Detection of Cardiac Electrical Instability Prior to Cardiac Arrest

Aneel Damaraju, Chiraag Kaushik, Andrew Pham, Kunal Rai, Tucker Reinhardt, Frank Yang

Mentors: Sebastian Acosta, PhD; Mubbasheer Ahmed, MD; Parag Jain, MD

Introduction

Congenital heart defects are one of the leading causes of birth defect related infant death (Oster et al. 2013). Infants affected by congenital heart defects must often spend time in intensive care units prior to or between treatments. While in intensive care units, these patients may experience unexpected cardiopulmonary arrests. These events are characterized by the breakdown of the heart's electrical function, leading to decreased flow of oxygenated blood and possible death. According to the American Heart Association, the survival rate for children in-hospital cardiac arrests is 36.8% ("AHA Cardiac Arrest Statistics" n.d.). In the event of cardiopulmonary arrest, prompt and decisive action by physicians is crucial for patient survival.

Physiological abnormalities in patients often manifest prior to these life-threatening events. Timely detection of these abnormalities offers the possibility of preventative treatment. Such proactive treatments circumvent the irreversible heart tissue damage that frequently occurs during cardiac arrest. This problem is challenging because these abnormalities can be imperceptible to the human eye, making them unavailable to caretakers. **In this project, we seek to develop an algorithm for early detection of electrical instability in the heart.** We examine the electrocardiogram (ECG) and more specifically, seek to quantify the variability of ECG signal prior to cardiac arrest.

Background

Hypoplastic Left Heart Syndrome:

In this project, we focus on infants born with hypoplastic left heart syndrome (HLHS), a condition in which the left side of the heart is underdeveloped. According to the Center for Disease Control and Prevention (CDC), 1,025 babies are born with HLHS annually (Mai et al. 2019). A comparison of normal and HLHS hearts is given in figure 1. In a healthy baby, the right side of the heart pumps oxygen-poor blood to the lungs, where it is re-oxygenated. The left side of the heart then pumps re-oxygenated blood to the body. In the case of HLHS, the left side of the heart cannot effectively pump this oxygen-rich blood to the remainder of the body. This can result in cardiac ischemia, a condition in which cardiac tissue is not properly supplied with blood. Cardiac ischemia results in instability of the heart's electrical conduction and eventual heart failure (McCallister et al. 1979; Shaw and Rudy. 1997).

