/ECG Encoder and

time heart sound analysis. Transformer Discriminator in classifying heart sounds

(Kendre, 2021).

End-to-End System

Testing the model's viability is crucial for ensuring the model's success in the real world.

Ideally, recording heart sounds are recorded with digital stethoscopes. These tools use

transducer technology to convert sound into an electrical signal. Over the past decade, this

technology has grown immensely (by cause of speech recognition). Modern phones have

S1 S1

S1 S1 the potential to record the sounds at a high resolution, given the microphone is located at

S2 S2 S2 the correct position relative to the heart. Such a device will prove extremely beneficial in

providing a diagnosis without the need for specialized equipment. Figure 7 shows a heart

sound recording from the phone microphone. The plot shows important biomarkers like S1

and S2, which suggest the microphone suitable for the classification task. This confirms

smartphone microphones do not record excessive amounts of noise that may hinder the

performance of the detection system. The smartphone app not only allows anyone to

record their heart sounds with any device with a microphone but does so without needing

any external equipment; such as a case or a stethoscope. Additionally, the app allows the

users to download and email the sound recordings, providing physicians and cardiologists

with deeper insight. This approach allows the app to also track heart activity over time,

allowing for more awareness of your fitness level, heart health, and emotional health.

Figure 7: Screenshot of created app recording a heart

sound from the built-in microphone (Kendre, 2021).

Discussion

Model Interpretation

Convolutional Encoder Activation Map

Figure 8 visualizes the channels of the Convolutional Encoder

layers which are responsible for extracting features from heart

Summed Weights

sounds. The color of the line represents the importance of the

feature relative to other features. Meaning, the lighter the color,

the more important the feature. The map illustrates that the

model pays more attention to peaks of higher amplitude in the

heart sound. This indicates that the layers are extracting latent

features from the signal. Specifically, the plot depicts the

PCG Signal

extraction of important biomarkers such as S1 and S2.

Furthermore, the extractions make it clear that the S1 sound is

more important, as those are the brightest throughout all layers.

This parallels medical knowledge, as most cardiovascular

anomalies occur in Systole, or at the start of S1. Figure 8: Visualization of CNN encoder when fed a heart sound

(Kendre, 2021).

Comparative System Evaluation Digital Stethoscope Comparison

In a clinical setting, physicians often use digital stethoscopes for

Cardiac

listening to heart sounds. By monitoring PCG characteristics such Device

Frequency Sample

Amplification Landmark Cost

Range (Hz) Rate (Hz)

as amplitude, pitch, and cyclic patterns, physicians differentiate Guide

between abnormalities in the heart sound. Additionally, these 3M Littman

devices allow their users to significantly amplify heart sounds by 20 – 2000 8000 Up to 24x No $499

3200

eliminating ambient noises by filtering background noise and

amplifying the sounds recorded by the sensor. However, the Eko Core 20 – 2000 4000 Up to 40x No $349

average heartbeat is between 60 and 100 bpm, meaning a heart

sound can occur anywhere from once a second to twice a second.

Hence, digital stethoscopes require high sampling rates for Jabes 20 – 1000 8000 Up to 20x No $229

collecting heart sounds. Table 4 conveys most digital stethoscopes

have a sampling rate of 4000-8000 Hz; however, modern phones

16000-

have sampling rates of 16000-48000. This high sampling rate Smartphone 20 – 2000

48000

Up to 40x Yes n/a

provides more accurate and granular control over traditional

digital stethoscopes. Table 4: The table displays different devices that record heart sounds digitals with

their respective features (Kendre, 2021).

Discussion Related Works

Comparative Model Evaluation Study

Classification

Beat types Dataset

Time

Results

techniques Efficiency

Comparing different methods, we observe the

85.7% was the best accuracy reached on the Sensitivity: 96.47%

PhysioNet

PhysioNet dataset and 61.1% was the best Nogueira et al. SVM N&A - Specificity: 72.65%

2016

Overall Score: 84.56%

accuracy reached on the PASCAL dataset.

Conversely, our proposed method archived 89.5%

Sensitivity: 86.73%

accuracy on the PhysioNet dataset and 73.7% PhysioNet

Krishnan et al. DNN N&A - Specificity: 84.75%

2016

accuracy on the PASCAL dataset. The increase in Accuracy: 85.65%

performance is attributed to our semi-supervised

approach, where we used adversarial Conditional Specificity: 95%

PhysioNet

Rubin et al. CNN N&A - Sensitivity: 73%

Generators to generate heart sounds. This exposed 2016

Overall Score: 84%

the Discriminator to a wide range of heart sounds

which assisted the model in optimizing for Bayes Net and

Specificity: 46.9%

generalized features. Additionally, we conducted a Singh et al. N&M PASCAL - Sensitivity: 69.73%

