Cardiac Electrocardiography Classification

2024 Data Science Challenge

Mentor: **Kerianne Pruett**
Creator: **Mikel Landajuela**

July 22, 2024

LLNL-PRES-XXXXXX
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Motivation

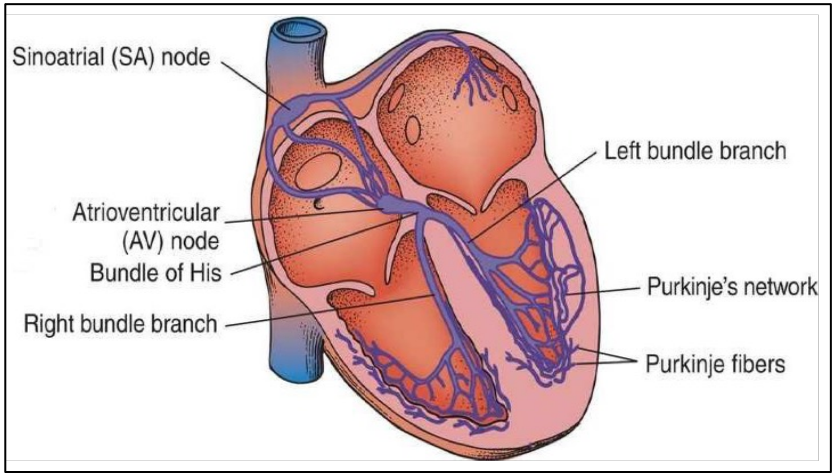
- Heart disease is the leading cause of death in the U.S.
- Every 34 seconds, one person in the U.S. dies from cardiovascular disease.
- Heart disease costs $229 billion every year in the U.S. alone.
- Death caused by abnormal electrical rhythms (arrhythmia).

Source: National Center for Chronic Disease Prevention and Health Promotion, Division for Heart Disease and Stroke Prevention

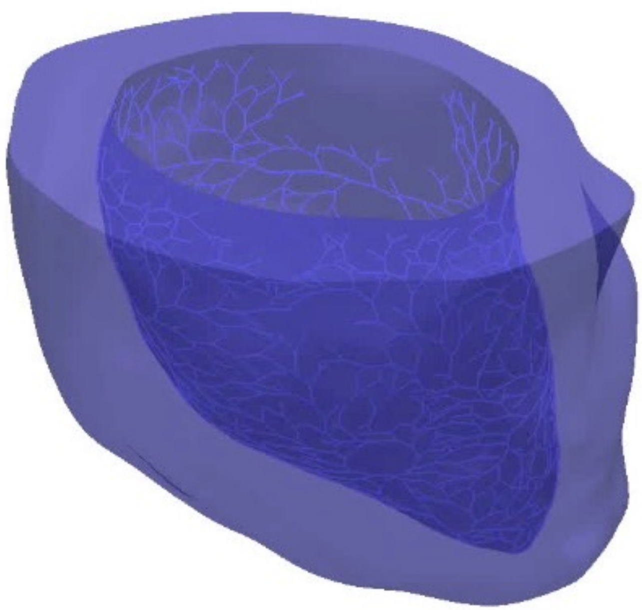
Background

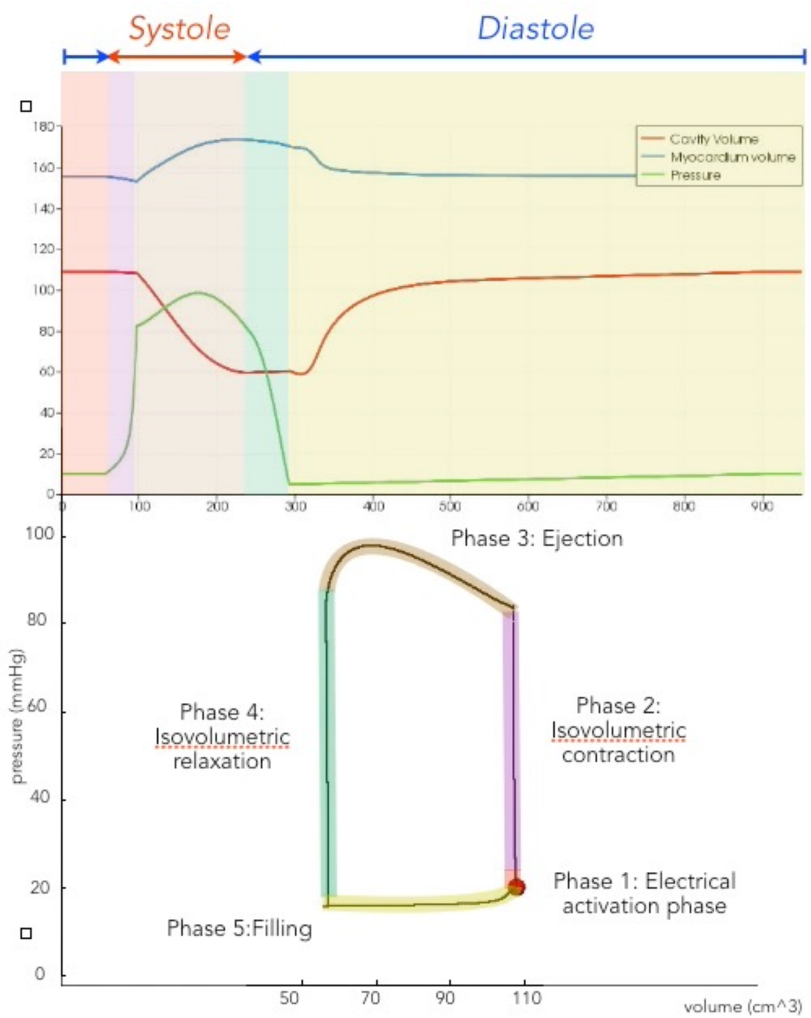
Cardiac Electrophysiology

- Electrical signals travel from the top to the bottom of the heart, causing contraction from the bottom to the top of the heart
  - Signal results from chemo-electro-mechanical coupling
- Cardiac cycle:
  - Electrical pulse originates from the SA node
    - The heart's natural pacemaker
  - Signal travels to AV node
    - Where it is delayed for ventricle filling
  - Pulse moves through specialized Purkinje Network
    - Facilitates rapid conduction of electrical signals
  - Enters the myocardium via the Purkinje-Muscle Junctions
    - Triggering the necessary reactions for mechanical contraction



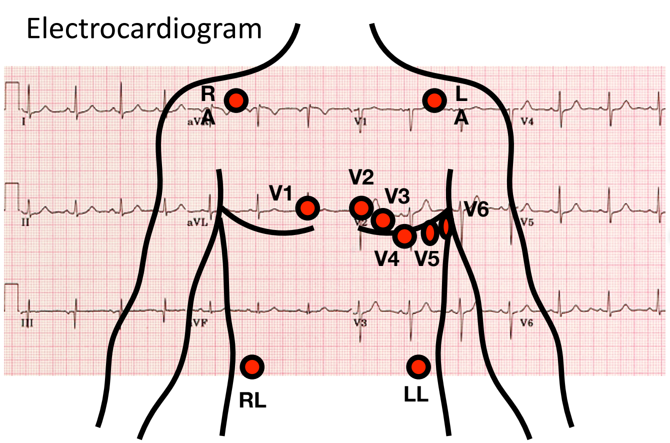
Cardiac cycle in the left ventricle

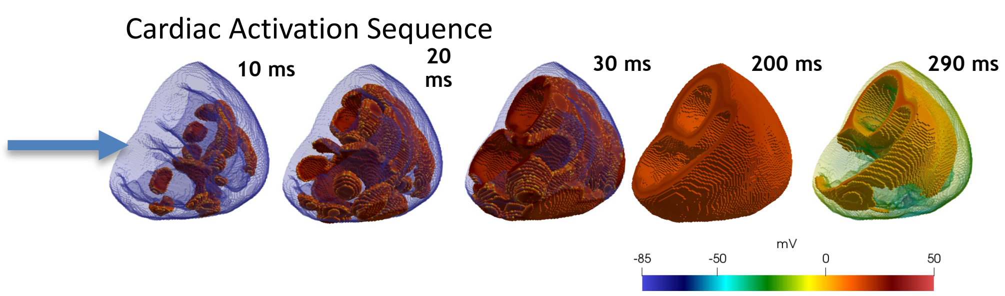




Electrocardiogram (ECG)