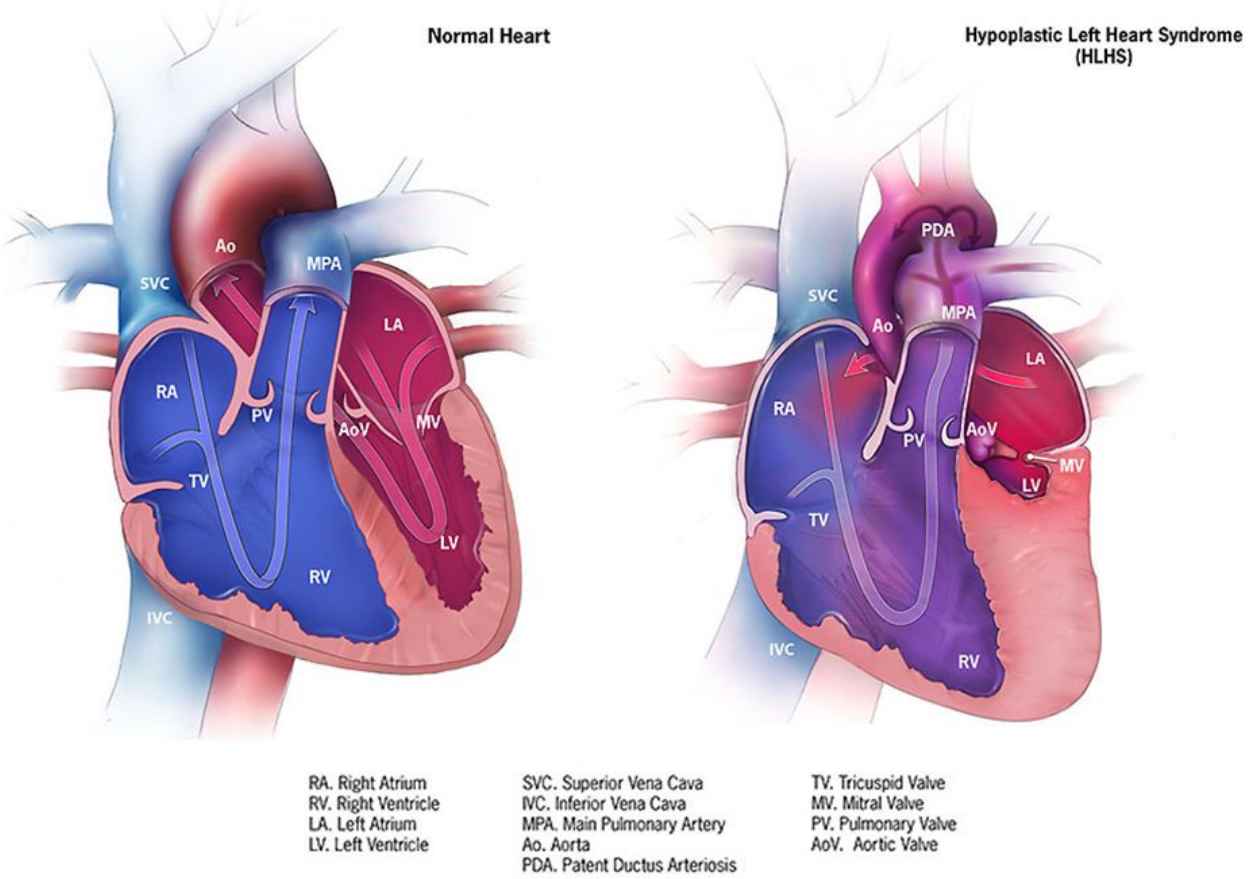

Figure 1: Comparison of normal heart (left) and HLHS heart (right). (www.cdc.gov).

Cardiac ischemia creates electrical instability in the electrocardiogram (ECG). This instability has been observed in various components of the ECG waveform, including prolonged or shortened PR intervals and QRS complexes and ST segment variability [Clifford et al. 2006; Vu et al. 2017]. Figure 2 depicts a sketch of a 1-lead ECG waveform, with various components labeled. **Notably, detection of ECG instability is not possible with conventional bedside monitoring.Therefore, development of an algorithm and system to detect ECG volatility has life-saving potential for patients.**

Approaching the electrocardiogram:

We seek to apply signal processing and machine learning techniques to detect electrocardiogram instability or arrhythmia. An arrhythmia is any abnormal cardiac rhythm. Feature extraction of the ECG waveform is a well studied topic. **Instead of focusing on cardiac signal feature extraction, we will approach the ECG signal holistically to quantify variability.** This work will yield an algorithm to detect the onset of ischemic instability, buying valuable time for caretakers to implement life-saving measures.

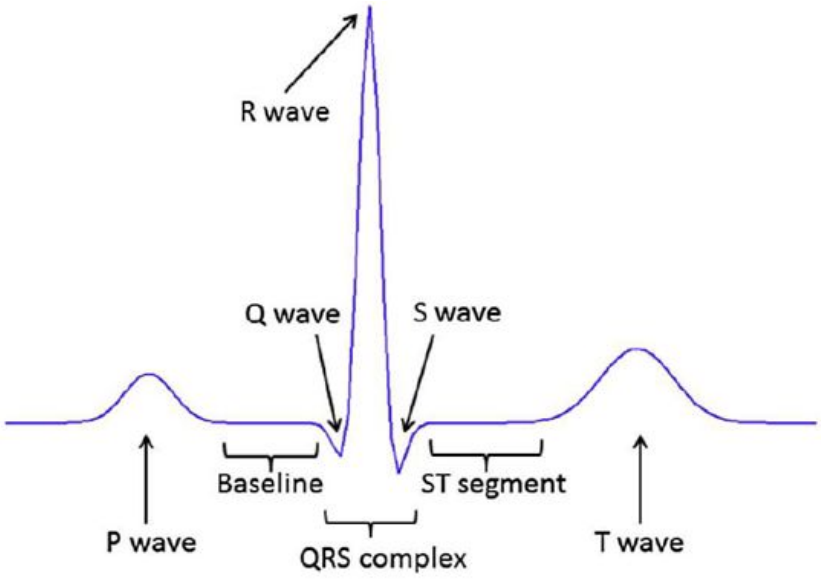

Figure 2: sketch of 1-lead ECG waveform, with components labelled (courtesy of Texas Children's Hospital).

Prior Work

Detection of cardiac arrhythmias to predict the likelihood of cardiac arrests has been explored by many publications. Most of the prior work in this field focuses on feature extraction of single heartbeats to determine if the beat is abnormal or not (Li and Boulanger 2020). This feature extraction looks at the different segments (PR, QRS, PP, QT, etc) of a heartbeat and determines inconsistencies in these segment intervals. For heartbeat classification, the shape of the heartbeat and waves present in a heartbeat are used to detect anomalies for diagnosis. Within feature extraction of heartbeats, there are two common approaches: extraction of morphological features and derived features. Morphological features are created through the observation of signals. Time intervals between Q and S waves, slope velocity of segments, valley and peak values, and areas for segments of the heartbeat are all used as morphological features. For derived features, vector cardiograms, discrete wavelet transforms, PCA components, eigenvectors, and other dimensionality reduction techniques are used. Most anomaly detection methods for heartbeats focus on using extracted features for modeling.

