

# Introduction

With the rapid growth of computational power and complex algorithms, we propose a novel approach to detect arrhythmias in Phonocardiograms (PCGs). Typically, Electrocardiograms are used to diagnose arrhythmias, requiring medical-grade equipment to recognize cardiac illnesses accurately. However, PCGs provide ease of access to everyone who has a device capable of recording audio, allowing medical professionals to treat arrhythmias in the developmental stages. The new design comprises two subsystems: one is based on the relationship between Electrocardiograms (ECGs) and PCGs, and the other between PCGs and arrhythmias. The association between ECGs and PCGs is amended to translate from one space to another, where ECGs become dimensionally reduced, then reconstructed into a PCG signal. The second subsystem uses a Generative Adversarial Networks (GAN), in which both arbitrary PCG signals and generated signals are fed into a discriminator that detects if an arrhythmia is present or if the signal is false.

# Background

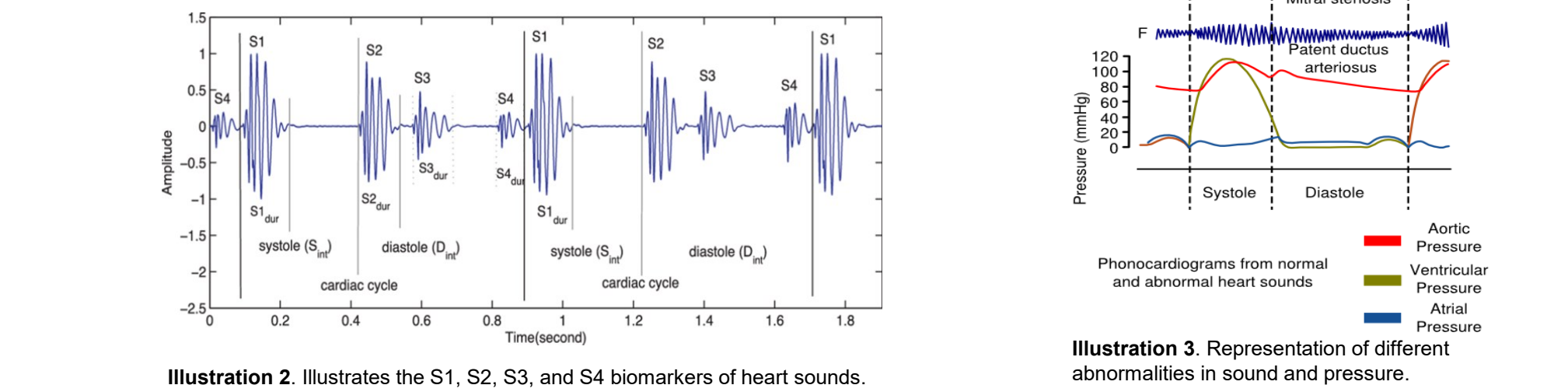
An estimated three million cases of arrhythmia occur in the United States yearly (Mayo Clinic), with 300,000 sudden deaths per year – an incidence rather high than stroke, lung cancer, or breast cancer (American Heart Association). Traditionally, non-invasive arrhythmia analysis is based on multiple electrodes that reflect the electrical activity on ECGs. This method, despite being accurate, limits the use case to hospitals and clinics with specialized equipment; thus, limiting the portability of diagnosing, let alone classification of the type of pathology.

Phonocardiograms (PCGs) are sounds that are created by the mechanical movement of the heart. This physical movement produces four distinct sounds: S1, S2, S3, S4, and murmurs. S1 and S2 are sounds created by a healthy heart; whereas S3, S4, and murmurs refer to diseases or anomalies. The first heart sound, S1, marks the start of Systole. Systole occurs when the heart muscle contracts and pumps blood from the chambers into the arteries. The second heart sound, S2, marks the end of Systole and the start of Diastole. Diastole is a phase of the heartbeat when the heart muscle relaxes and allows the chambers to fill with blood.

# Related Works

Although heart sound databases do exist, these datasets are limited by the number of pathologies that are collected, often having to divide the dataset into two categories: normal and abnormal. Currently, only three major supervised PCG datasets exist: PhysioNet Classification of Heart Sound Challenge dataset, PASCAL Heart Sound Challenge dataset, and the Littman Heart Sound and Murmur Library. Additionally, little effort has been done to increase the development and labeling of comprehensive datasets of PCG signals that cover the complete range of pathologies.