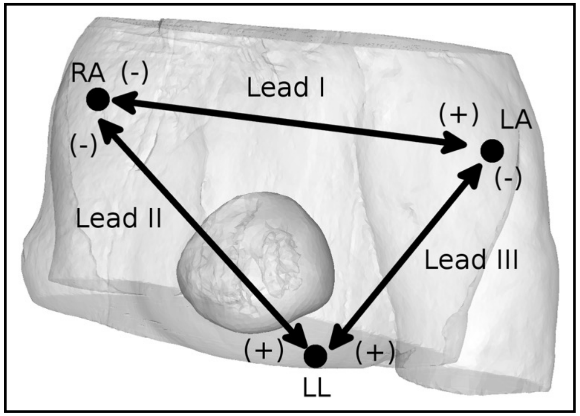
- ECGs play an essential role in diagnosing ventricular arrhythmias
- Non-invasive and most cost-effective detector
- ECGs can detect a wide variety of diseases
- Malfunctions in the cardiac cycle may be identified by an ECG





ECG Measurements

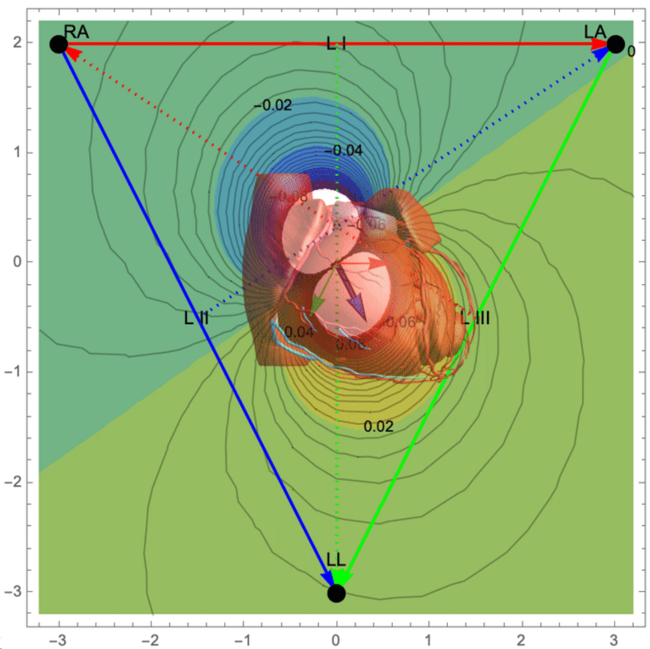
- As the electrical signal travels through the torso, it can be detected and recorded by an ECG
- The activation wave generates an external potential that propagates through the torso
- The ECG consists of measurements of that signal in different torso locations
  - Electrodes are metallic disks in contacts with the skin
  - Leads are measurements received from pairs of electrodes

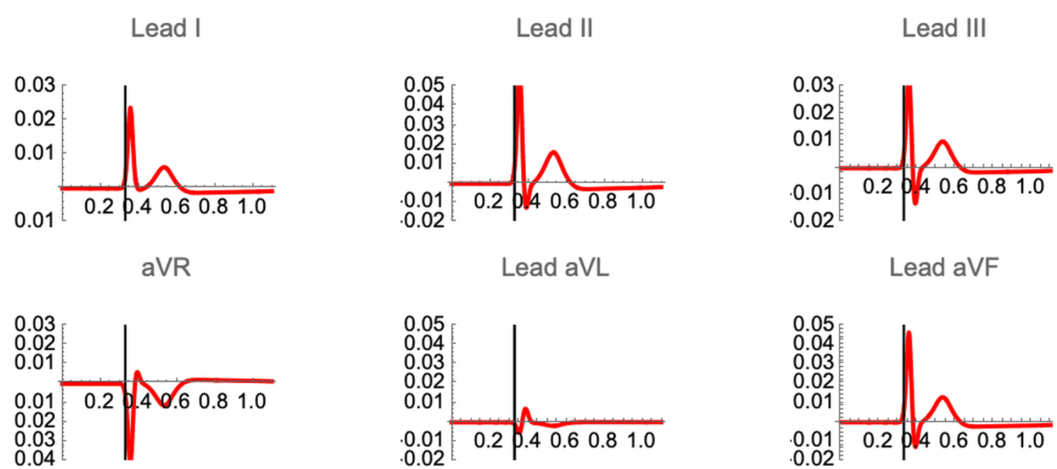


$$
\begin{cases}
X_{:,1} \rightarrow RA \\
X_{:,2} \rightarrow LA \\
X_{:,3} \rightarrow LL \\
X_{:,4} \rightarrow RL \\
X_{:,5} \rightarrow V1 \\
X_{:,6} \rightarrow V2 \\
X_{:,7} \rightarrow V3 \\
X_{:,8} \rightarrow V4 \\
X_{:,9} \rightarrow V5 \\
X_{:,10} \rightarrow V6
\end{cases}
\text{ and }
\begin{cases}
\text{Lead } I & : \; LA - RA \\
\text{Lead } II & : \; LL - RA \\
\text{Lead } III & : \; LL - LA \\
\text{Lead } aVR & : \; \frac{3}{2}(RA - Vw) \\
\text{Lead } aVL & : \; \frac{3}{2}(LA - Vw) \\
\text{Lead } aVF & : \; \frac{3}{2}(LL - Vw) \\
\text{Lead } V1 & : \; V1 - Vw \\
\text{Lead } V2 & : \; V2 - Vw \\
\text{Lead } V3 & : \; V3 - Vw \\
\text{Lead } V4 & : \; V4 - Vw \\
\text{Lead } V5 & : \; V5 - Vw \\
\text{Lead } V6 & : \; V6 - Vw
\end{cases}
$$

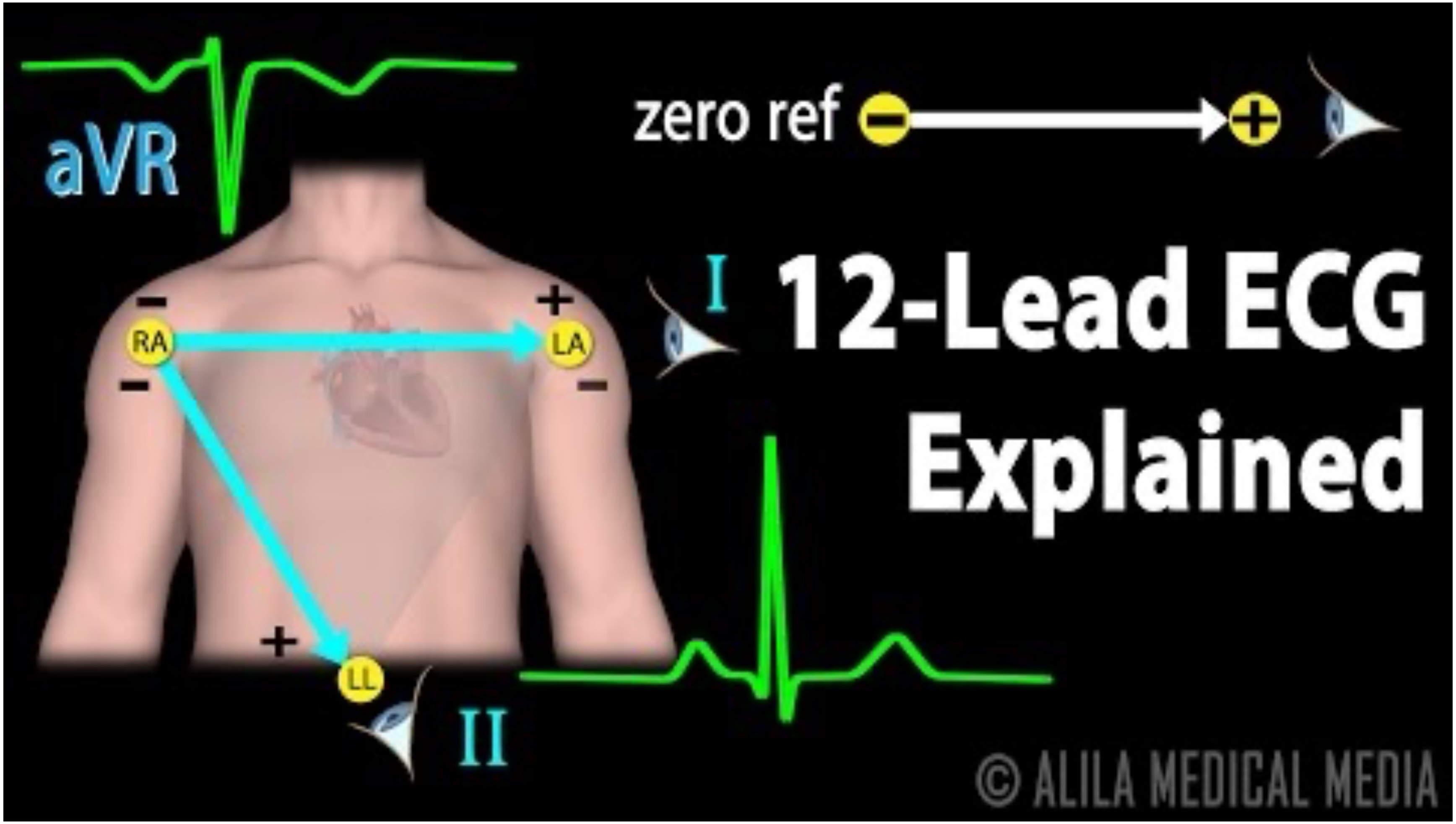
Cardiac signal through different leads

- Electrical phenomena inside heart causes dipole vector
  - Positive and negative pole rotating inside your chest
  - Generates electric potential that propagates through chest, organs, and skin
  - Detected by electrode on chest
  - 3D vector - 3 leads minimum to reconstruct vector