Deep learning models have replaced feature extraction based methods for heartbeat classification problems in recent years. Stacked long short-term memory (LSTM) networks have often been used in place of feature extraction to determine thresholds for irregular heartbeats (Chauhan and Vig 2015). Further, LSTM networks have been shown to be able to identify different types of anomalies which is important when the signals are being inspected holistically rather than segment/feature-wise. The Chauhan and Vig paper uses the probability distribution of the LSTM network error to provide a determination of whether a heartbeat is anomalous or not. LSTM networks have also been implemented for continuous classification of heartbeats in

online settings due to low computation overhead on personal wearable devices (Saadatnejad, Oveisi, and Hashemi 2020).

Another neural network based approach for heartbeat anomaly detection is autoencoders(Putra et al. 2019). Putra et al. used autoencoder network structures for feature extraction for heartbeat classification. Autoencoder networks are often used for dimensionality reduction with higher information retention than classical methods. Kieu et al. used sparsely-connected recurrent neural network autoencoders for time series in an ensemble fashion, which was the first implementation for unsupervised tasks (Kieu et al. 2019).

The data used in this project comes from pediatric patients who have been through heart surgeries so their electrocardiogram data is more unstable than that of typical patients. This makes morphological feature extraction much more difficult, as peaks and segments are unable to be identified. Because of this difficulty, our project aims to study instability in the entire waveform, rather than in any specific component. In addition, this project intends to build on the conclusions presented by a TCH publication, which determined that ST-segment instability was associated with cardiopulmonary events (Vu et al. 2017).

Issues that prior work faced in modeling anomalies in ECG data included imbalanced data, unlabeled data, and signal noise. We will touch on how our project addresses these issues in future sections. This project will build on the developments of continuous classification and anomaly detection in ECG data. The projected contribution of this project is to focus on using the entire heartbeat to determine an index of volatility in real time, rather than utilizing morphological feature extraction to determine if a given heartbeat is an anomaly.

Objectives

1. Process and align the electrocardiogram waveforms in time.
2. Explore the data for patient-specific and patient-independent trends prior to cardiac arrests.
3. Develop an algorithm to quantify instability in the heart's electrical activity.
4. Build a classifier or anomaly detection method to detect instability changes and sound an alarm prior to a cardiac event.

Data Science Pipeline Design Plan

Data Wrangling

Structuring and Preparing Data

The first step in our pipeline is to transform the provided ECGs into individual heartbeats. This requires identifying the peaks in the ECGs and then partitioning the ECGs based on those identified peaks.

To further process the data, we need to transform the heartbeat signals to the same length. We chose to linearly interpolate the heartbeat vectors to a length of 100 based on the recommendation of Dr. Acosta. We stored the partitioned, interpolated signals as a single 3D array for each patient, with one dimension for the heartbeat index, one dimension for the lead number, and one dimension for the individual interpolated heartbeat data points.

Cleaning Data

We also need to remove obvious outliers from the provided ECGs during this step. Obvious outliers in this context are portions of the ECG that are much too long to be heartbeats. This may include points where the signal was so weak that the heartbeat detector could not detect a peak, or a portion of time where the leads were disconnected which would correspond to a flat signal.

Noise Removal

In prior work that has used ECG signals, denoising is implemented prior to modeling in order to allow for better detection of morphological features in the heartbeats. Since this project will be focused on inspection of heartbeats without feature extraction and with respect to the variability of the heartbeat, traditional noise removal has potential to remove variability. Dr. Acosta has recommended that noise removal for these patients be limited to a determination of whether a lead is producing a valid signal or not from the patient (i.e. lead is plugged in or set up correctly or not).

Preprocessing

During the data wrangling step, the heartbeats were interpolated to be 100-dimensional. However, preliminary data exploration reveals that the vast majority of the variance can be explained by 15 or fewer dimensions. We want to reduce the dimensionality of our data to reduce our model complexity, improve efficiency, and improve the ease of computing variability metrics for the data.

Dimensionality Reduction

We first applied Principle Component Analysis (PCA) to the individual leads of single patients to visualize where the variance lies in our data. To capture any potential nonlinearities in the data, we also decided to apply a variational autoencoder to reduce the dimensionality of our data.