In diagnosing heart sounds, two major challenges arise: localization and classification. Localization aims to find the position of biomarkers in heart sounds. By doing this, heart sounds can be segmented into signals containing a single heart sound. Furthermore, classification attempts to categorize heart sounds into normal and abnormal groups by exploiting the information extracted from localization. Conventional heart sound localization and classification methods involve time and/or frequency and are typically dependent on machine learning algorithms to enhance the results. These algorithms typically include artificial neural networks (ANNs), support vector machines (SVMs), self-organizing maps (SOMs), and are limited to the number of samples and pathologies covered in a given dataset. This leads to a surface-level analysis of the heart sounds.



# Objectives

**Problem:** Current detection methods have limited performance in pathologies and lack real-time classification capabilities.

**Question:** Is it possible to use Generative Adversarial Networks (GANs) to accurately detect arrhythmias in PCGs to decrease the undiagnosed rate of arrhythmias?

**Hypothesis:** If a novel heart sound analysis system is developed to detect a variety of arrhythmias, then arrhythmias will have a decreased undiagnosed rate.

**Solution:** The new semi-supervised approach is composed of two subsystems; the first subsystem uses a discriminator, in which PCG signals are classified into categories based on arrhythmias in the signal. The second subsystem – a generator, aims to generate data such that, when fed into the discriminator, the discriminator will classify the generated data as abnormal or normal. While the discriminator aims to adversarially classify the generated data as fake and classify the real PCG signals to their respective categories. This adversarial approach allows the discriminator to not only extract PCG signal specific features, but also introduces the discriminator potential noisy signals that should not be classified.

# Engineering Goals

- 1. Increase the Number of Cardiovascular Pathologies analyzed in heart sounds-** Develop a model to construct heart sounds from pre-existing datasets that cover the complete range of pathologies that are likely to be encountered in clinical settings.
- 2. Develop an End-to-End System** - Create a system that is able to record, analyze, and predict (end-to-end system) heart sounds for Cardiovascular modalities without specialized equipment.
- 3. Real-World Testing** - Test the end-to-end system in a real-world environment to ensure the practicality and generality of the system.