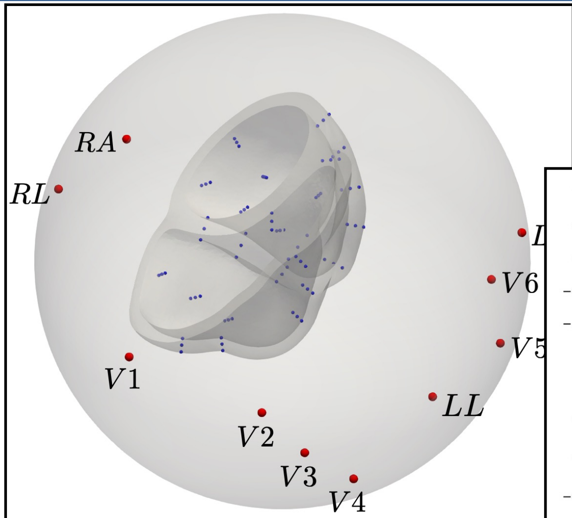


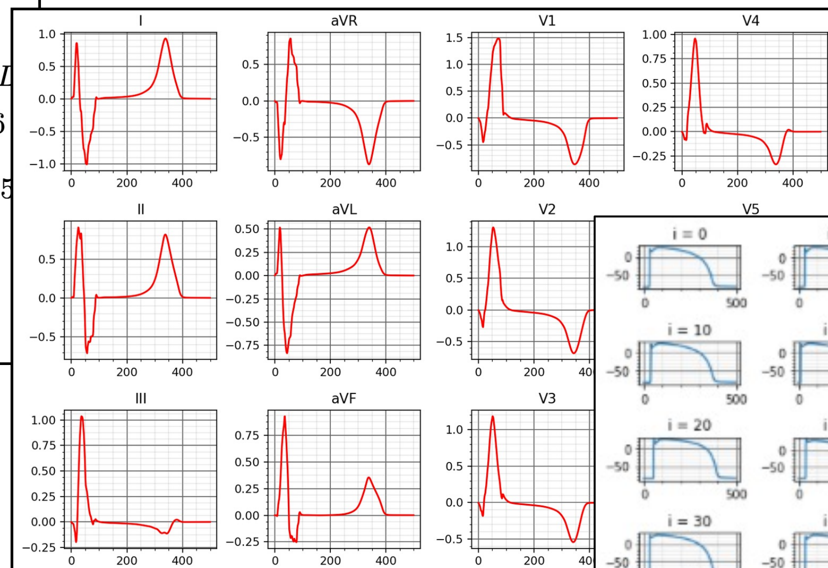


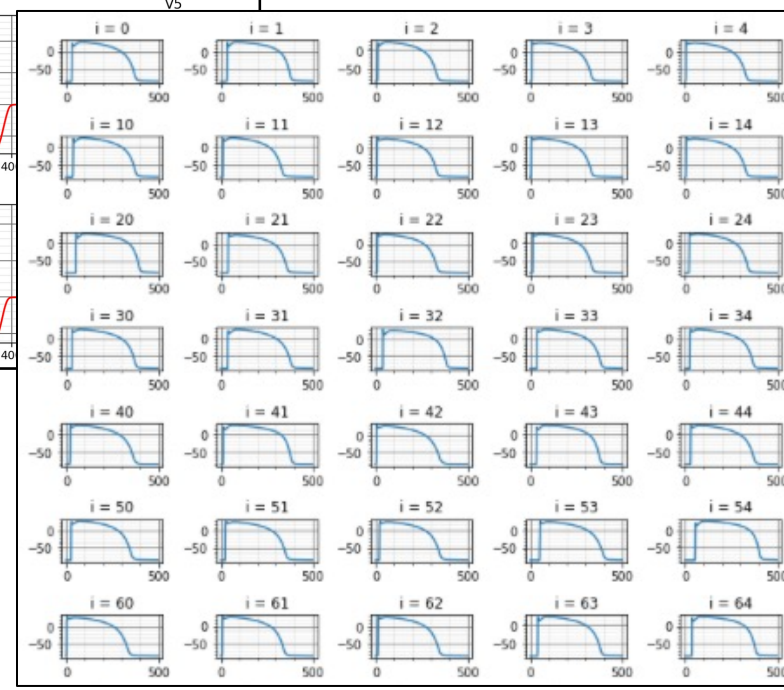
ECG Leads Explained



Transmembrane voltages for understanding activation map at various spots within the heart

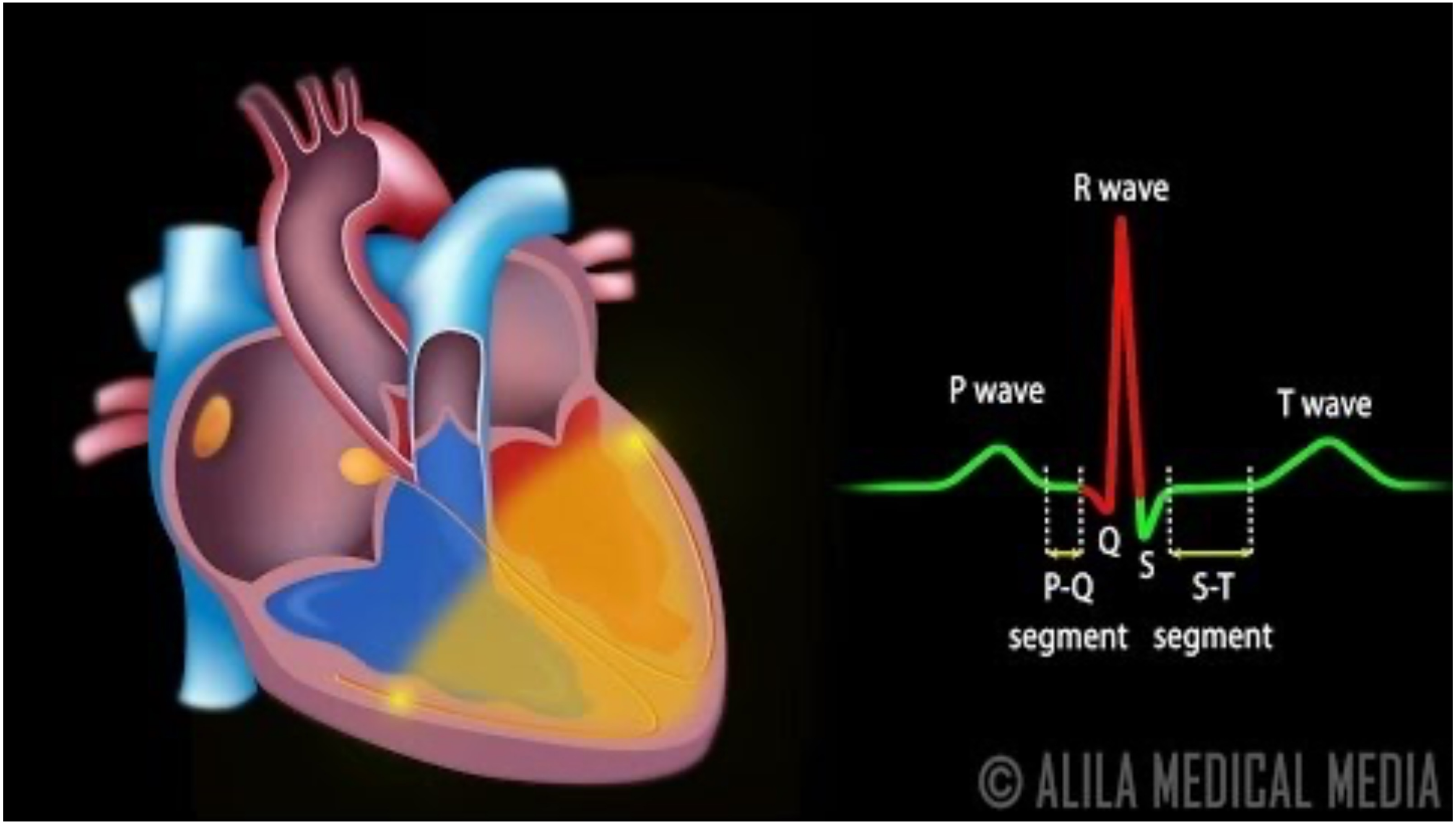






- Cardiac myocytes: special muscle cells in the heart
- Action potentials: contraction of myocytes initiated by electrical impulses