To assess the effectiveness of our dimensionality reduction methods, we will first reduce the data from 100 to 10 dimensions using the specified technique. Then, we will reconstruct the data to the original 100 dimensions, and then compute the mean squared error between the original signal and the reconstructed signal. Whichever technique yields the lower mean squared error will be selected for further use.

Data Exploration

In the early stages of the pipeline, we need to visualize the initial and intermediate data in order to both validate our work and improve subsequent steps in the pipeline. Data exploration in the data wrangling section may include plotting small sections of time to observe individual heartbeats. In the preprocessing section, exploration includes plotting the eigenvalues of the data covariance matrix to see how much variance is captured in the first k dimensions. Furthermore, visualization of the data in a low-dimensional space (2 or 3 dimensions) may be used to inform and motivate the modeling phase of the pipeline.

Modeling

The model's input is 1 minute of ECG data. The model will output an *index of volatility* within 1 minute of receiving the input.

One potential idea for modeling is to use Gaussian Mixture Models, which can be used for density estimation and clustering of the heartbeats. The intuition behind this idea is that regular heartbeats would locally cluster following a normal distribution, and that abnormal heartbeats might be farther away from this normal cluster and possibly closer to a cluster of abnormal heartbeats. With the knowledge of these two distributions - the closest cluster as well as the distance from that cluster - we can calculate an index of volatility. Traditional measures of

variance, complexity and non-stationarity (such as entropy, autocorrelation, and variance) will be analyzed for trends over time.

Validation

We set aside five patients from the dataset to use as a testing data. However, we plan to acquire more ECGs from Dr. Acosta to use as additional testing and training data.
Once we finalize our model, we will convert the ECGs of those 5 patients into "streams" of data, wherein we generate 1-minute time signals from this ECG data that we can feed into our model. Over the course of the 6 hour stream, we will plot the index of volatility as a function of time.

We will assess the model accuracy by observing the index of volatility over the 6 hour time period. We expect the index of volatility to be close to zero during the first 1-2 hours, then approach one in the half hour before the cardiac arrest (i.e. the 6 hour mark).

Description of Data

The dataset used in this project was obtained from the Sickbay database at Texas Children's Hospital (TCH). Before use, all data was de-identified to minimize the risk of exposure to health information which is protected by the HIPAA.

The specific physiological time series chosen for the project comes from 44 pediatric patients in the intensive care unit at TCH. Each of the 44 patients were born with hypoplastic left heart syndrome and were operated on shortly after birth. The 44 patients used in this study each experienced at least one cardiac arrest, which protects our project from having to deal with imbalanced data. The data provided for each of the 44 patients are as follows:

- The **electrocardiogram (ECG)** signal measured at each of 4 different leads on the patient's body. For each patient, six hours of ECG data leading up to a cardiac arrest are given. Each ECG signal is measured at a sampling rate of 240 Hz. The provided data is not labeled, but the end time of each ECG time series (after the 6 hours) corresponds to the onset of a cardiac arrest.
- The **heart rate** measured over the same six-hour period as the above ECG data in beats per minute. The data was recorded at a sampling rate of 2 Hz. However, before given to our team, the data was resampled using linear interpolation to the 240 Hz sampling rate of the ECG signals. This data was not available for all of the 44 patients (the missing data is characterized below).
- A signal of **timestamps** (in seconds) which is aligned with the 240 Hz sampling rate of the ECG signals.

In total, our dataset consists of over 264 hours of ECG data at each of the 4 leads (for a total of over 1056 hours of ECG data), and heart rate information corresponding to this entire

duration. If needed, our sponsor may provide data from additional patients (or other features such as ST segment elevations and premature ventricular contractions) at a later date. A short sample of the raw ECG data is shown below:

