2023 Data Science Challenge

Cardiac Electrocardiography using Machine Learning

Co-lead DSC Organizer:

Mikel Landajuela

LLNL-PRES-845447

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

Agenda

- Background
- Task 1: Perform binary classification for healthy heartbeat vs. irregular
- Task 2: Perform multiclass classification to diagnose the irregular heartbeats
- Task 3: Activation Map Reconstruction from ECG
- Task 4: Transmembrane Potential Reconstruction from ECG
- Tips and Tricks
- Q/A

Background

Heart disease is the leading cause of death in the United States

- Heart disease is the leading cause of death for men, women, and people of most racial and ethnic groups in the United States.

- Every 34 seconds, one person in the US succumbs to cardiovascular disease, and it costs the country approximately $229 billion each year.

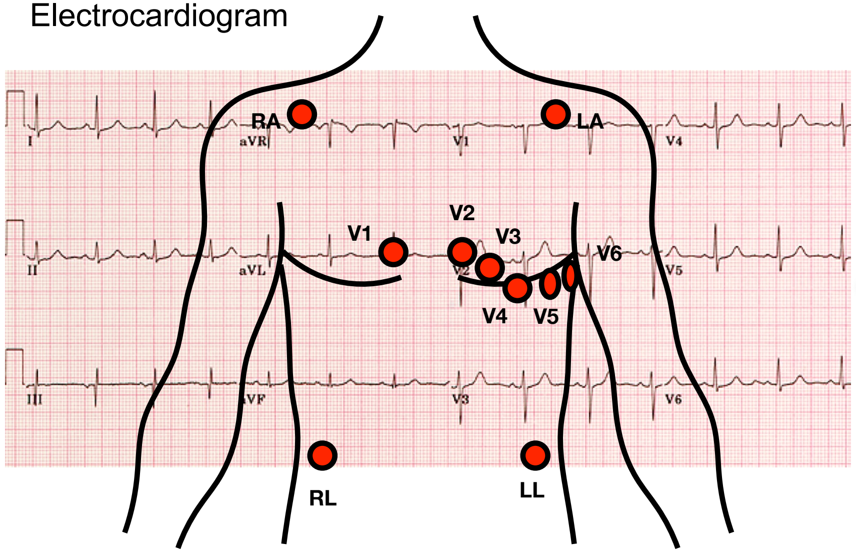
- Most sudden death is caused by abnormal electrical rhythms, or arrhythmia.

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

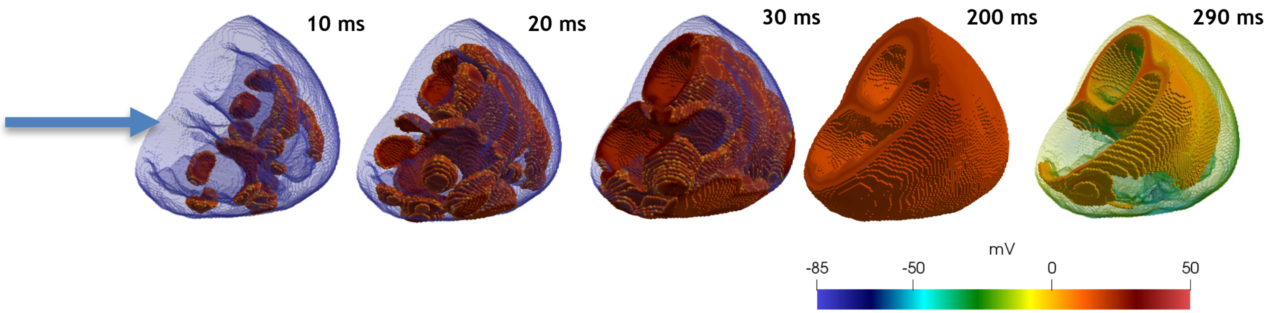
The electrocardiogram (ECG) plays an important role in diagnosing ventricular arrhythmias

- The electrocardiogram (ECG) plays an important role in diagnosing ventricular arrhythmias because it is non-invasive and cost-effective while still able of distinguishing a wide variety of diseases such as ventricular myocardial infarction and bundle branch blocks.
- In this year's data challenge, we'll delve into the exciting field of machine learning and how can it be used to aid cardiologists in their decision-making processes using ECG recordings.