Heartbeat Segments



Heart math

- Mono-domain equations:
  - Reaction-diffusion type parabolic PDE

$$\begin{cases} \partial_t V + I_{ion}(\boldsymbol{w}, V) = \frac{1}{\chi C_m} \nabla \cdot (\boldsymbol{\sigma} \nabla V) + I_{app} & \text{in } \Omega \\ \partial_t \boldsymbol{w} = \boldsymbol{g}(\boldsymbol{w}, V) & \text{in } \Omega \end{cases}$$

Alternants and spiral breakup in a human ventricular tissue model - Ten Tusscher 2006 with endo-, mid- and epicardial cell heterogeneity

- Pseudo-ECG:
  - Averaging the potential propagation through the torso (Poisson problem)

Bioelectricity: a quantitative approach - Plonsey et al, 2007

$$\phi_e(\boldsymbol{x}) \approx \frac{1}{4\pi \boldsymbol{\sigma}_B} \int_{\Omega_H} \frac{\nabla \cdot (\boldsymbol{\sigma} \nabla V)(\boldsymbol{y})}{\|\boldsymbol{x} - \boldsymbol{y}\|} d\boldsymbol{y}$$

- Equations can help model electrical activations in heart
  - Can invert to go from ECG to what is going on inside the heart
  - Machine learning can save time and resources to do this!

Challenge

Challenge Problem Overview

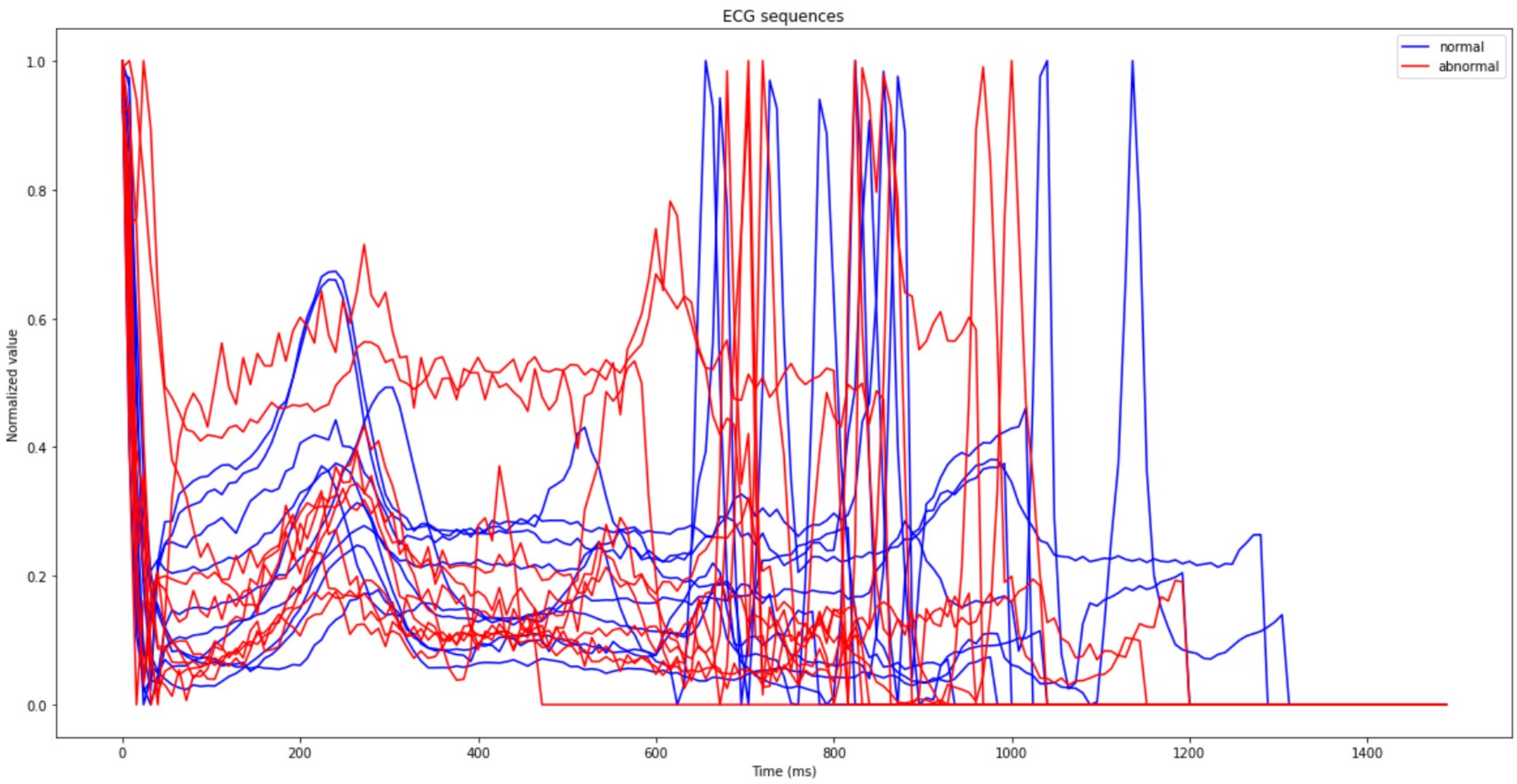
- Machine learning (ML) for cardiac electrocardiography
  - ML can aid cardiologists in their decision-making processes
  - ML can improve the accuracy and efficiency of diagnosing heart conditions using ECG data
  - Full ECG data and spatio-temporal activation maps of the heart are expensive and difficult to get, but ML can help!

- 4 tasks of increasing difficulty
  - Identifying abnormal vs. normal heart rhythms
  - Identifying heart conditions
  - Activation map reconstruction
  - Transmembrane potential reconstruction