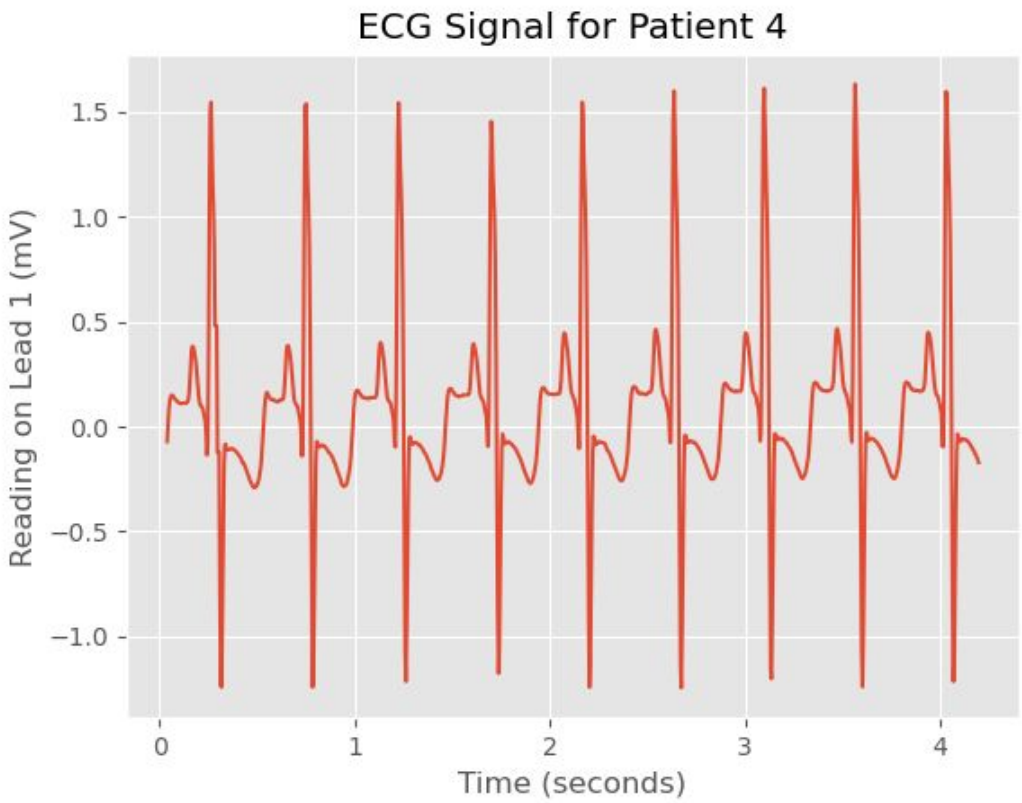

Figure 3 : Short segment of the provided ECG data for patient 4

There is no missing data in the provided ECG data for any patient, but there are some sections which display a flat signal, indicating that the leads were disconnected from the skin during measurement. Such segments will be removed in the data cleaning section of our pipeline. Moreover, there are 3 patients for whom no heart rate information is available, and the heart rate data for several other patients is missing for certain chunks of time. Information about the missing heart rate data is summarized below:

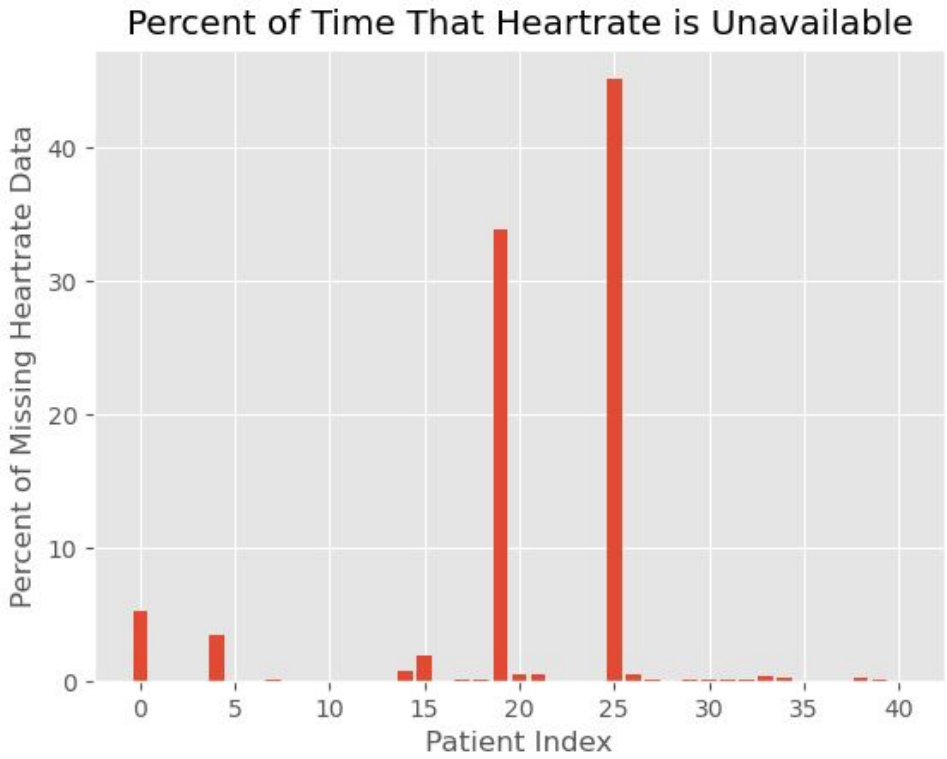

Figure 4 : Percentage of total time that heart rate data is not available

As shown in the figure above, for a few of the patients, the heart rate monitor readings are not available for certain sections of the entire 6 hour interval. For three of the patients, more than five percent of the total heart rate data is unavailable.