# Methods

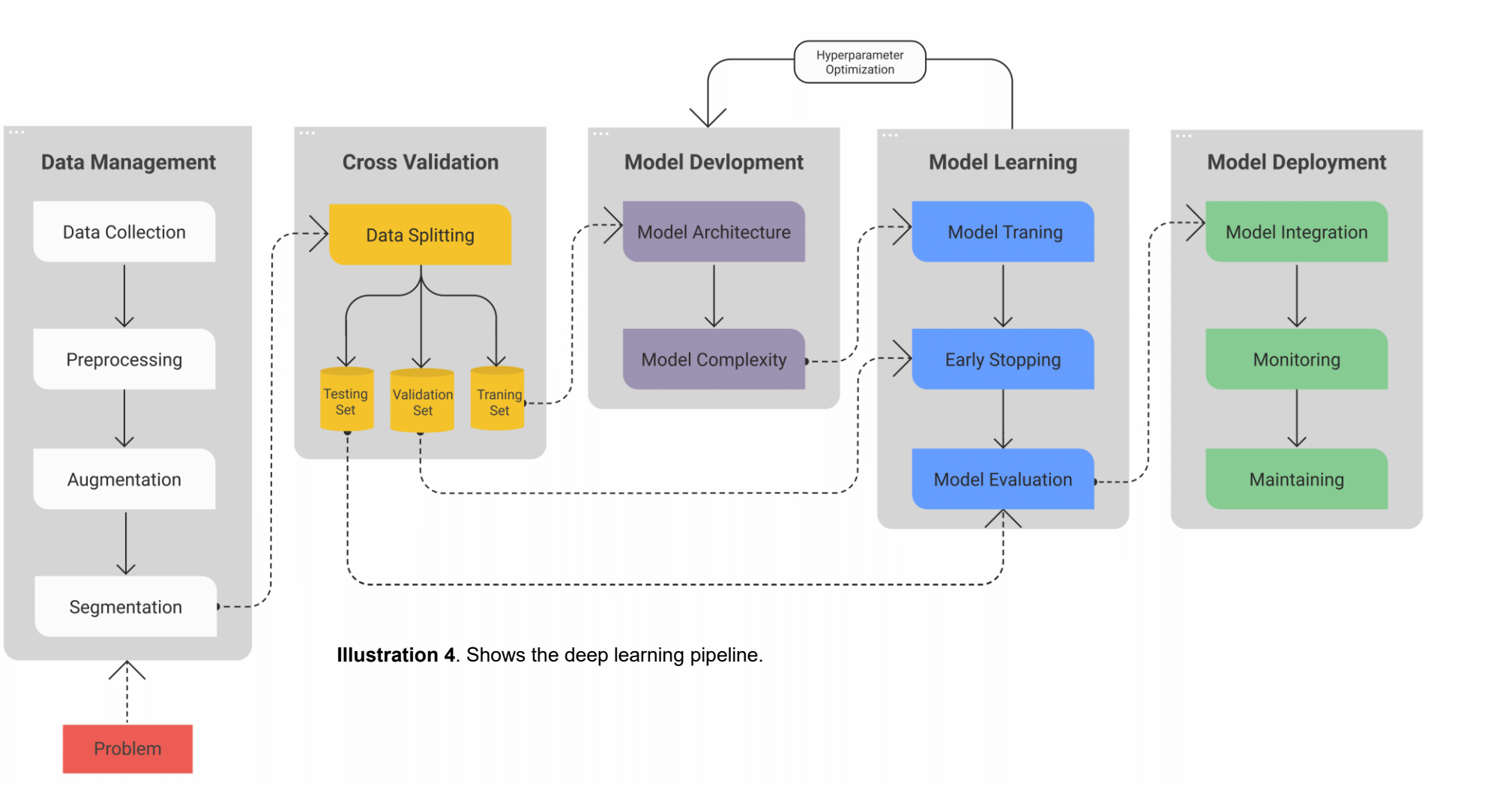


Figure 4: Matrix of accuracy between categories (pathologies) classified by the model (predicted label) and the true categories (true label) (Figure taken by Kendre, 2021).

# Employing Adversarial Machine Learning and Computer Audition for Smartphone-Based Real-Time Arrhythmia Classification in Heart Sounds

# Data Management

Dataset	Dataset Type	Lengths	Environment & Recording Quality	Pathologies Ratios
Classification of Heart Sound Recordings - PhysioNet 2016	PCG & ECG	5-120 seconds	Extremely noisy and low signal quality	<b>Normal:</b> 3541 (77.1%) <b>Abnormal:</b> 551 (12.0%) <b>Noisy:</b> 501 (10.9%)
PASCAL 2011	PCG	1-30 seconds	Noisy and taken from digital stethoscope	<b>Normal:</b> 351 (40.0%) <b>Murmur:</b> 29 (14.7%) <b>Extra:</b> 86 (9.8%) <b>Unlabeled:</b> 247 (28.1%) <b>Stenosis:</b> 5 (1.3%) <b>Septal:</b> 1 (0.3%) <b>Ejection:</b> 5 (18.8%) <b>Coronary:</b> 1 (6.3%) <b>Prosthetic:</b> 1 (6.3%) <b>Regurgitation:</b> 2 (12.5%) <b>Pericarditis:</b> 1 (6.3%) <b>Gallop:</b> 2 (12.5%)
Littman Heart Sound & Murmur Library	PCG	2 seconds	Clean and taken from digital stethoscope	<b>Normal:</b> 9028 (24.2%) <b>CD:</b> 5486 (19.7%) <b>MI:</b> 5250 (18.9%) <b>HTF:</b> 4007 (17.8%) <b>STTC:</b> 2055 (9.5%)
PTB-XL	ECG	10 seconds	n/a	

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 number of categories in the dataset (Table taken by Kendre, 2021).

