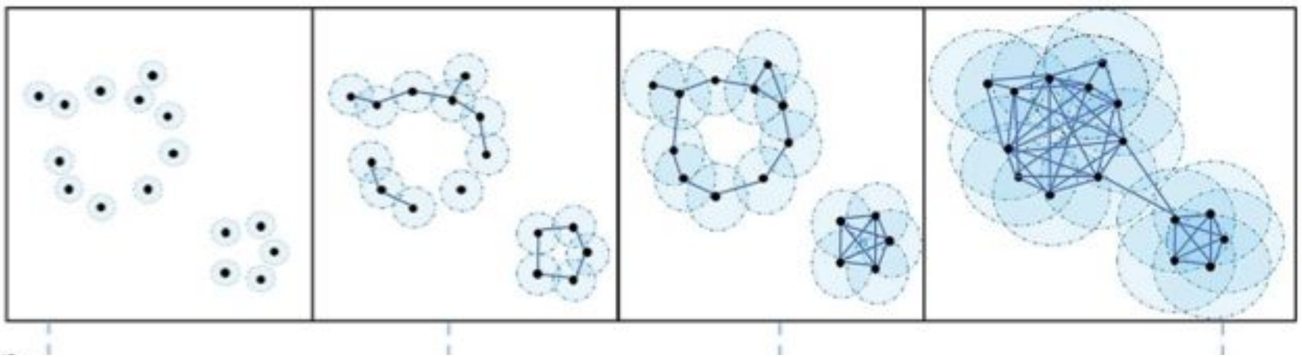
An Exploration of Topological Data Analysis (TDA) for Arrhythmia Detection

Abby Headstrom & Abby Irish

Topological Data Analysis

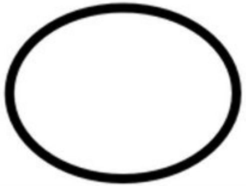
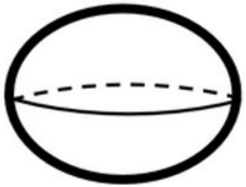
- Characterization of data using its geometric structure
- Persistent Homology: specifically looks at the structure of holes
- Relevance to machine learning



Persistent Homology

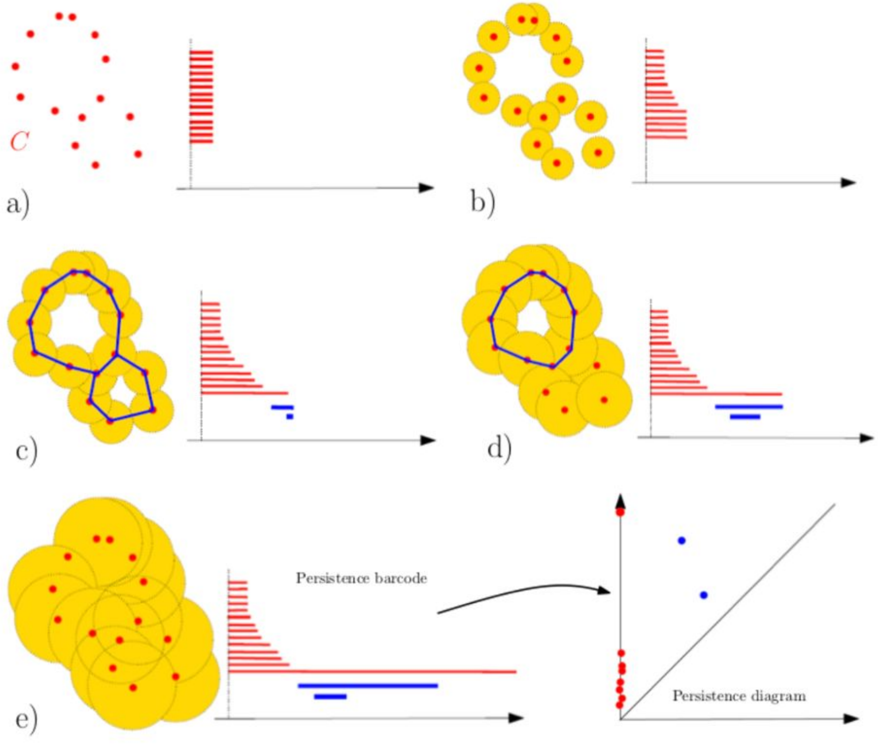
- In a given nested sequence $X1 \subseteq X2 \subseteq . . . \subseteq Xn$, the inclusion $Xi \subseteq Xi'$ for $i \leq i'$ produces a linear map $Hk(Xi) \rightarrow Hk(Xi')$ on the corresponding k-th homology. As the time parameter *i* increases, persistent homology tracks elements of $Hk(Xi)$, most often represented with a persistence diagram (PD) in the Cartesian plane R^2.
- Each point (x, y) in a PD corresponds to a feature appearing at scale x (birth) and disappearing at scale y (death), and has a persistence value of (y - x).