Electrocardiogram



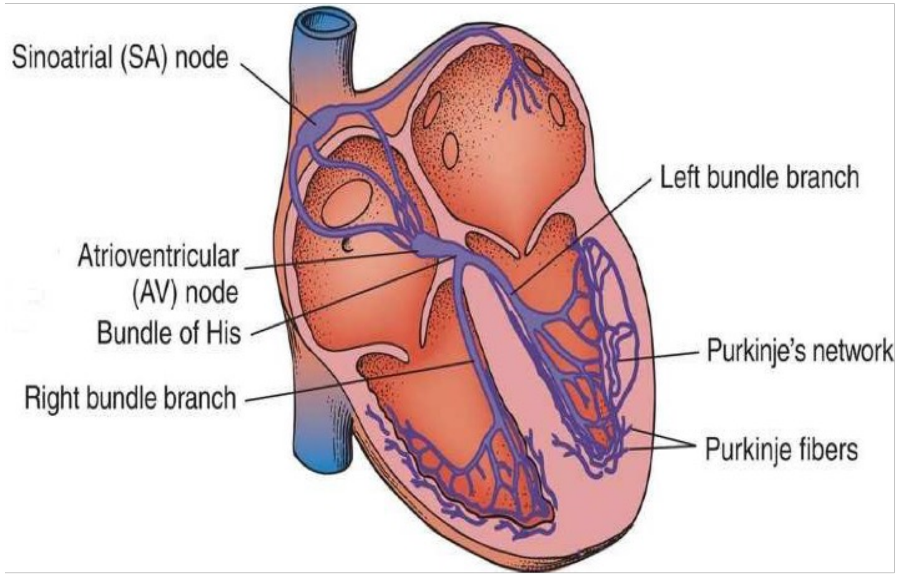
Cardiac Activation Sequence



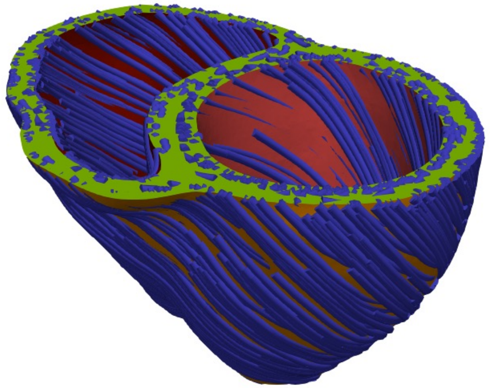
Basics of cardiac electrophysiology

- The activity of the heart results from a chemo-electro-mechanical phenomenon
- The typical cardiac cycle
  - The electrical impulse originates from the SA node, which acts as the heart's natural pacemaker
  - From there, it travels to the AV node, where it is delayed to allow for ventricle filling.
  - Next, the impulse moves through the specialized Purkinje Network, which facilitates rapid conduction of electrical signals
  - Finally, it enters the myocardium via the Purkinje-Muscle Junctions, triggering the necessary reactions for mechanical contraction

  If there is any malfunction in this process, it may be identified by an ECG

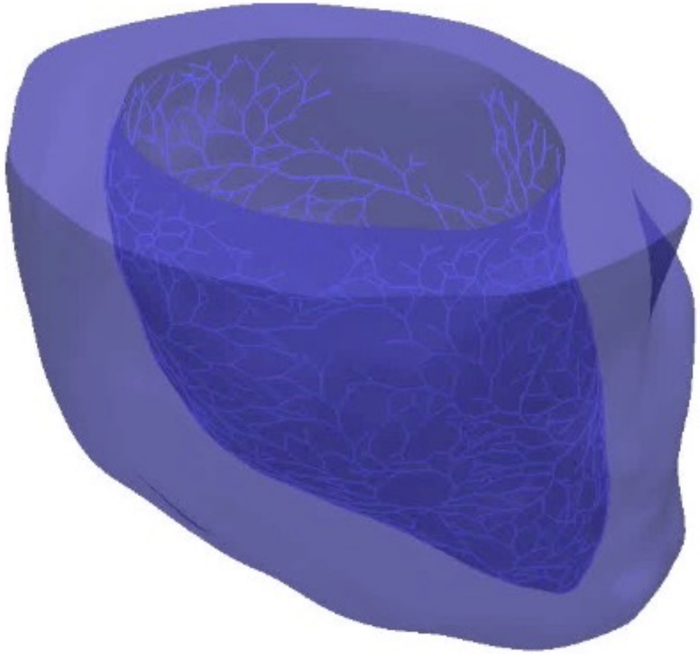


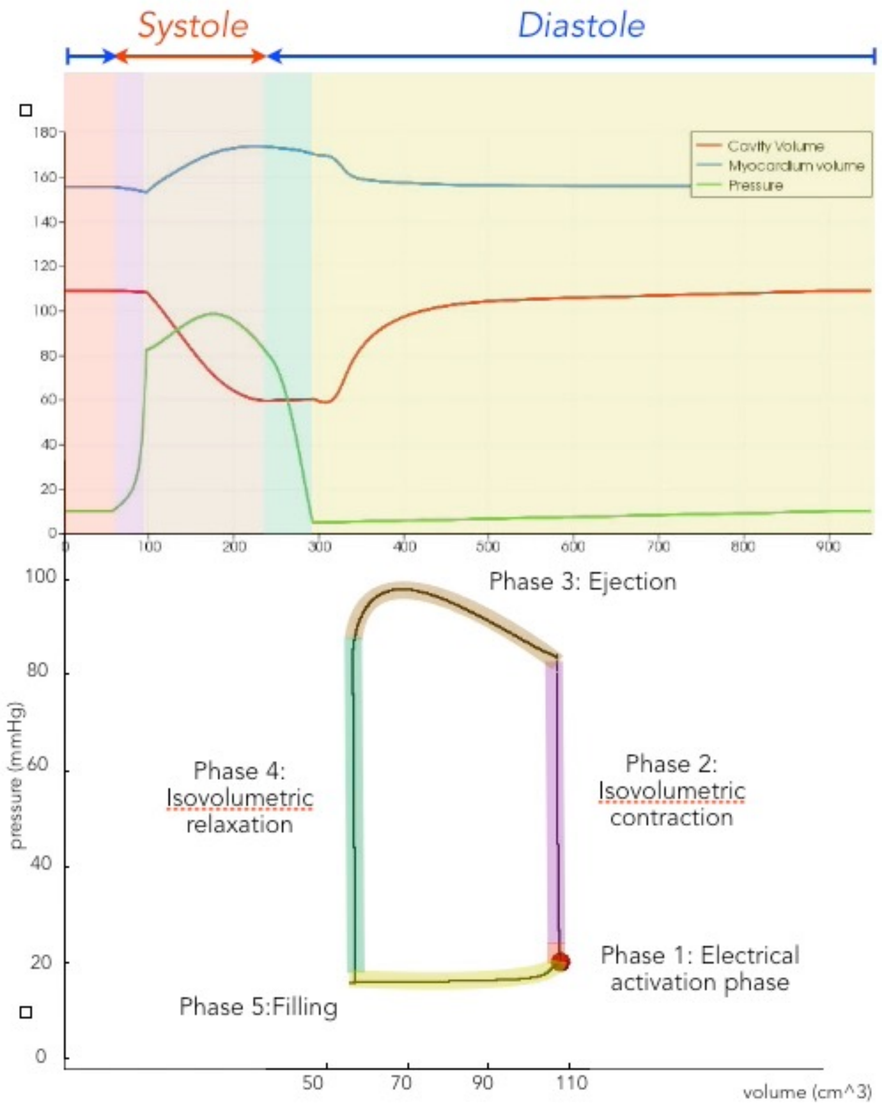
http://medical-dictionary.thefreedictionary.com