# Model Development

We propose using Generative Adversarial Networks (GANs) along with Convolutional Neural Network (CNNs) and Transformers 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 attention mechanisms that give 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-trained, to compress the PCG signals into a latent space, 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 optimized towards the PCG's latent space. This novel concept forces the Generators to produce equivalent latent spaces for ECG and PCG (given both signals were recorded simultaneously). After pre-training, the ECG CNN Encoder is fed ECG data from categorized arrhythmia datasets.

# Distribution Table

Dataset	Accuracy	F1 Score	Sensitivity	Specificity	PPV	NPV
PASCAL	73.7±0.02	75.0±0.04	75.0±0.05	72.2±0.04	75.0±0.02	72.2±0.04
PhysioNet 2016	89.5±0.03	64.4±0.05	48.0±0.06	85.7±0.02	96.5±0.04	85.7±0.03
Synthesized PTB-XL (AF)	83.0±0.03	94.3±0.03	99.5±0.03	90.3±0.03	99.3±0.03	91.8±0.03
Combined	<b>95.0±0.03</b>	<b>94.6±0.04</b>	<b>94.3±0.04</b>	<b>99.5±0.03</b>	<b>90.3±0.03</b>	<b>99.3±0.02</b>

Table 3: The table displays the metrics collected on the testing set for all classes in the dataset (Table taken by Kendre, 2021).

# Confusion Matrix

	Normal	Murmur	Gallop	AF
Normal	0.959	0.0	0.041	0.0
Murmur	0.029	0.964	0.007	0.0
Gallop	0.094	0.062	0.944	0.0
AF	0.0	0.0	0.05	0.0
Noisy	0.0	0.046	0.0	0.954

Figure 4: Matrix of accuracy between categories (pathologies) classified by the model (predicted label) and the true categories (true label) (Figure taken by Kendre, 2021).

# 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-time heart sound analysis.

Figure 6: Time complexity of the PCG/ECG Encoder and Transformer Discriminator in classifying heart sounds (Figure taken by Kendre, 2021).

# Cross-Validation

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 70% of training, 10% of validation, and 20% of testing.

# Model Learning

**Mean Squared Error:**  
 $MSE(x, y) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$   
**Cross Entropy Loss:**  
 $CE(x, class) = -\log(\frac{\exp(x[class])}{\sum_{class} \exp(x[class])}) = -x[class] + \log(\sum_{class} \exp(x[class]))$   
**Accuracy:**  
 $ACC = \frac{TP + TN}{TP + TN + FP + FN}$   
**Sensitivity:**  
 $TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR$   
**Specificity:**  
 $TNR = \frac{TN}{N} = \frac{TN}{TN + FP} = 1 - FPR$   
**Positive Predictive Value:**  
 $PPV = \frac{TP}{TP + FP} = 1 - FDR$   
**Negative Predictive Value:**  
 $NPV = \frac{TN}{TN + FN} = 1 - FOR$   
**F1 Score:**  
 $F_1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$   
**Time Complexity:**  
 $O(n) = \frac{1}{n} \sum_{i=1}^n m(\log(n) - n)_i$

**Real-World Testing & Model Deployment** The model's viability is crucial for ensuring the model's success in the real world. Thus, we need to conduct trials with the end-to-end system to ensure we can deploy and integrate the model with ease.

Table 2: Metrics shows in the table are used to evaluate the model on the testing set to ensure generalization and comparability (Table taken by Kendre, 2021).