Logit Boost

Accuracy: 61.1%

statistical significance test (t-test) to show the

probability our proposed method's results are due

Sensitivity: 75.0.3%

to random chance. The test concluded the results N, M, G, No PASCAL Specificity: 72.2%

were statistically significant as all p-values were Accuracy: 73.7.0%

less than 0.05. This implies that the null

hypothesis can be rejected, and the results are Sensitivity: 48.0%

Proposed GAN, CNN, PhysioNet

statistically significant. N&A

2016 ~1000 cps Specificity: 85.7%

Method Transformer Accuracy: 89.5%

Comparative Analysis of Prediction Metrics

Sensitivity: 94.3%

PhysioNet

Krishnan et al. N, M, G, AF,

Figure 9: Combined Specificity: 99.5%

No

Proposed Method Comparison Accuracy: 95.0%

between

proposed

method and

state-of-the art Table 5: The table displays studies with their proposed classification techniques,

Singh et al. Accuracy

PASCAL

(Kendre, 2021). beat types (types of heart sounds), dataset, time efficient, and results. N(ormal),

Specificity

Proposed Method A(bnormal), M(urmur), G(allop), No(isy) (Kendre, 2021).

Sensitivity

0 20 40 60 80 100

Percent Correct (%)

Conclusions

We proposed a Generative Adversarial Network (GAN), composed of a Convolution

Transformer Generator and a Transformer Discriminator to detect abnormal heart sounds

in a recording. The results from model testing and evaluation, along with results from the t-

test revealed the proposed method reached better performance than the previous state-

of-the-art methods. The introduction of heart sounds analysis with ECGs allowed for

increased arrhythmia labels for classification and in a time-efficient manner. Furthermore,

the proposed method showed real-world deployment capabilities for autonomous heart

sound abnormality detection with recordings collected from a phone microphone.

In terms of future development, we propose conducting prospective clinical trials with

patients that have different types of arrhythmias. This will allow us to truly test the

generalization capabilities of the model and smartphone app in the real world. Depending

on these results, we may opt to develop a low-cost DIY and clinical solution for increased

sensitivity in heart recordings. Also, applicable fields include medical emergencies that are

time constraint (ER) and developing rural communities that don’t have access to

arrhythmia expertise.

Applications with the development of the multiview approach include language and time

series processing. Specifically, we can train models to convert language to speech and

speech to language without the need for a supervised dataset of language A and language Figure 10: Location of standard cardiac

landmarks used for auscultation with

B. Rather, the model can be trained to convert language A to an intermediary language phone at Mitral area (Kendre, 2021).

(language C), this language can then be converted into language B. Moreover, we can train

the model to reconstruct speech recording directly from electrical signals (EEGs) from the

auditory cortex or reconstruct vision from the visual cortex.

The object of this study was to create a fast and accurate end-to-end heart sound

arrhythmia detection system, capable of detecting abnormalities in real-time without

specialized equipment. While also increasing the number of cardiovascular pathologies

classified. Our proposed method accomplishes exemplary statistics in abnormality

detection and shows promising results in increased heart sound synthesis. Hopefully, this

study will shed light on abnormality detection techniques and give birth to applications

with signal construction.

References

Ali, M. N., El-Dahshan, E.-S. A., & Yahia, A. H. (2017). Denoising of Heart Sound Signals Using Discrete Wavelet Transform. Circuits,

Systems, and Signal Processing, 36(11), 4482–4497. https://doi.org/10.1007/s00034-017-0524-7

Babu, K. A., Ramkumar, B., & Manikandan, M. S. (2017). S1 and S2 heart sound segmentation using variational mode decomposition.

TENCON 2017 - 2017 IEEE Region 10 Conference, 1629–1634. https://doi.org/10.1109/TENCON.2017.8228119

David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, and Antonio Torralba. GAN Dissection:

Visualizing and Understanding Generative Adversarial Networks. arXiv:1811.10597 [cs], December 2018. arXiv: 1811.10597.

Latif, S., Usman, M., Rana, R., & Qadir, J. (2020). Phonocardiographic Sensing using Deep Learning for Abnormal Heartbeat Detection.

ArXiv:1801.08322 [Cs]. http://arxiv.org/abs/1801.08322 Special thanks to my brother ACHYUT

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

PhysioNet/CinC Challenge 2016: Training Sets. (n.d.). Retrieved January 11, 2021, from https://archive.physionet.org/pn3/challenge/2016/

Smilkov, D., Thorat, N., Kim, B., Viégas, F., & Wattenberg, M. (2017). SmoothGrad: Removing noise by adding noise. ArXiv:1706.03825 [Cs,

Stat]. http://arxiv.org/abs/1706.03825

Subasi, A. (2019). Biomedical Signals. In Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques (pp. 27–87).

Elsevier. https://doi.org/10.1016/B978-0-12-817444-9.00002-7

Thoms, L.-J., Collichia, G., & Girwidz, R. (2019). Real-life physics: Phonocardiography, electrocardiography, and audiometry with a

smartphone. Journal of Physics: Conference Series, 1223, 012007. https://doi.org/10.1088/1742-6596/1223/1/012007

Zhang, W., Han, J., & Deng, S. (2017). Heart sound classification based on scaled spectrogram and partial least squares regression. Biomedical

Signal Processing and Control, 32, 20–28. https://doi.org/10.1016/j.bspc.2016.10.004