- The myocardium is a complex tissue comprised of cardiomyocytes organized in fibers and laminar collagen sheets, and it exhibits an anisotropic architecture with highly anisotropic electrical and mechanical material properties.

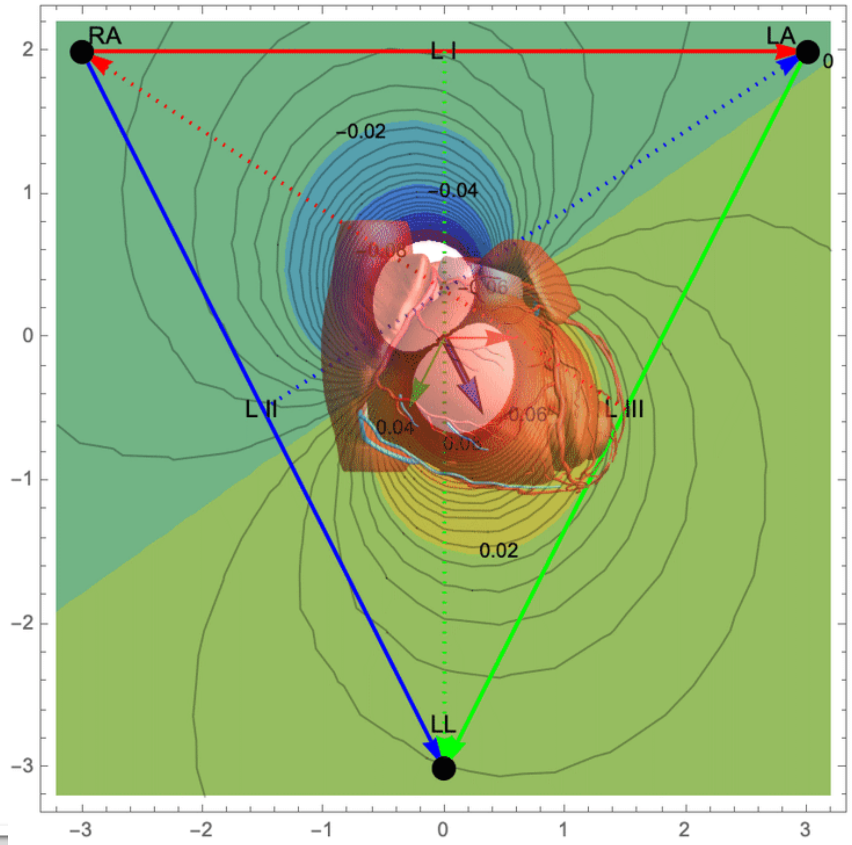
The full cardiac cycle in the left ventricle