- https://github.com/landajuela/cardiac_challenge

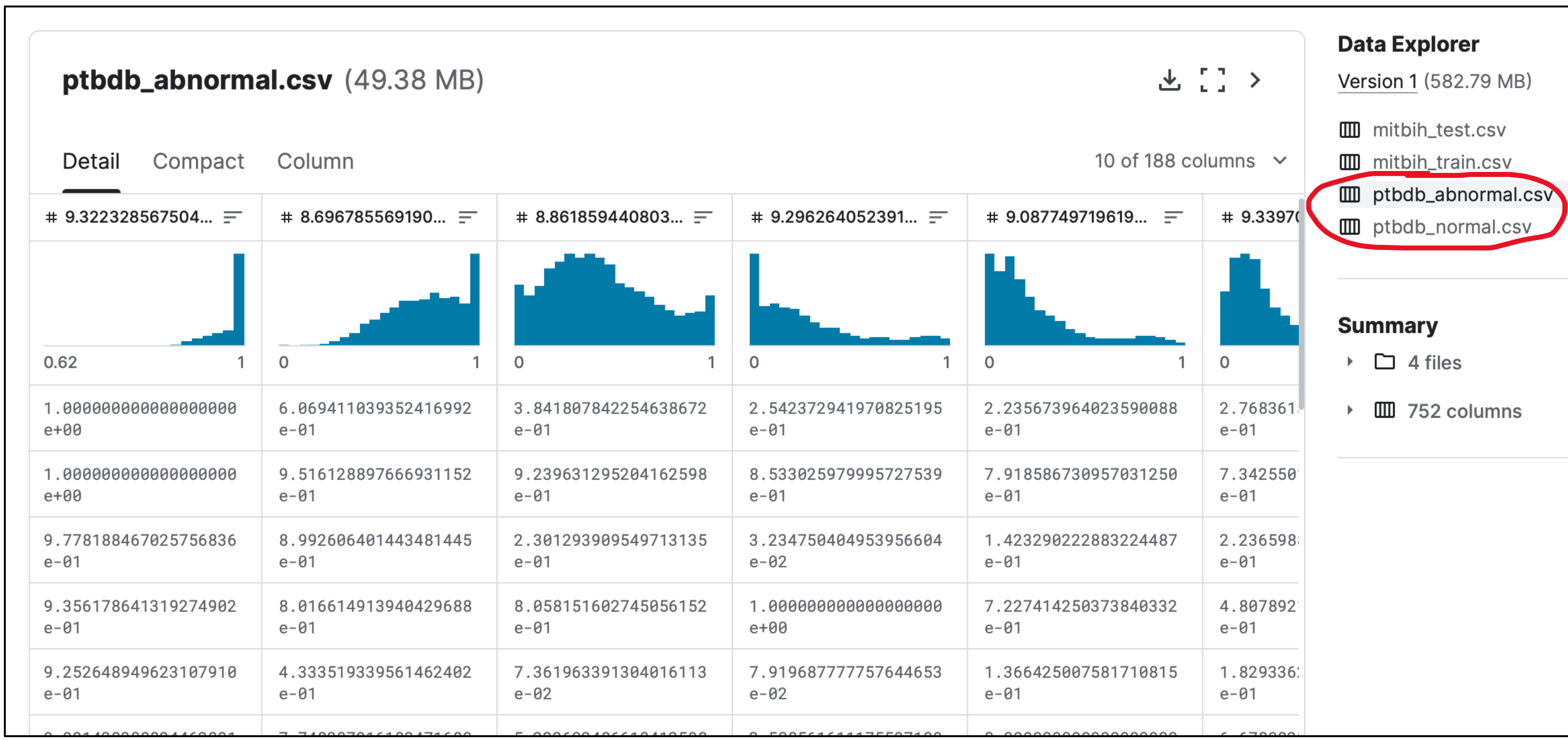
Task 1

Task 1: Binary classification for healthy vs. irregular heartbeats



Task 1: Data

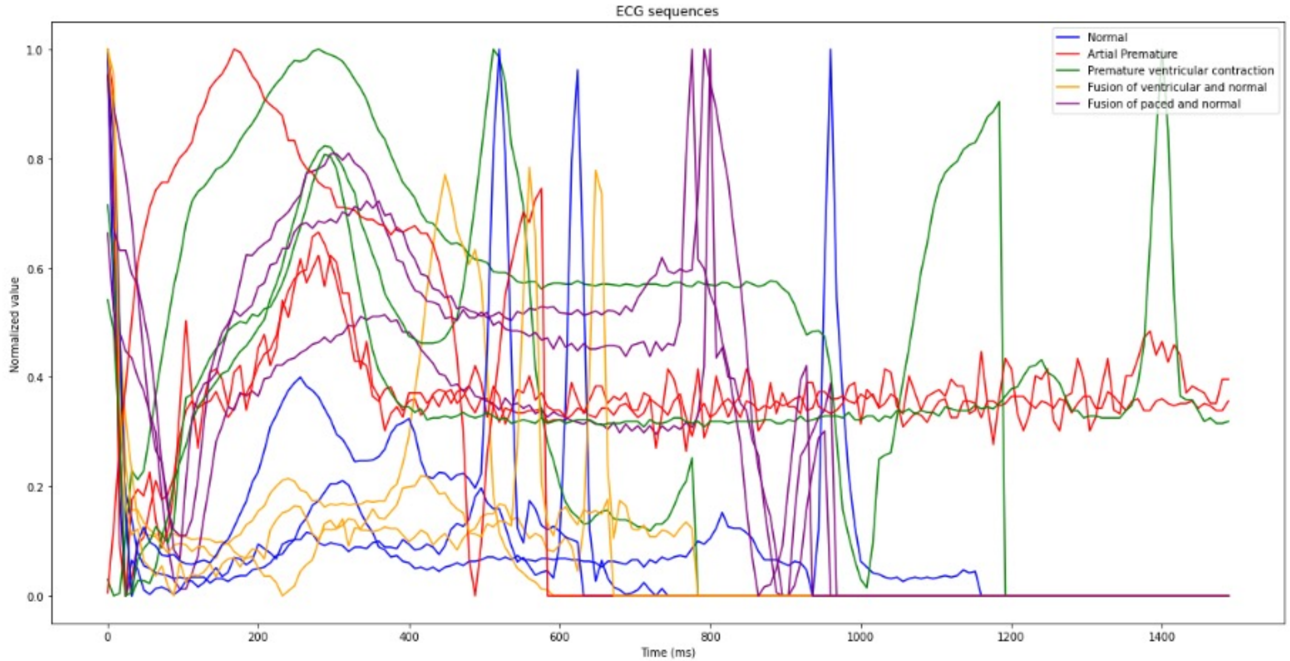
- Use the following ECG dataset for binary classification
  - https://www.kaggle.com/datasets/shayanfazeli/heartbeat



Task 2

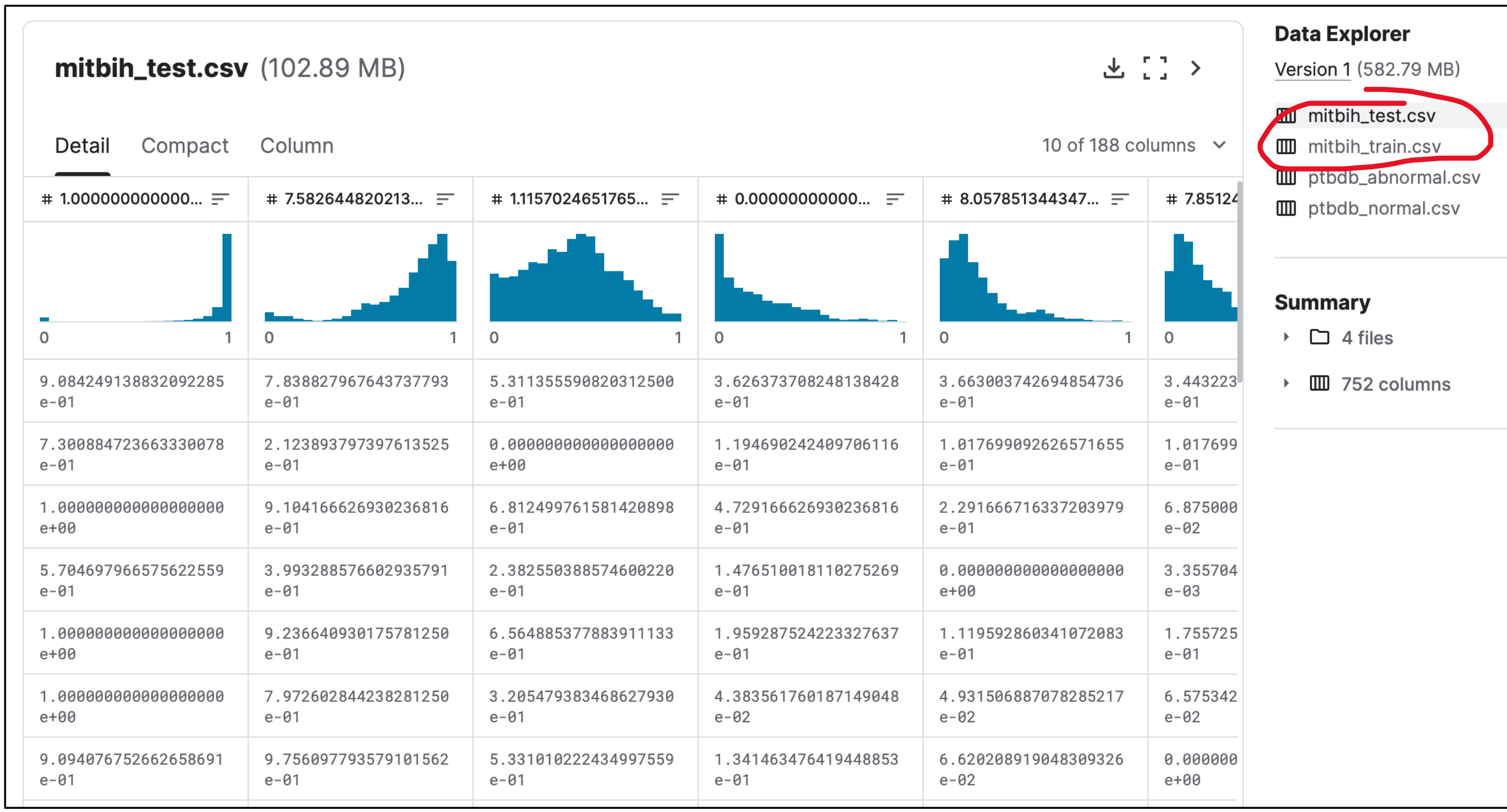
Task 2: multi-class classification on ECG data