Betti Numbers

				
β_0	1	1	1	1
β_1	0	1	0	2
β_2	0	0	1	1
β_3	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Persistence Diagrams and Barcodes

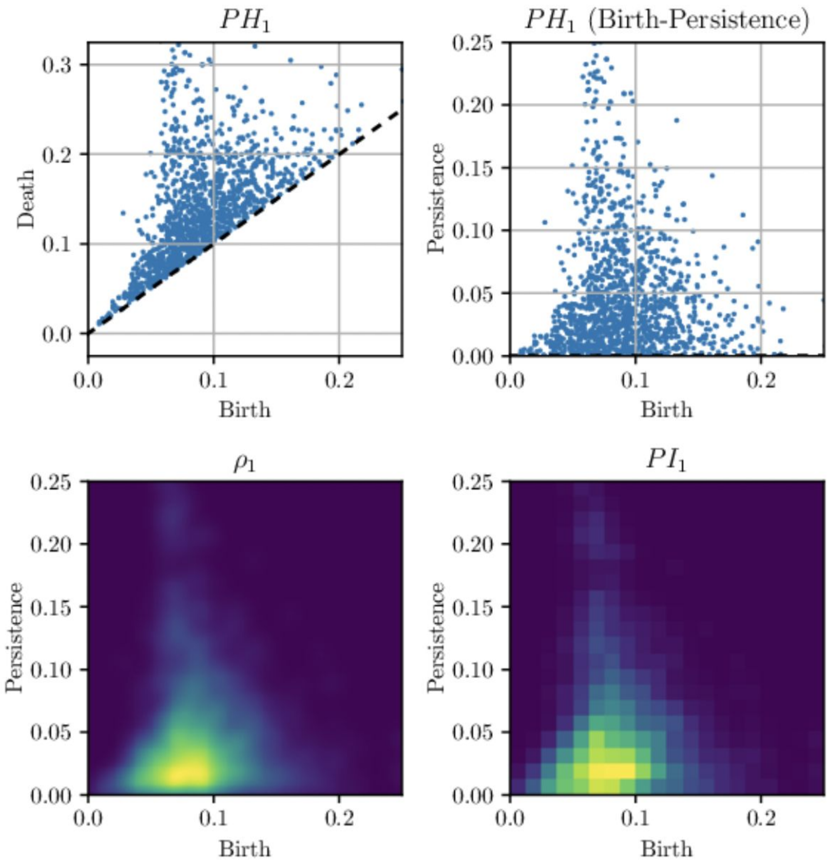
In the example below we consider the filtration given by a union of growing balls centered on the finite set of points C.



Persistence Diagram → Persistence Image

- Need a vector representation in order to perform machine learning methods such as SVM, logistic regression, etc – Persistence Images provide this
  - Involves a linear transformation of the PD, *B*, from birth-death coordinates to birth-persistence coordinates: $\boldsymbol{T: R^2 \to R^2}$ **defined as** $\boldsymbol{T(x, y) = (x, y - x)}$
  - Center the points usually using a Gaussian distribution function:
  - Fix a nonnegative weighting function f: $R^2 \to R$ to produce the corresponding persistence surface and ensure a stable transformation:

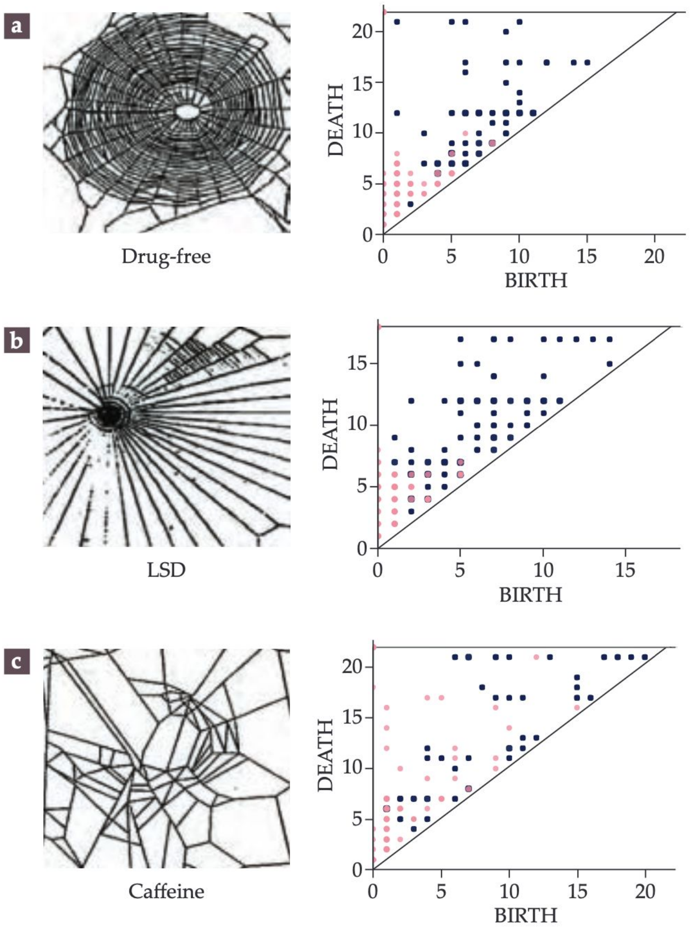
    $\boldsymbol{\rho_B(z) = \Sigma_{u \in T(B)} f(u)\varphi_u(z)}$
  - Reduce the surface to a finite-dimension vector by overlaying it on a grid and integrating over each pixel to get the PI.



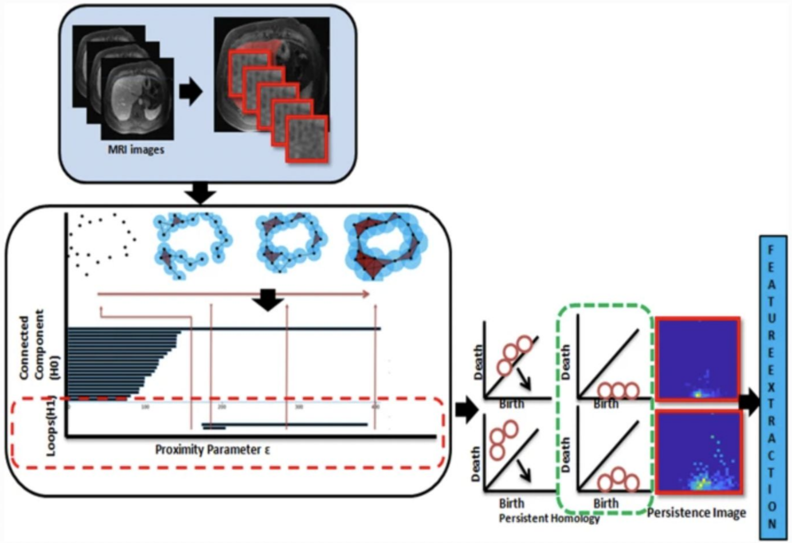
PH_1
Death
Birth
PH_1 (Birth-Persistence)
Persistence
Birth
ρ_1
Persistence
Birth
PI_1
Persistence
Birth