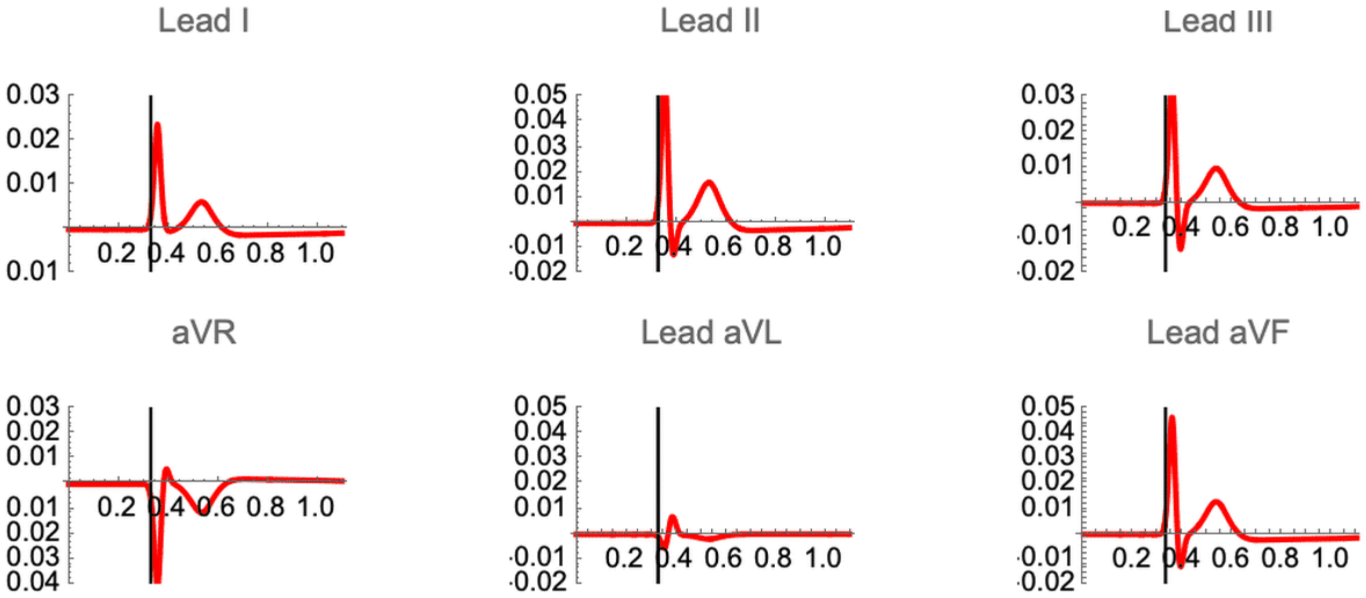


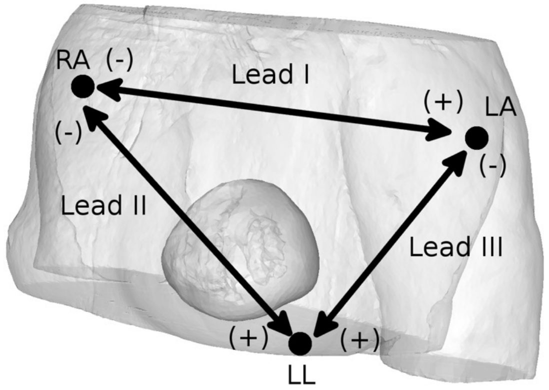


As the electrical signal travels through the torso, it can be detected and recorded by an ECG

- The activation wave generates an external potential that propagates through the torso
- The ECG consists of measurements of that signal in different torso locations called leads



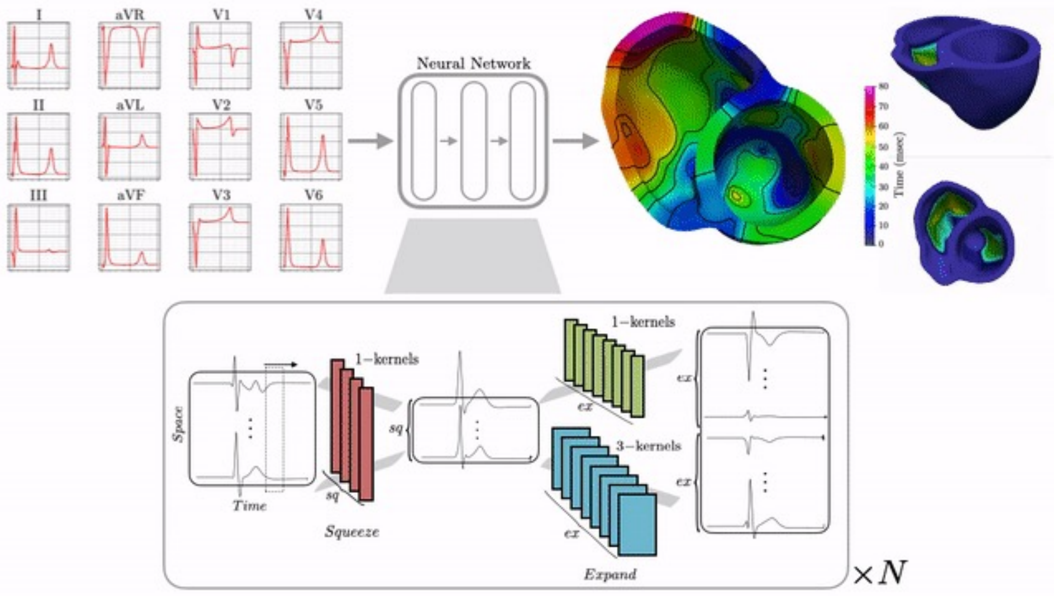




DSC Tasks

DSC 2023 is to develop machine learning tools for Cardiac Electrocardiography

- The Data Science Challenge involves **four tasks of increasing complexity** to develop machine learning tools for diagnosing heart conditions using electrocardiograms (ECG).
- The tasks range from identifying heart conditions from ECG profiles to reconstructing a complete spatio temporal activation map of the heart. The goal is to improve the accuracy and efficiency of diagnosing heart conditions using ECGs.



Task 1 : perform binary classification for healthy heartbeat vs. irregular

- Get familiar working with ECG data by using the dataset below to perform binary classification for healthy heartbeat vs. irregular
- Link : https://www.kaggle.com/datasets/shayanfazeli/heartbeat

Task 2 : perform multiclass classification to diagnose the irregular heartbeats

- Diagnosing an irregular heartbeat by using the to perform multiclass classification to diagnose the irregular heartbeats.
- Link : https://www.kaggle.com/datasets/shayanfazeli/heartbeat

-N : Non-ecotic beats (normal beat)
-S : Supraventricular ectopic beats
-V : Ventricular ectopic beats
-F : Fusion Beats
-Q : Unknown Beats

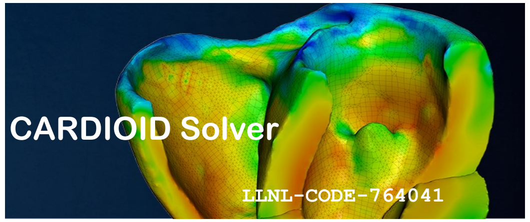
Task 3 and 4

Data Augmentation through detailed Multiphysics simulations

While there are many ECG recordings available for training machine learning models to classify heart conditions, obtaining full datasets of ECG and spatio-temporal activation maps of the heart is expensive and difficult.

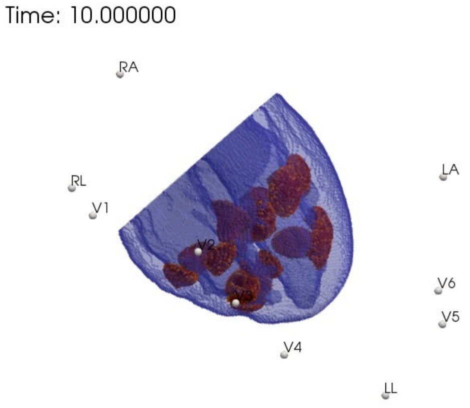
To create these datasets, techniques such as data augmentation and synthetic data approaches are needed.