Preliminary Data Exploration

Raw Data Visualization

To understand the signals generated by the four-lead ECGs, we take small sections of the data as shown in Figure 3, and plot the signals on top of each other. Figure 5 shows the most common method for visualizing these signals, by overlaying the slice of all four signals onto a grid. By segmenting the data, we are able to visually compare the stability in the four leads at various points in time up until the cardiac arrest. From these plots, we can compare the peaks heights for different leads, as well the information contained in the signals for each lead.

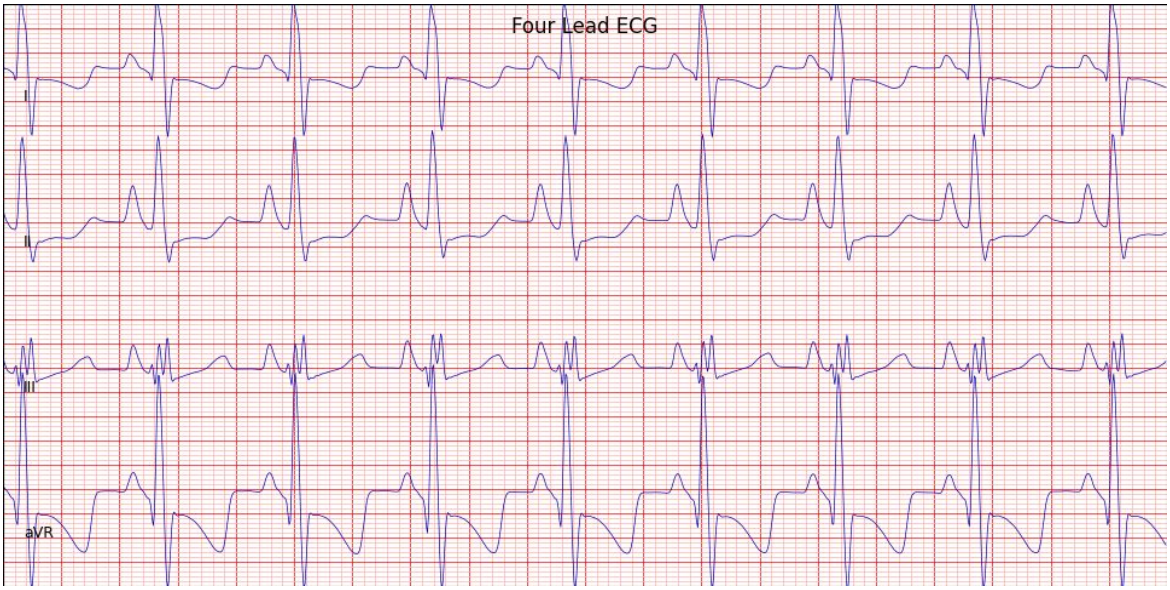

Figure 5: A time slice of the ECG signals from all four leads simultaneously

Peak Detection

During our preprocessing steps, each ECG signal is separated in time by heartbeat. For each patient there are tens of thousands of heartbeats, so we require summary statistics to judge both the heartbeats themselves, as well as the efficacy of our peak-finding algorithm. Both of the visualizations for this can be seen in Figure 5, which first explores the heartbeat lengths in the form of a histogram. We can see that the heartbeat lengths are almost normally distributed, with a slight skew noted by the increased variance on the right hand side of the mean. The right subplot in Figure 5 compares our algorithm's calculated heart rate to the heart rate recorded by a monitor on the patient as the ECG was being recorded, and speaks well for the accuracy of our algorithm.