# ROC/AUC

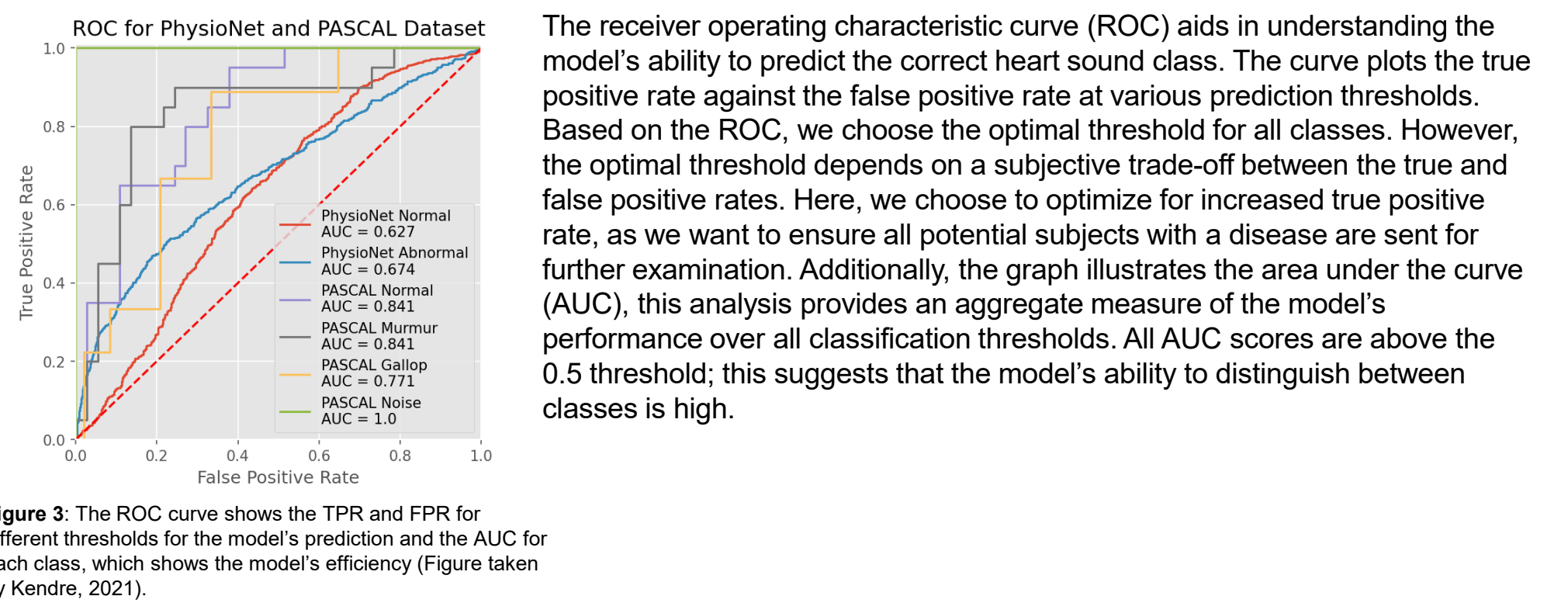


Figure 3: The ROC curve shows the TPR and FPR for different thresholds for the model's prediction and the AUC for each class, which shows the model's efficiency (Figure taken by Kendre, 2021).

# 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 (Figure taken by 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 the potential to record the sounds at a high resolution, given the microphone is located at 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 (Figure taken by Kendre, 2021).

# Model Interpretation

Figure 8 visualizes the channels of the Convolutional Encoder layers which are responsible for extracting features from heart 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 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 parses 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 (Figure taken by Kendre, 2021).

# Comparative System Evaluation

Device	Frequency Range (Hz)	Sample Rate (Hz)	Amplification	Cardiac Landmark Guide	Cost
3M Littman 3200	20 – 2000	8000	Up to 24x	No	\$499
Eko Core	20 – 2000	4000	Up to 40x	No	\$349
Jabes	20 – 1000	8000	Up to 20x	No	\$229
Smartphone	20 – 2000	16000-48000	Up to 40x	Yes	n/a

Table 4: The table displays different devices that record heart sounds digitally with their respective features (Table taken by Kendre, 2021).

# Comparative Model Evaluation

Comparing different methods, we observe the 85.7% was the best accuracy reached on the PhysioNet dataset and 61.1% was the best accuracy reached on the PASCAL dataset. Conversely, our proposed method achieved 89.5% accuracy on the PhysioNet dataset and 73.7% accuracy on the PASCAL dataset. The increase in performance is attributed to our semi-supervised approach, where we used adversarial Conditional Generators to generate heart sounds. This exposed the Discriminator to a wide range of heart sounds which assisted the model in optimizing for generalized features. Additionally, we conducted a statistical significance test (t-test) to show the probability our proposed method's results are due to random chance. The test concluded the results were statistically significant as all p-values were less than 0.05. This implies that the null hypothesis can be rejected, and the results are statistically significant.