By leveraging decades of research on cardiac electrophysiology and training neural networks on detailed simulations, we can gain a better understanding of the heart and its activity.



Cardioid solver @ LLNL https://github.com/LLNL/cardioid

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



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

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

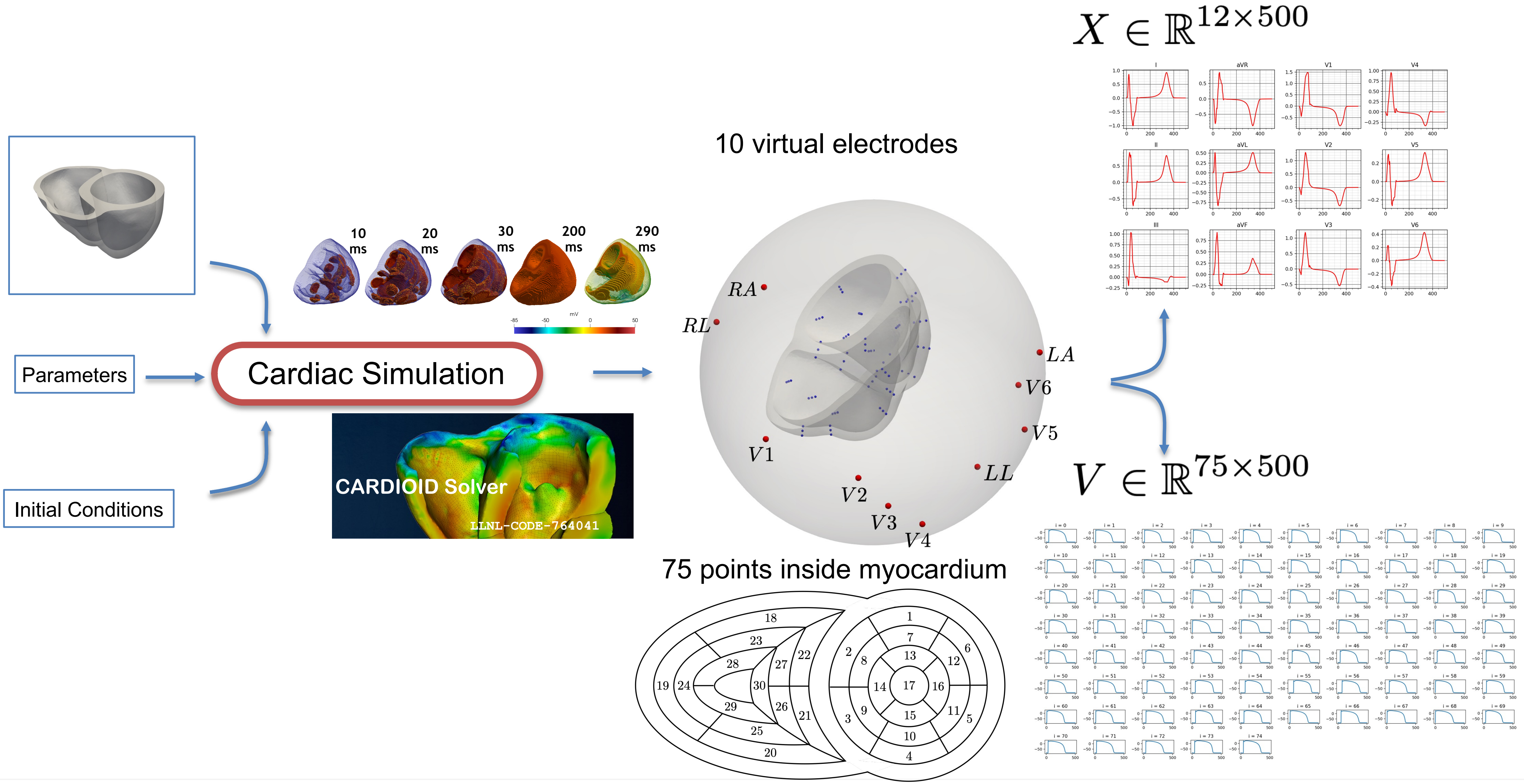
Alternants and spiral breakup in a human ventricular tissue model - Ten Tusscher 2006 with endo-, mid- and epicardial cell heterogeneity

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

Bioelectricity: a quantitative approach - Plonsey et al, 2007

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

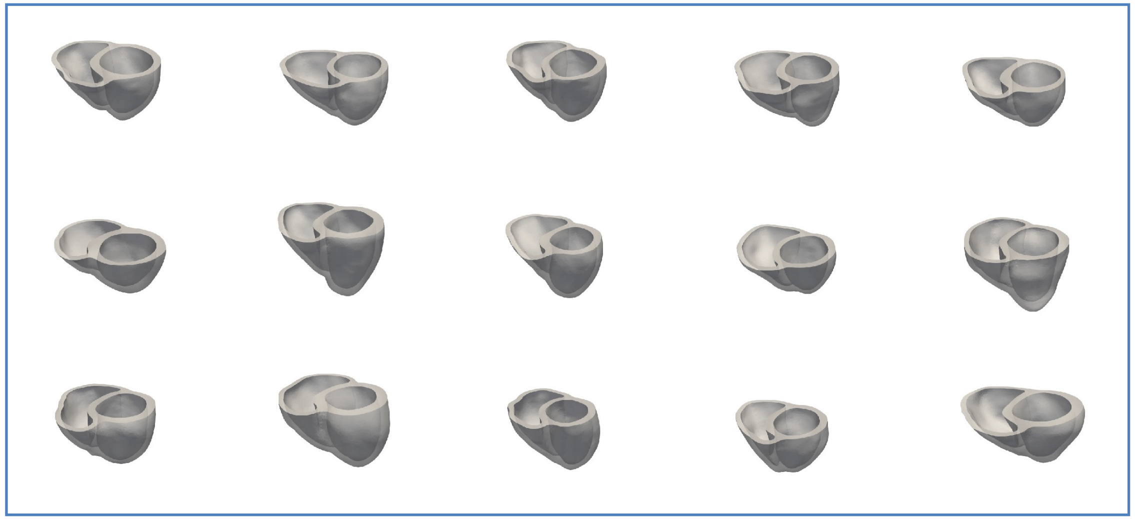
Data augmentation through detailed Multiphysics simulations