- Further break each heartbeat class into multiple classes:
  - N: Non-ectopic beats (normal beats)
  - S: Supraventricular ectopic beats
  - V: Ventricle ectopic beats
  - F: Fusion beats
  - Q: Unknown beats



Task 2: Data

- Use the same Kaggle data from task 1, new dataset
  - Last column includes heartbeat class label

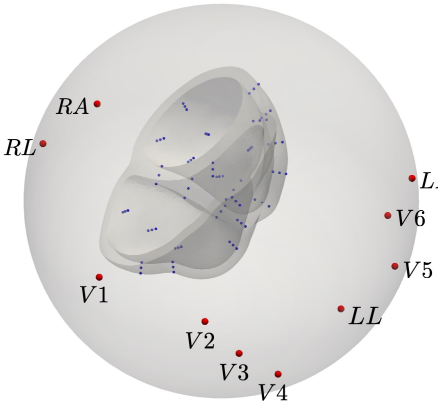


# 1.000000000000...	# 7.582644820213...	# 1.1157024651765...	# 0.00000000000...	# 8.057851344347...	# 7.85124
0 1	0 1	0 1	0 1	0 1	0
9.084249138832092285 e-01	7.838827967643737793 e-01	5.311355590820312500 e-01	3.626373708248138428 e-01	3.663003742694854736 e-01	3.443223 e-01
7.300884723663330078 e-01	2.123893797397613525 e-01	0.000000000000000000 e+00	1.194690242409706116 e-01	1.017699092626571655 e-01	1.017699 e-01
1.000000000000000000 e+00	9.104166626930236816 e-01	6.812499761581420898 e-01	4.729166626930236816 e-01	2.291666716337203979 e-01	6.875000 e-02
5.704697966575622559 e-01	3.993288576602935791 e-01	2.382550388574600220 e-01	1.476510018110275269 e-01	0.000000000000000000 e+00	3.355704 e-03
1.000000000000000000 e+00	9.236640930175781250 e-01	6.564885377883911133 e-01	1.959287524223327637 e-01	1.119592860341072083 e-01	1.755725 e-01
1.000000000000000000 e+00	7.972602844238281250 e-01	3.205479383468627930 e-01	4.383561760187149048 e-02	4.931506887078285217 e-02	6.575342 e-02
9.094076752662658691 e-01	9.756097793579101562 e-01	5.331010222434997559 e-01	1.341463476419448853 e-01	6.620208919048309326 e-02	0.000000 e+00

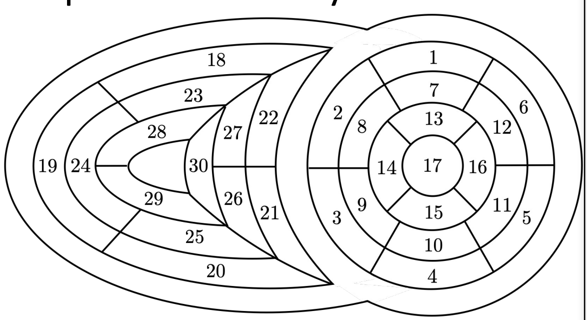
Task 3

Task 3: Activation Map Reconstruction
Understand the time when each part of the heart is activated

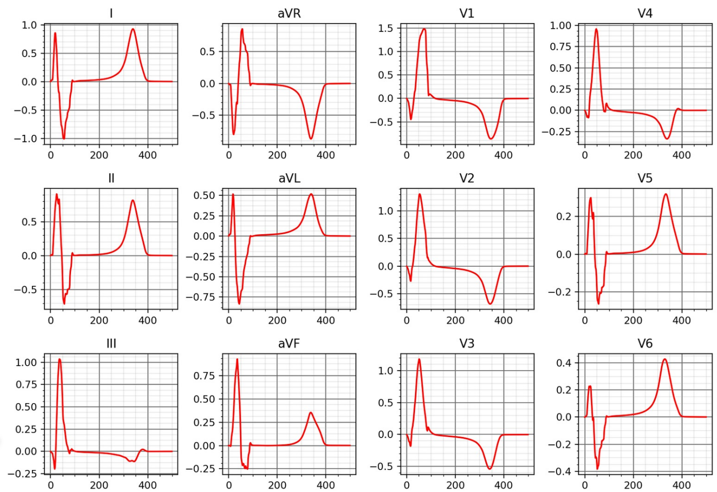
10 virtual electrodes



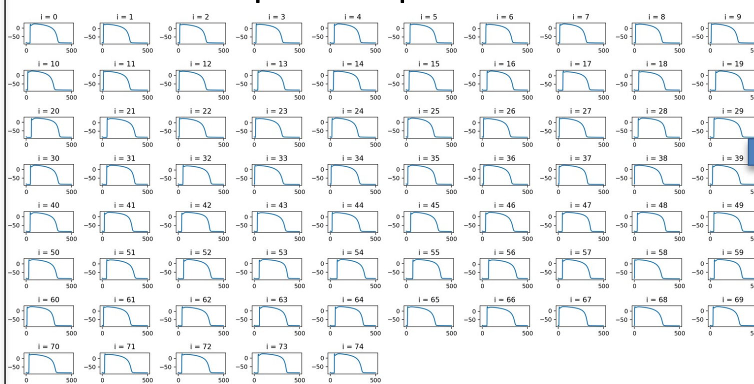
75 points inside myocardium



ECG data from 12 leads



Activation map for 75 points

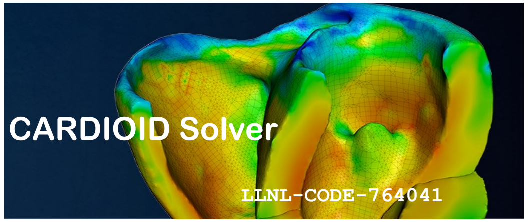


Understand the activation potential in heart

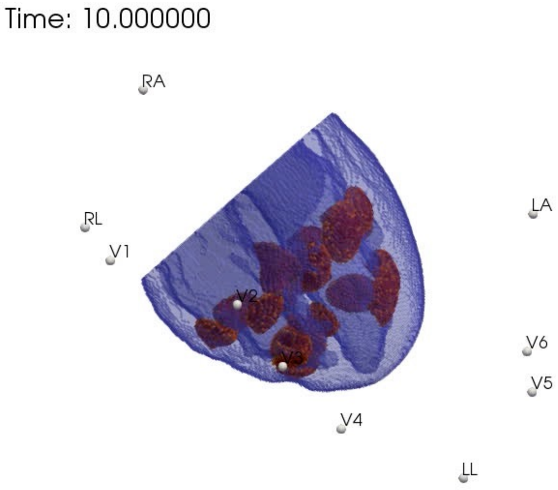
Task 3: Data Simulation

- Use LLNL's Cardioid solver:
  - https://github.com/LLNL/cardioid
  - Package for simulating cardiac electrophysiology, cardiac mechanics, torso-ECGs, cardiac meshing and fiber generation

Richards, David et al.(2013). Towards real-time simulation of cardiac electrophysiology in a human heart at high resolution. Computer methods in biomechanics and biomedical engineering

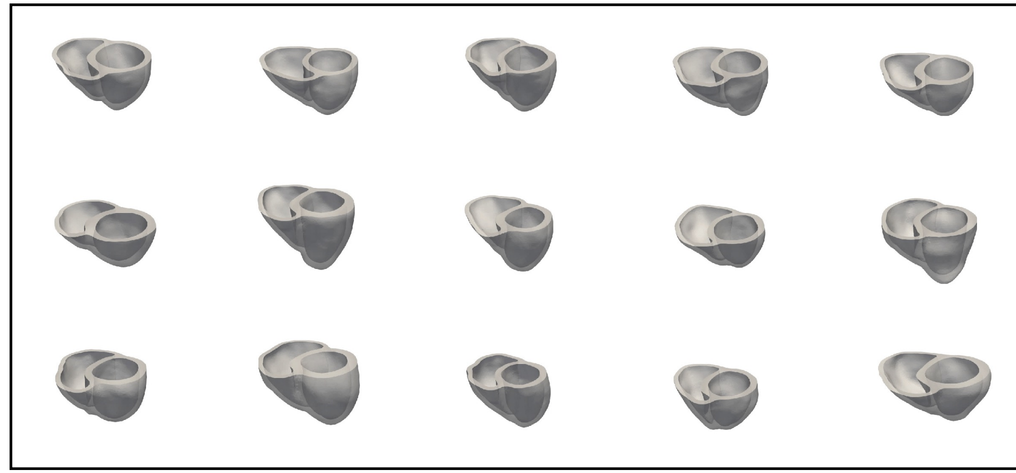


- Data augmentation and synthetic data approaches can help obtain very difficult data to get from a real heart
  - We can understand better what is happening in the heart by leveraging machine learning techniques trained on decades worth of research on cardiac electrophysiology
- You can then invert ECG to go backwards and understand what is going on inside the heart
  - Reconstructing the signal happening inside the heart