Study	Classification techniques	Beat types	Dataset	Time Efficiency	Results
Nogueira et al.	SVM	N & A	PhysioNet 2016	-	Sensitivity: 96.47% Specificity: 72.65% Overall Score: 84.56%
Krishnan et al.	DNN	N & A	PhysioNet 2016	-	Sensitivity: 86.73% Specificity: 84.75% Accuracy: 85.65%
Rubin et al.	CNN	N & A	PhysioNet 2016	-	Specificity: 98% Sensitivity: 73% Overall Score: 84%
Singh et al.	Bayes Net and Logit Boost	N & M	PASCAL	-	Specificity: 46.9% Sensitivity: 69.73% Accuracy: 61.1%
	N, M, G, No	PASCAL			Sensitivity: 75.0.3% Specificity: 72.2% Accuracy: 73.7.9%
Proposed Method	GAN, CNN, Transformer	N & A	PhysioNet 2016	~1000 cps	Sensitivity: 48.0% Specificity: 85.7% Accuracy: 89.5%
	N, M, G, AF, No	Combined			Sensitivity: 94.3% Specificity: 99.5% Accuracy: 95.0%

Table 5: The table displays studies with their proposed classification techniques, beat types, types of heart sounds, dataset, time efficiency, and results (Normal, Abnormal, Murmur, Gallop, Noisy) (Table taken by Kendre, 2021).

# Conclusion

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 B. Rather, the model can be trained to convert language A to an intermediary language (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.

- Up to 20x times cheaper
- Provides 6 times better clarity
- Up to 34% better accuracy with a wider range of pathologies
- 10 times faster in diagnosis

# Future Applications

- Deploying the model with an app that is available to 3rd world countries that can't afford to conduct in-depth testing regularly
  - Integrate and serve the model using FastAPI
- Use abnormal heart sound unsupervised datasets as a basis of categorical arrhythmia classification
  - Using low dimensional visualization techniques like t-SNE or UMAP
  - Cluster data using methods like K means and hierarchical clustering
- Create a classifiable latent representation of PCG signal biomarkers that can be represented with accuracy and precision
  - Created by VAE that are fed the PCG signals directly instead of a spectrogram
- Investigate training a heart sound discriminator from generated PCG data from ECG datasets
- Reconstructing speech (wav) recordings from the human auditory cortex (EEG) using techniques used for PCG construction