Dataset of Simulated Intracardiac Transmembrane Voltage Recordings and ECG Signals

Sample the space of ECG-Activation pairs

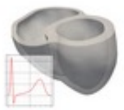
- 15 clinical bi-ventricular geometries
- 29 clinically-inspired activation patterns
- 3 different combinations of tissue conductivities
- 3 different values of G_{Kr}
- 2 Basic Cycle Lengths
- Randomized samples over the space of inner activation points



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

Total: 16140 data points (5.504 GB)

LIBRARY DIGITAL COLLECTIONS UC San Diego



ITEM

Dataset of Simulated Intracardiac Transmembrane Voltage Recordings and ECG Signals

Part of: Lawrence Livermore National Laboratory (LLNL) Open Data Initiative

Name: Anirudh, Rushil; Landajuela, Mikel; Blake, Robert

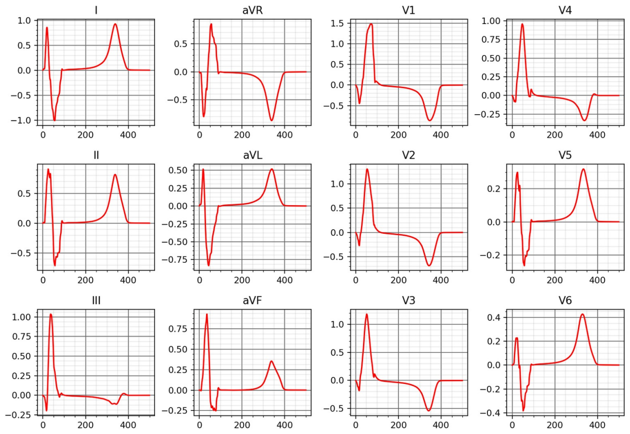
Date: 2018 to 2020

Topic: Cardiology; Computational-cardiology; Bio-signals; Electrocardiogram; Intracardiac-electrical-imaging

https://library.ucsd.edu/dc/object/bb29449106

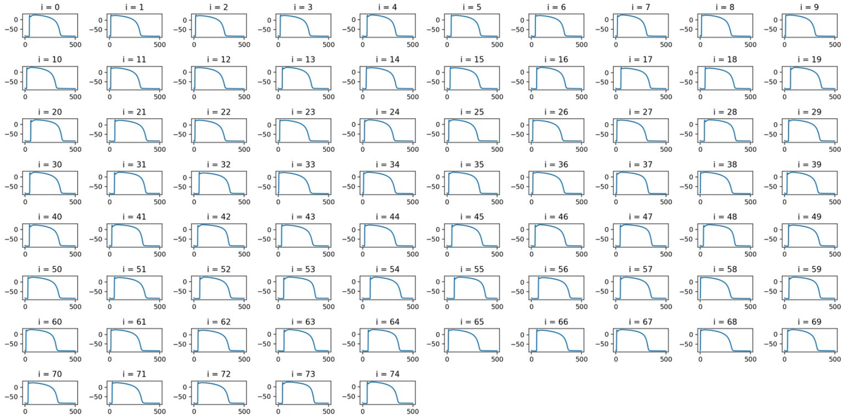
Deep intracardiac electro imaging : Sequence-to-Sequence Mapping

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



Neural Network

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



Task 3 : Activation map reconstruction (sequence transduction problem)

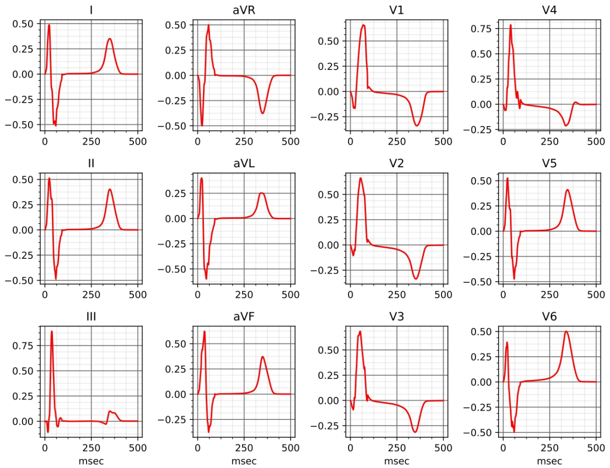
Given $X \in \mathbb{R}^{12\times 500}$ reconstruct $A \in \mathbb{R}^{75}$ with $A_i = \min_j V_{ij} > 0, \quad 1 \leq i \leq 75$