Applications of TDA

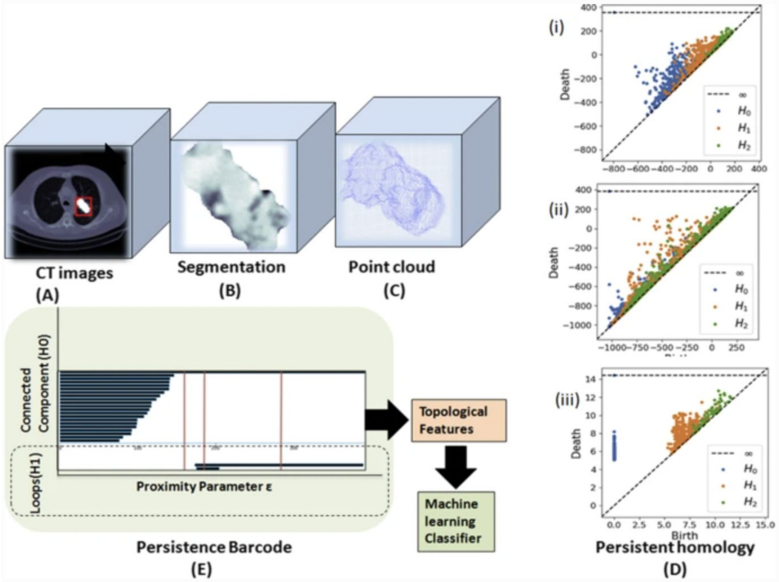
- Shape extraction
- Medical image processing
  - High-dimensional MRIs and CT scans
  - Tumor detection
- Time-series analysis
  - ECGs
- Physics and dynamical systems
- Literature reviews
- Machine learning
  - Similarities to CNNs



TDA and Healthcare



Workflow of algebraic topology-based machine learning with MRI imaging signal as input



Heart Arrhythmia Detection

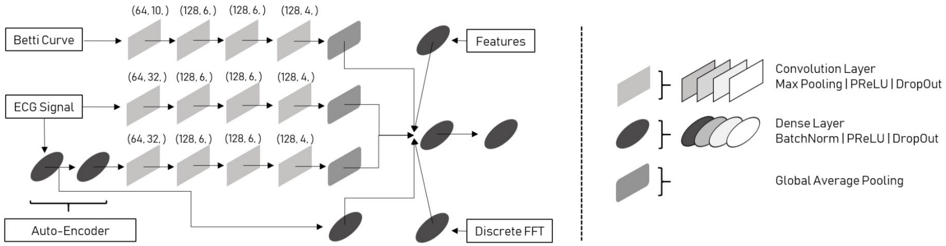


Figure 4: General Overview of Deep-Learning Architecture

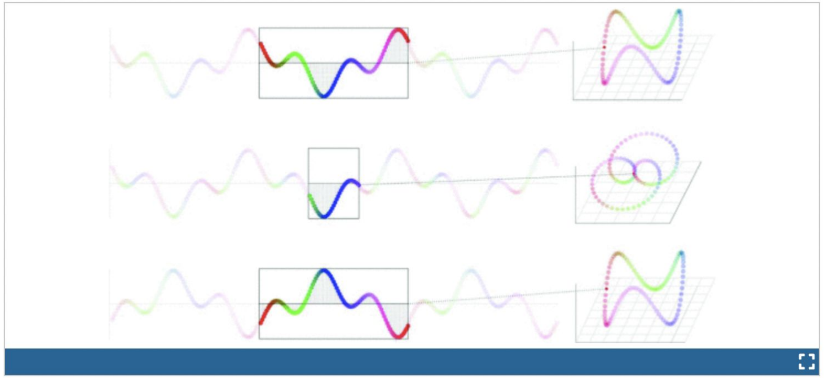


Fig. 2.
Multi-dimensional scaling to R^3 of the high-dimensional point clouds generated from delay embeddings with varying window sizes (respectively 100, 35, and 100 time units). Here, the embedding dimensions are set equal to the window sizes and the delay parameter τ is set to 1. Projection to R^3 is only used for visualization.