## References

- [1] 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.
- [2] Prakash D, Uma Mageshwari T, Prabakaran K, and Suguna A. *Detection of Heart Diseases by Mathematical Artificial Intelligence Algorithm Using Phonocardiogram Signals*.
- [3] PhysioNet/CinC Challenge 2016: Training Sets.
- [4] Nabina N Rawther and Jini Cheriyan. Detection and Classification of Cardiac Arrhythmias based on ECG and PCG using Temporal and Wavelet features. 4(4):6, 2015.
- [5] Shahid Ismail, Imran Siddiqi, and Usman Akram. Localization and classification of heart beats in phonocardiography signals —a comprehensive review. *EURASIP Journal on Advances in Signal Processing*, 2018(1):26, December 2018.
- [6] Omer Deperlioglu, Utku Kose, Deepak Gupta, Ashish Khanna, and Arun Kumar Sangaiah. Diagnosis of heart diseases by a secure Internet of Health Things system based on Autoencoder Deep Neural Network. *Computer Communications*, 162:31–50, October 2020.
- [7] Dinesh Surukutla, Karan Bhanushali, and Trupti Patil. Cardiac Arrhythmia Detection Using CNN. *SSRN Electronic Journal*, 2020.
- [8] Abdulhamit Subasi. Biomedical Signals. In *Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques*, pages 27–87. Elsevier, 2019.
- [9] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning Deep Features for Discriminative Localization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2921–2929, Las Vegas, NV, USA, June 2016. IEEE.
- [10] Palani Thanaraj Krishnan, Parvathavarthini Balasubramanian, and Snehalatha Umapathy. Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network. *Physical and Engineering Sciences in Medicine*, 43(2):505–515, June 2020.
- [11] Abhishek Das, Harsh Agrawal, C. Lawrence Zitnick, Devi Parikh, and Dhruv Batra. Human Attention in Visual Question Answering: Do Humans and Deep Networks Look at the Same Regions? *arXiv:1606.03556 [cs]*, June 2016. arXiv: 1606.03556.
- [12] Pranav Rajpurkar, Awni Y. Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y. Ng. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. *arXiv:1707.01836 [cs]*, July 2017. arXiv: 1707.01836.
- [13] Tharindu Fernando, Houman Ghaemmaghami, Simon Denman, Sridha Sridharan, Nayyar Hussain, and Clinton Fookes. Heart Sound Segmentation using Bidirectional LSTMs with Attention. *IEEE Journal of Biomedical and Health Informatics*, 24(6):1601–1609, June 2020. arXiv: 2004.03712.
- [14] A. Almasi, M. B. Shamsollahi, and L. Senhadji. A dynamical model for generating synthetic Phonocardiogram signals. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5686–5689, Boston, MA, August 2011. IEEE.
- [15] V. Kalaivani. DIAGNOSIS OF ARRHYTHMIA DISEASES USING HEART SOUNDS AND ECG SIGNALS. *Russian Journal of Cardiology*, (1-ENG):35–41, January 2014.
- [16] Chris Olah, Alexander Mordvintsev, and Ludwig Schubert. Feature Visualization. *Distill*, 2(11):10.23915/distill.00007, November 2017.
- [17] M Finlay. The "mobile-phonocardiogram", a new tool in the arrhythmia clinic. *Heart*, 92(7):898–898, May 2006.
- [18] Varsha Garg, Arpit Mathur, Nishant Mangla, and Aman Singh Rawat. Heart Rhythm Abnormality Detection from PCG Signal. In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pages 1–5, Noida, India, August 2019. IEEE.
- [19] L. Jordaens. A clinical approach to arrhythmias revisited in 2018: From ECG over noninvasive and invasive electrophysiology to advanced imaging. *Netherlands Heart Journal*, 26(4):182–189, April 2018.
- [20] Serkan Kiranyaz, Morteza Zabihi, Ali Bahrami Rad, Turker Ince, Ridha Hamila, and Moncef Gabbouj. Real-time phonocardiogram anomaly detection by adaptive 1D Convolutional Neural Networks. *Neurocomputing*, 411:291–301, October 2020.