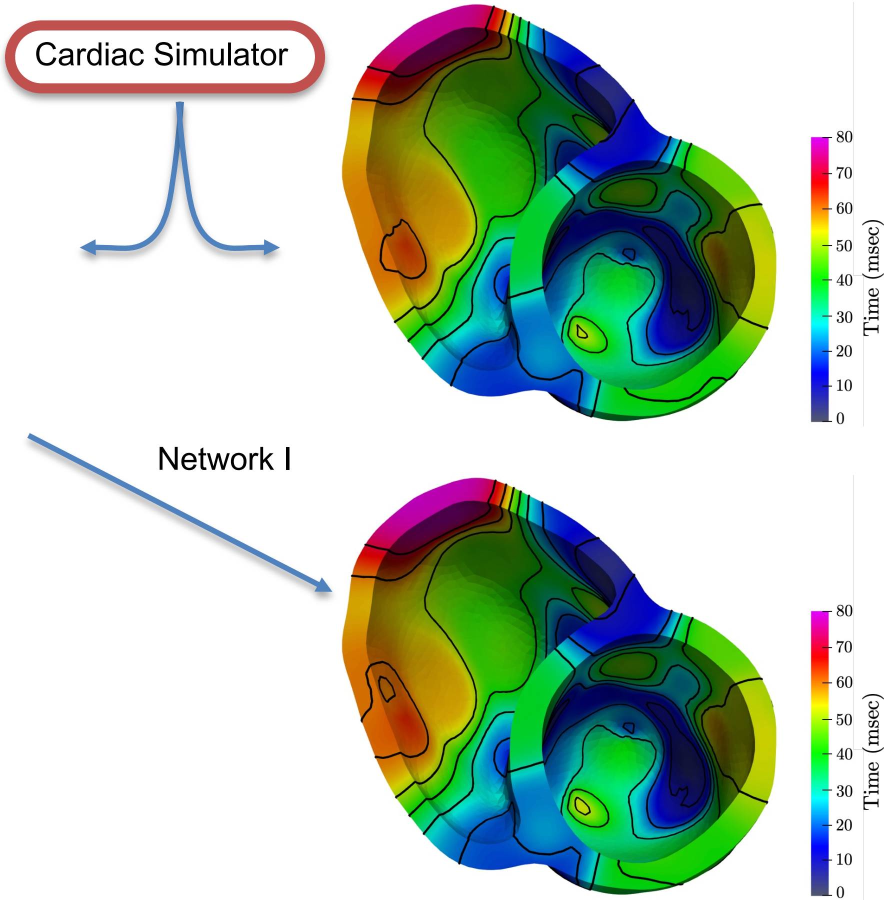
Task 4 : Transmembrane potential reconstruction (regression per time step problem)

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

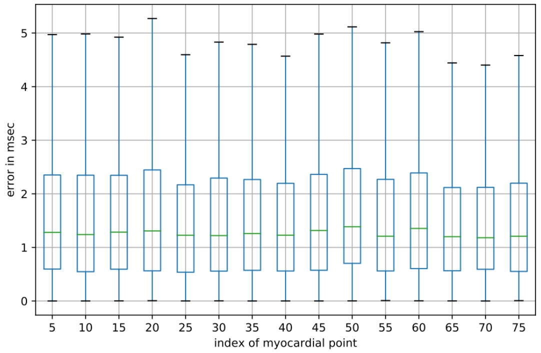
Task 3 : Activation Map Reconstruction from ECG

- Ground truth:



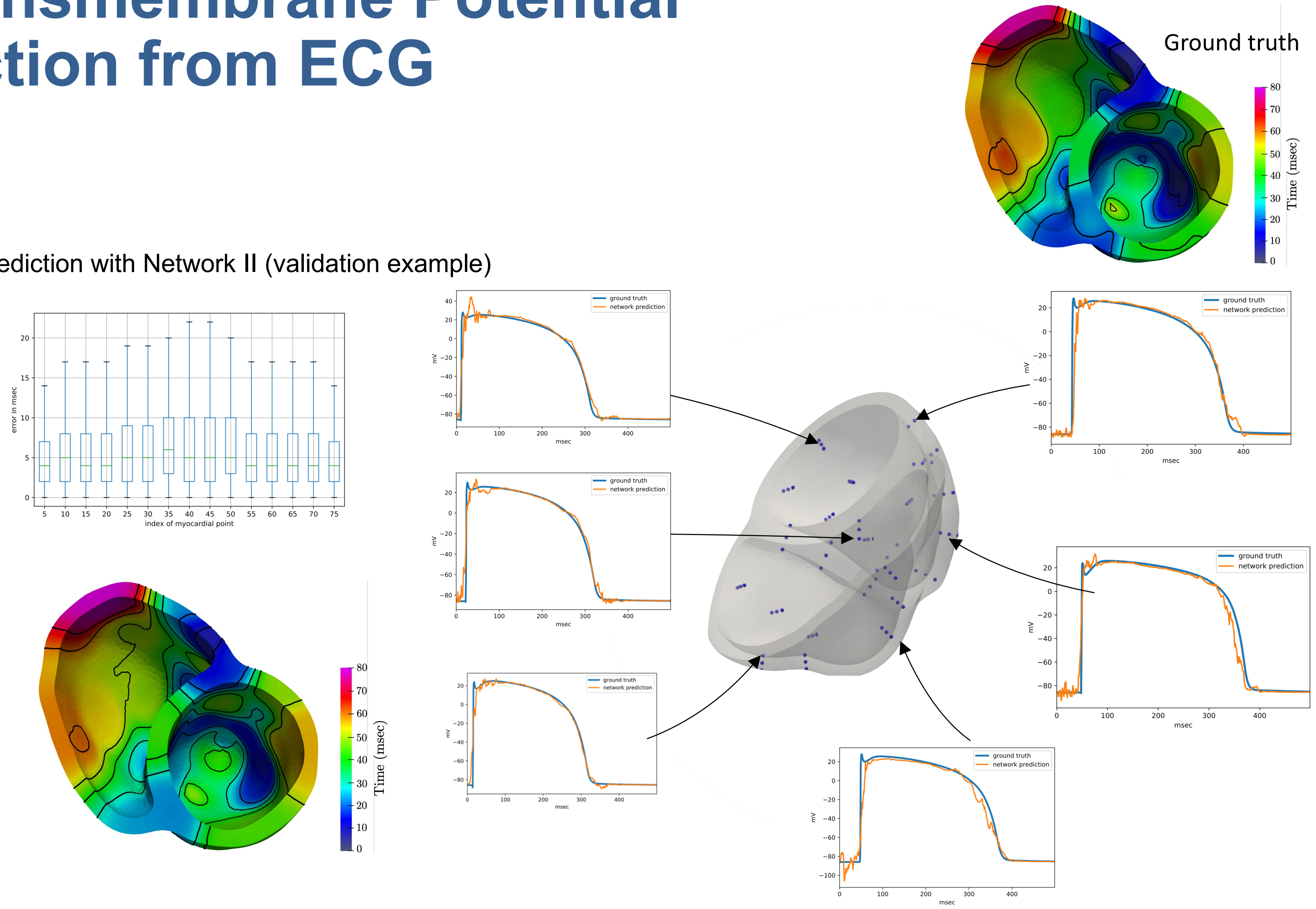


Prediction with Network I (validation example)



Task 4 : Transmembrane Potential Reconstruction from ECG

Prediction with Network II (validation example)



Tips and Tricks

Neural Network Architecture Choice

- Recurrent Neural Network (GRUs,LSTMs)

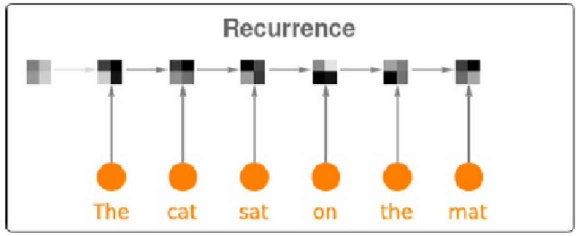
  Natural for sequences

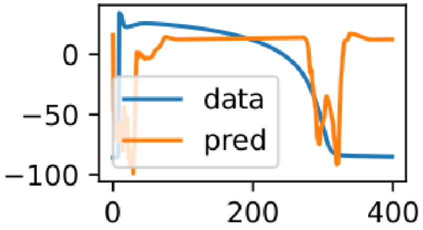
  They are able to maintain its hidden state and learn dependencies over time, are Turing complete and are able to deal with sequences of any length

  Working personal experience:

  Difficult to train for long sequences

  Good only when we deal with rather short sequences (10-100 time steps)





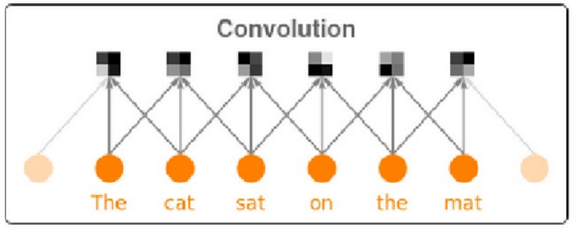
1D Convolutional Neural Networks

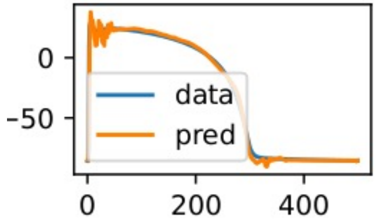
Convolutions are natural for images, but it turns outs they can model patterns in time series too!

Working personal experience:

Great performance

Faster to train





Architecture: 1D Squezeenet

Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size - Iandola et al, 2016

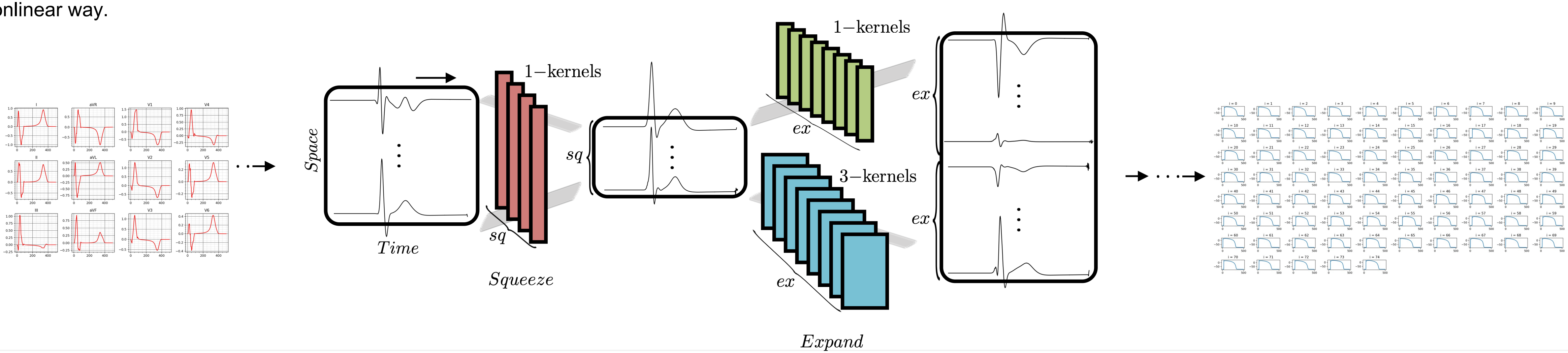
Fully convolutional, Accurate, Very Light Models

- 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.



Q/A