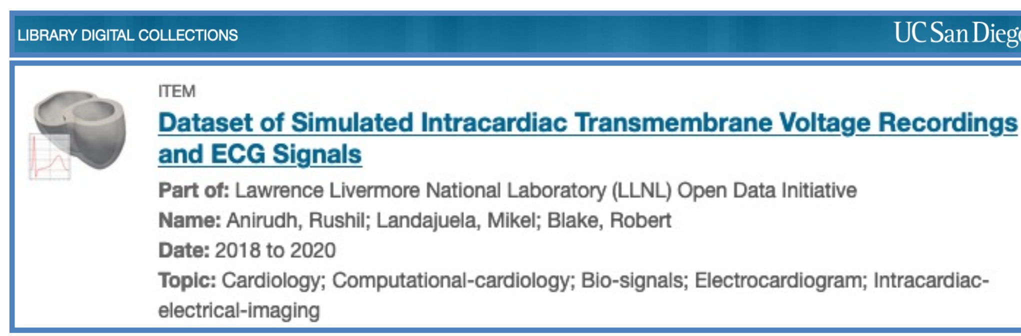


Task 3: Data

- LLNL dataset of simulated intracardiac transmembrane voltage recordings and ECG signals using Cardioid:
  - https://library.ucsd.edu/dc/object/bb29449106
  - 16,140 data points (5.504 GB)
- Data samples the space of ECG activation pairs:
  - 15 clinical bi-ventricular geometries
  - 29 clinically-inspired activation patterns
  - 3 different combinations of tissue conductivities
  - 3 values of G_{Kr}
  - 2 basic cycle lengths
  - Randomized samples over the space of inner activation points

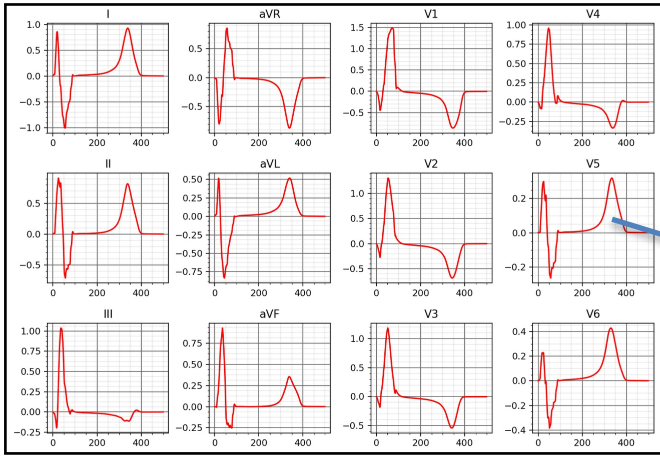


From *Model-based generation of large databases of cardiac images: synthesis of pathological cine mr sequences from real healthy cases - Duchateau et al, 2018*



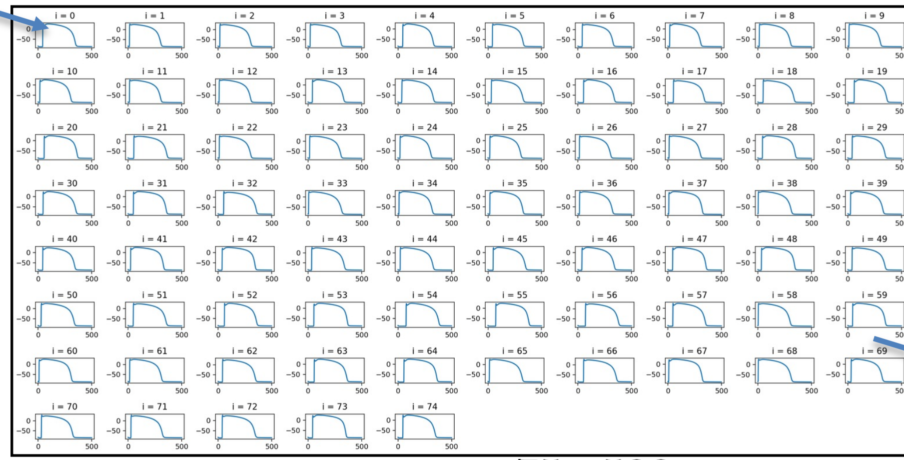
Task 3: Process

Understand the time when each part of the heart is activated



$$X \in \mathbb{R}^{12\times 500}$$

Train neural network on truth simulated ECG and activation maps



$$V \in \mathbb{R}^{75\times 500}$$

Predict spike time on activation map

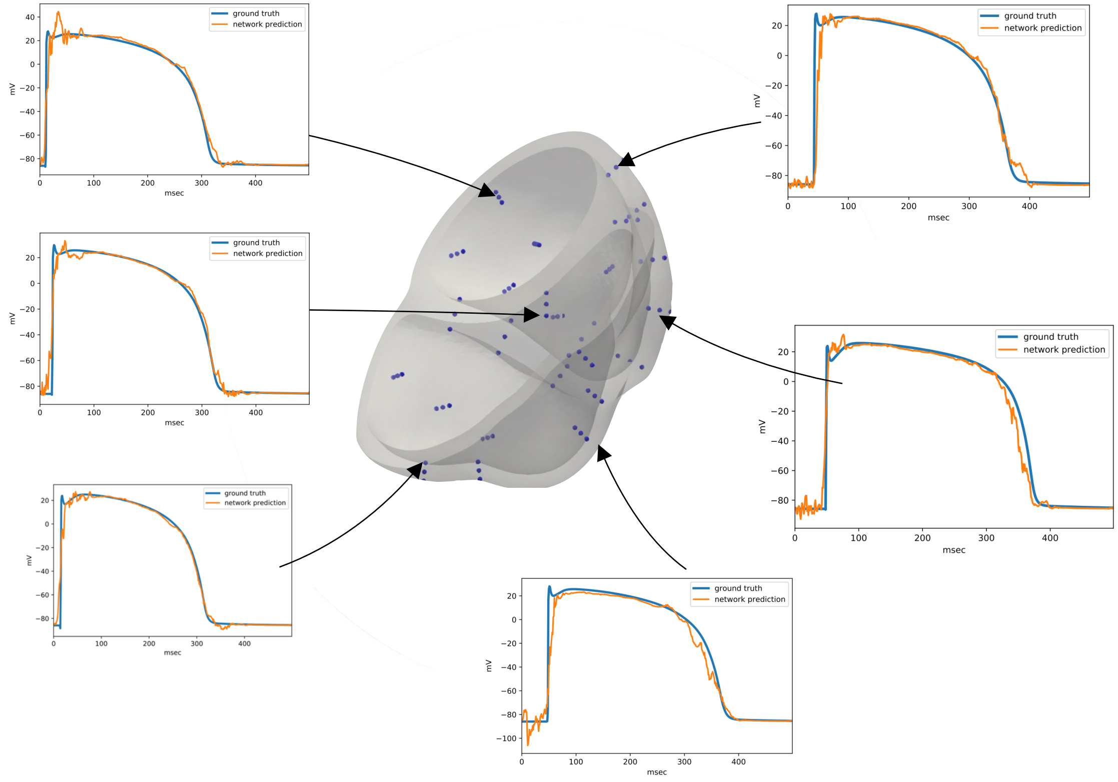
Task 3 : Activation map reconstruction (sequence transduction problem)

Given $X \in \mathbb{R}^{12\times 500}$ reconstruct $A \in \mathbb{R}^{75}$ with $A_i = \min_j V_{ij} > 0, \quad 1 \le i \le 75$

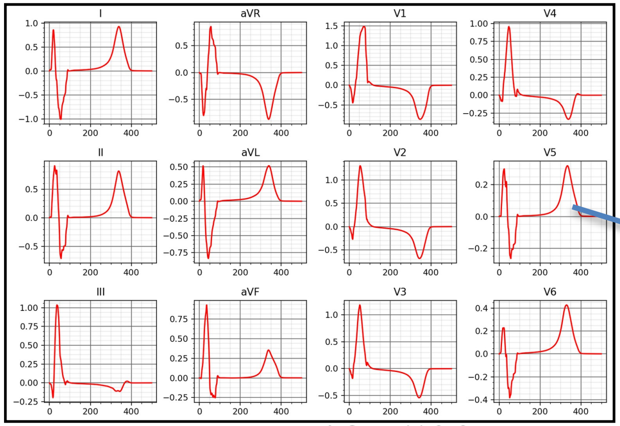
Task 4

Task 4: Transmembrane Potential Reconstruction

- Same dataset as before, but now reconstruct the entire transmembrane potential curve