- [21] P. Lubaib and K.V. Ahammed Muneer. The Heart Defect Analysis Based on PCG Signals Using Pattern Recognition Techniques. *Procedia Technology*, 24:1024–1031, 2016.
- [22] Lars-Jochen Thoms, Giuseppe Collichia, and Raimund Girwidz. Real-life physics: phonocardiography, electrocardiography, and audiometry with a smartphone. *Journal of Physics: Conference Series*, 1223:012007, May 2019.
- [23] Preecha Yupapin, Wardkein, Preecha Yupapin, Phanphaisarn, Koseeyaporn, Roeksabutr, Roeksabutr, Wardkein, and Koseeyaporn. Heart detection and diagnosis based on ECG and EPCG relationships. *Medical Devices: Evidence and Research*, page 133, August 2011.
- [24] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. *arXiv:1311.2901 [cs]*, November 2013. arXiv: 1311.2901.
- [25] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WaveNet: A Generative Model for Raw Audio. *arXiv:1609.03499 [cs]*, September 2016. arXiv: 1609.03499.
- [26] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. SmoothGrad: removing noise by adding noise. *arXiv:1706.03825 [cs, stat]*, June 2017. arXiv: 1706.03825.
- [27] Siddique Latif, Muhammad Usman, Rajib Rana, and Junaid Qadir. Phonocardiographic Sensing using Deep Learning for Abnormal Heartbeat Detection. *arXiv:1801.08322 [cs]*, July 2020. arXiv: 1801.08322.
- [28] Stephanie Ger and Diego Klabjan. Autoencoders and Generative Adversarial Networks for Imbalanced Sequence Classification. *arXiv:1901.02514 [cs, stat]*, August 2020. arXiv: 1901.02514.
- [29] Chaoqiang Zhao, Qiyu Sun, Chongzhen Zhang, Yang Tang, and Feng Qian. Monocular Depth Estimation Based On Deep Learning: An Overview. *Science China Technological Sciences*, 63(9):1612–1627, September 2020. arXiv: 2003.06620.
- [30] Amy T. Dao. Wireless laptop-based phonocardiograph and diagnosis. *PeerJ*, 3:e1178, August 2015.
- [31] Sumair Aziz, Muhammad Umar Khan, Majed Alhaisoni, Tallha Akram, and Muhammad Altaf. Phonocardiogram Signal Processing for Automatic Diagnosis of Congenital Heart Disorders through Fusion of Temporal and Cepstral Features. *Sensors*, 20(13):3790, July 2020.
- [32] Md. Khayrul Bashar, Samarendra Dandapat, and Itsuo Kumazawa. Heart Abnormality Classification Using Phonocardiogram (PCG) Signals. In *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 336–340, Sarawak, Malaysia, December 2018. IEEE.
- [33] Cristhian Potes, Saman Parvaneh, Asif Rahman, and Bryan Conroy. Ensemble of Feature-based and Deep learning-based Classifiers for Detection of Abnormal Heart Sounds. September 2016.
- [34] S1 and S2 Heart Sound Recognition Using Deep Neural Networks. *IEEE Transactions on Biomedical Engineering*, 64(2):372–380, February 2017.
- [35] Grzegorz Redlarski, Dawid Gradolewski, and Aleksander Palkowski. A System for Heart Sounds Classification. *PLoS ONE*, 9(11):e112673, November 2014.
- [36] Chengyu Liu, David Springer, Qiao Li, Benjamin Moody, Ricardo Abad Juan, Francisco J Chorro, Francisco Castells, José Millet Roig, Ikaro Silva, Alistair E W Johnson, Zeeshan Syed, Samuel E Schmidt, Chrysa D Papadaniil, Leontios Hadjileontiadis, Hosein Naseri, Ali Moukadem, Alain Dieterlen, Christian Brandt, Hong Tang, Maryam Samieinasab, Mohammad Reza Samieinasab, Reza Sameni, Roger G Mark, and Gari D Clifford. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, 37(12):2181–2213, December 2016.
- [37] Ye Jia, Ron J. Weiss, Fadi Biadisy, Wolfgang Macherey, Melvin Johnson, Zhifeng Chen, and Yonghui Wu. Direct speech-to-speech translation with a sequence-to-sequence model. *arXiv:1904.06037 [cs, eess]*, June 2019. arXiv: 1904.06037.
- [38] Mohammed Nabih Ali, EL-Sayed A. El-Dahshan, and Ashraf H. Yahia. Denoising of Heart Sound Signals Using Discrete Wavelet Transform. *Circuits, Systems, and Signal Processing*, 36(11):4482–4497, November 2017.
- [39] K. Ajay Babu, Barathram Ramkumar, and M. Sabarimalai Manikandan. S1 and S2 heart sound segmentation using variational mode decomposition. In *TENCON 2017 - 2017 IEEE Region 10 Conference*, pages 1629–1634, Penang, November 2017. IEEE.
- [40] Wenjie Zhang, Jiqing Han, and Shiwen Deng. Heart sound classification based on scaled spectrogram and partial least squares regression. *Biomedical Signal Processing and Control*, 32:20–28, February 2017.

- [41] M. Abo-Zahhad, Mohammed Farrag, Sherif N. Abbas, and Sabah M. Ahmed. A comparative approach between cepstral features for human authentication using heart sounds. *Signal, Image and Video Processing*, 10(5):843–851, July 2016.
- [42] Elmar Messner, Matthias Zohrer, and Franz Pernkopf. Heart Sound Segmentation—An Event Detection Approach Using Deep Recurrent Neural Networks. *IEEE Transactions on Biomedical Engineering*, 65(9):1964–1974, September 2018.
- [43] Patrick Esser, Robin Rombach, and Björn Ommer. Taming Transformers for High-Resolution Image Synthesis. *arXiv:2012.09841 [cs]*, February 2021. arXiv: 2012.09841.
- [44] Photo Posters - Create Custom Photo Posters | Walgreens Photo.
- [45] Ye Jia, Ron J. Weiss, Fadi Biadsy, Wolfgang Macherey, Melvin Johnson, Zhifeng Chen, and Yonghui Wu. Direct speech-to-speech translation with a sequence-to-sequence model. *arXiv:1904.06037 [cs, eess]*, June 2019. arXiv: 1904.06037.