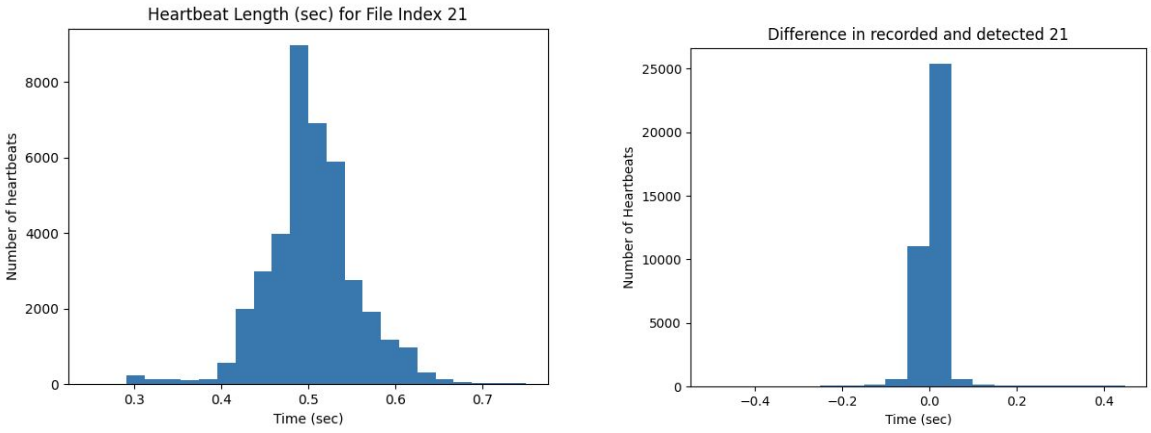

Figure 6: (Left) A histogram outlining the distribution of heartbeat lengths detected by our peak detection algorithm. (Right) A histogram comparing the length of heartbeats recorded and detected in the same patient.

Dimensionality Reduction

The final visualization technique used in exploring this data was reducing the dimension of the input data. By reducing the data dimensions to two, we can plot the data to determine trends along the data. One dimensionality reduction technique used here is principal component analysis. We use PCA to understand how much information is stored in the first two dimensions. As seen in Figure 6, around 90% of the information in each heartbeat can be captured by the first two components, which allows us to feel comfortable about using a two dimensional projection of our data for visualization purposes. Figure 7, shows the projection of the data using another common dimensionality reduction algorithm, UMAP. An interesting conclusion from UMAP is that as the samples increase in time, i.e. get closer to the time of the heart attack, they tend to move away from the cluster of points representing stable heart rates. This provides promising insight for the future steps of our project, as we can see that UMAP is able to find differences in the stable and unstable heartbeats.

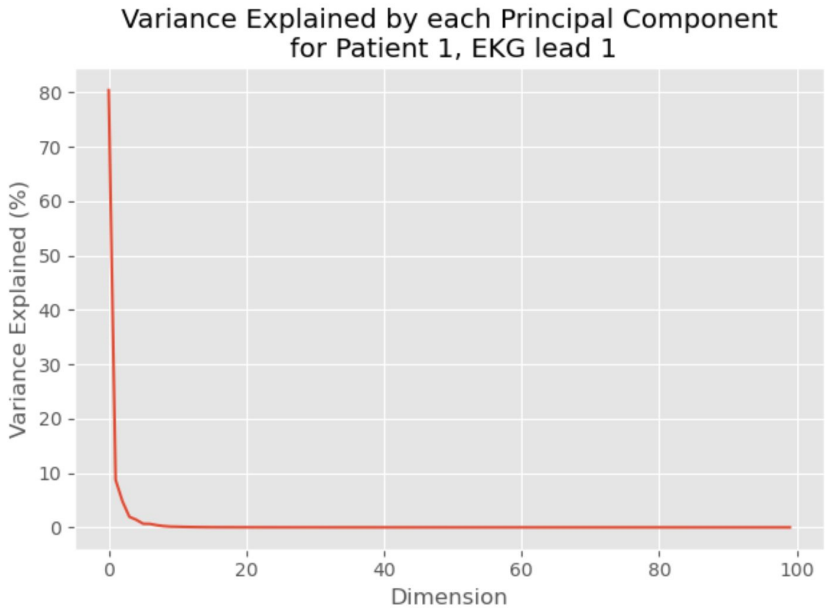

Figure 7: By performing PCA on all vectors for a given patient, we can determine how much information is stored within a low dimensional orthogonal projection of the data.