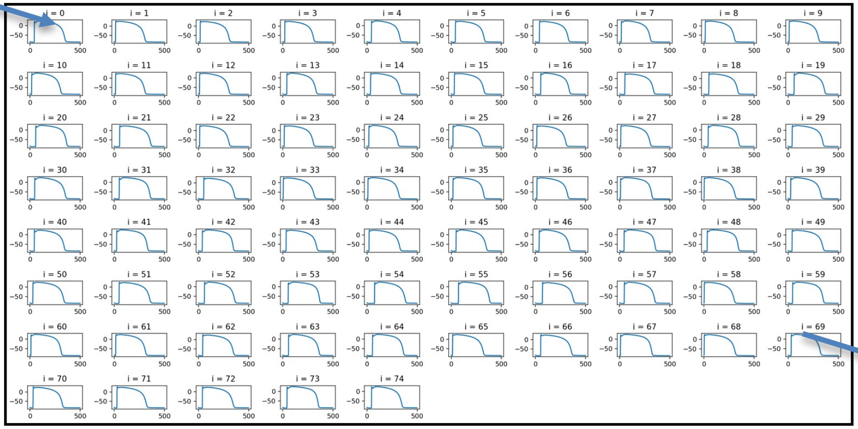


Task 4: Process



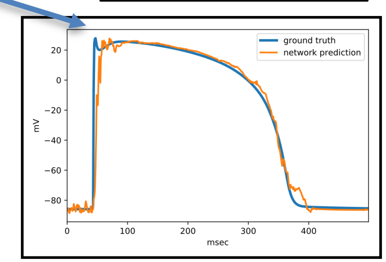
$X \in \mathbb{R}^{12\times 500}$

Train neural network on truth simulated ECG and activation maps



$V \in \mathbb{R}^{75\times 500}$

Predict activation function curve for all 75 points



Task 4 : Transmembrane potential reconstruction (regression per time step problem)

Given $X \in \mathbb{R}^{12\times 500}$ reconstruct $V \in \mathbb{R}^{75\times 500}$

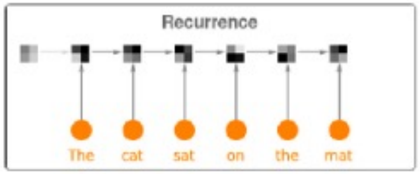
Starting points

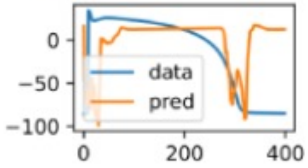
Task 1 and 2

- Download data
- Task 1:
  - The 2 datasets are separated between normal and abnormal heartbeats already
  - Assign a classification label to normal and abnormal heartbeats
  - Do a train/test split
- Task 2:
  - Files are already train/test split between the 2 files
  - Files already have the classification labels
  - Learn which labels correspond to which class

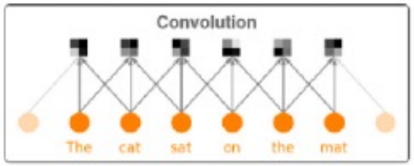
Neural Networks

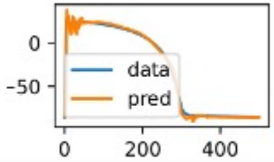
- Recurrent Neural Network (GRUs,LSTMs)
  - Natural for sequences
  - They are able to maintain its hidden state and learn dependencies over time, are Turing complete and are able to deal with sequences of any length
  - Working personal experience:
    - Difficult to train for long sequences
    - Good only when we deal with rather short sequences (10-100 time steps)





- 1D Convolutional Neural Networks
  - Convolutions are natural for images, but it turns outs they can model patterns in time series too!
  - Working personal experience:
    - Great performance
    - Faster to train





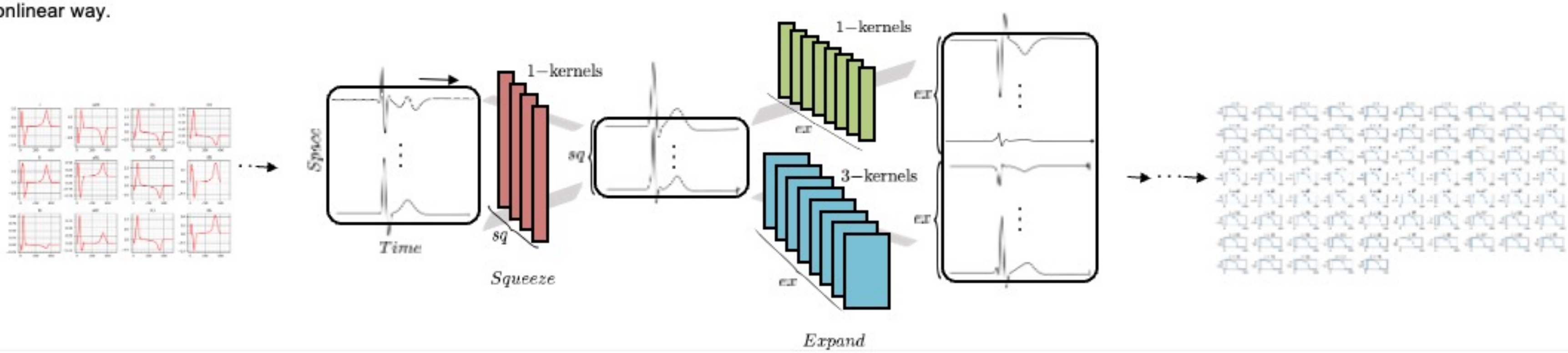
1D Sqeezenet

- *Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size* - Iandola et al, 2016
  - Fully convolutional, Accurate, Very Light Models
- 2 different Networks with the same root:

Network I
Conv 1D: filters = 64, kernel = 3, stride = 2, padding = 1
MaxPool 1D: kernel = 3, stride = 2, padding = 1
Fire : sq = 16, ex = 64
Fire : sq = 16, ex = 64
MaxPool 1D: kernel = 3, stride = 2, padding = 1
Fire : sq = 32, ex = 128
Fire : sq = 32, ex = 128
MaxPool 1D: kernel = 3, stride = 2, padding = 1
Fire : sq = 48, ex = 192
Fire : sq = 48, ex = 192
Fire : sq = 64, ex = 256
Fire : sq = 64, ex = 256
Droput : p = 0.1
Conv 1D: filters = 75, kernel = 3, stride = 2, padding = 0
Average pool 1D
Conv 1D: filters = 75, kernel = 3, stride = 2, padding = 0

Network II
Conv 1D: filters = 64, kernel = 3, stride = 1, padding = 1
MaxPool 1D: kernel = 3, stride = 1, padding = 1
Fire : sq = 16, ex = 64
Fire : sq = 16, ex = 64
MaxPool 1D: kernel = 3, stride = 1, padding = 1
Fire : sq = 32, ex = 128
Fire : sq = 32, ex = 128
MaxPool 1D: kernel = 3, stride = 1, padding = 1
Fire : sq = 48, ex = 192
Fire : sq = 48, ex = 192
Fire : sq = 64, ex = 256
Fire : sq = 64, ex = 256
Droput : p = 0.1
Conv 1D: filters = 75, kernel = 1, padding = 0

- *Fire* module : allow for both temporal and spatial information derived from the ECG signal to be combined and reorganized in a nonlinear way.



Resources

- Challenge problem GitHub: https://github.com/landajuela/cardiac_challenge/tree/main
- Task 1 & 2 Kaggle dataset: https://www.kaggle.com/datasets/shayanfazeli/heartbeat/data
- Task 3 & 4 simulated LLNL dataset: https://library.ucsd.edu/dc/object/bb29449106
- Optional resources describing LLNL datasets:
  - Simulation code GitHub: https://github.com/LLNL/cardioid
  - ML on cardiac data repo: https://github.com/landajuela/cardiac_ml
  - "digestible" article describing data: https://medium.com/@mikel.landajuela.larma/machine-learning-for-cardiac-electrocardiography-a20661669937
- Slack project with multiple channels for getting help from LLNL leadership team and peers: "2024 Data Science Challenge"