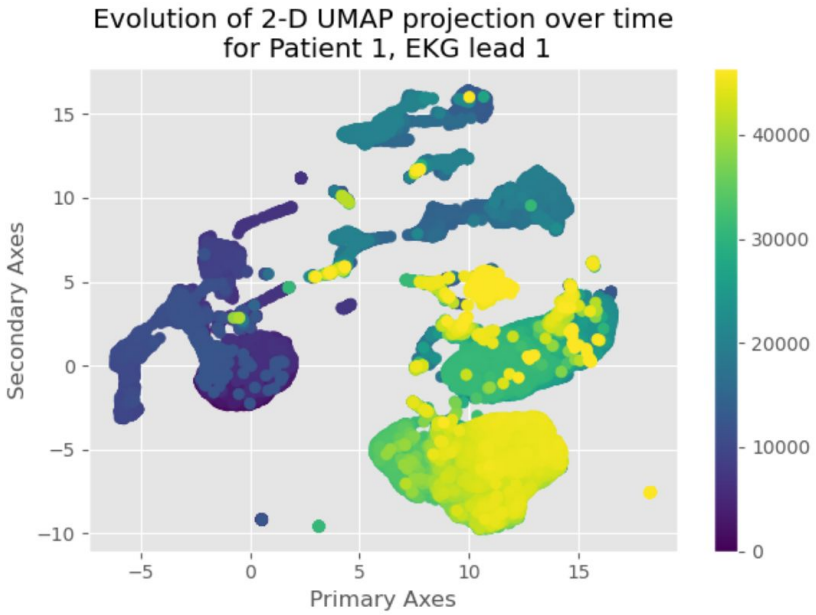

Figure 8: Using UMAP to visualize the data for a two dimensional projection of the heartbeats for a single patient. The colorbar shows what index the heartbeat plotted is, and how close the patient was to cardiac arrest.

Future Work

Future work involves employing different dimensionality reduction techniques such as auto-encoders to attempt to capture important variance information in the ECG. Further, we will apply machine learning models to quantify variance in an index of volatility. We seek to develop a computationally inexpensive algorithm to allow for real time analysis of ECG instabilities.

Citations

1. Oster, M. E., K. A. Lee, M. A. Honein, T. Riehle-Colarusso, M. Shin, and A. Correa. 2013. "Temporal Trends in Survival among Infants with Critical Congenital Heart Defects." *Pediatrics* 131 (5). https://doi.org/10.1542/peds.2012-3435.
2. "AHA Cardiac Arrest Statistics." n.d. Accessed September 20, 2020. https://cpr.heart.org/AHAECC/CPRAndECC/ResuscitationScience/UCM_477263_AHA-Cardiac-Arrest-Statistics.
3. Mai CT, Isenburg JL, Canfield MA, et al. for the National Birth Defects Prevention Network. National population-based estimates for major birth defects, 2010-2014. Birth Defects Res 2019; 1– 16. https://doi.org/10.1002/bdr2.1589.

4. McCallister LP, Liedtke AJ, Hughes HC. Ischemic injury to the conducting system of the heart. Involvement of myocardial lysosomes. *J Thorac Cardiovasc Surg*. 1979;77(5):647-661.
5. Shaw RM and Rudy Y. Electrophysiologic effects of acute myocardial ischemia. A mechanistic investigation of action potential conduction and conduction failure. *Circ Res 80*: 124–138, 1997.
6. https://www.cdc.gov/ncbddd/heartdefects/hlhs.html#:~:text=Hypoplastic%20left%20heart%20syndrome%20(HLHS,type%20of%20congenital%20heart%20defect
7. Clifford, Gari D., et al. *Advanced Methods and Tools for ECG Data Analysis*. Artech House, 2006.
8. Vu EL, Rusin CG, Penny DJ, et al. A Novel Electrocardiogram Algorithm Utilizing ST-Segment Instability for Detection of Cardiopulmonary Arrest in Single Ventricle Physiology: A Retrospective Study. *Pediatr Crit Care Med*. 2017;18(1):44-53.
9. Li, Hong Zu, and Pierre Boulanger. 2020. "A Survey of Heart Anomaly Detection Using Ambulatory Electrocardiogram (ECG)." *Sensors* 20 (5).
10. Chauhan, Sucheta, and Lovekesh Vig. 2015. "Anomaly Detection in ECG Time Signals via Deep Long Short-Term Memory Networks." *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. https://doi.org/10.1109/dsaa.2015.7344872.
11. Saadatnejad, Saeed, Mohammadhosein Oveisi, and Matin Hashemi. 2020. "LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices." *IEEE Journal of Biomedical and Health Informatics* 24 (2): 515–23.
12. Putra, Bayu Wijaya, Muhammad Fachrurrozi, M. Rudi Sanjaya, Firdaus, Anita Muliawati, Akhmad Noviar Satria Mukti, and Siti Nurmaini. 2019. "Abnormality Heartbeat Classification of ECG Signal Using Deep Neural Network and Autoencoder." *2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. https://doi.org/10.1109/icimcis48181.2019.8985206.
13. Kieu, Tung, Bin Yang, Chenjuan Guo, and Christian S. Jensen. 2019. "Outlier Detection for Time Series with Recurrent Autoencoder Ensembles." *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. https://doi.org/10.24963/ijcai.2019/378.
14. Vu, Eric L., Craig G. Rusin, Dan J. Penny, Kathy K. Kibler, Ronald Blaine Easley, Brendan Smith, Dean Andropoulos, and Ken Brady. 2017. "A Novel Electrocardiogram Algorithm Utilizing ST-Segment Instability for Detection of Cardiopulmonary Arrest in Single Ventricle Physiology: A Retrospective Study." *Pediatric Critical Care Medicine: A Journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies* 18 (1): 44–53.