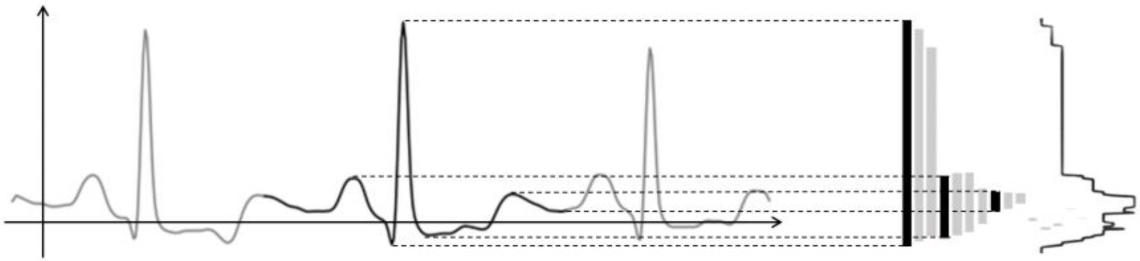
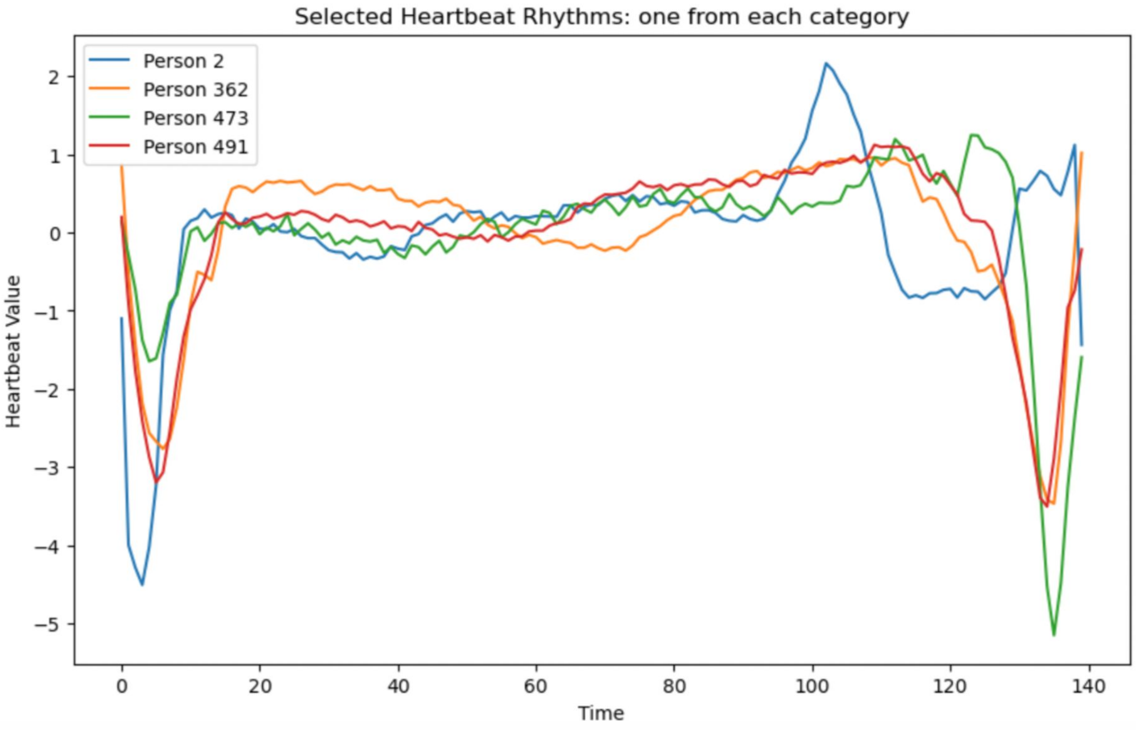


Figure 1: From three Consecutive Heartbeats to their corresponding persistence barcode and Betti curve

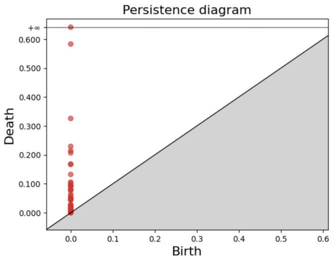
ECG 5000 Dataset

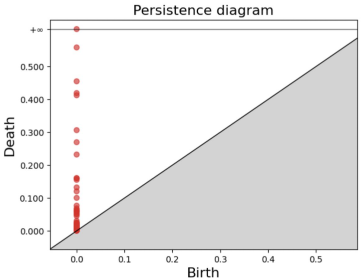
Goal: Classify 4 classes of arrhythmias using TDA

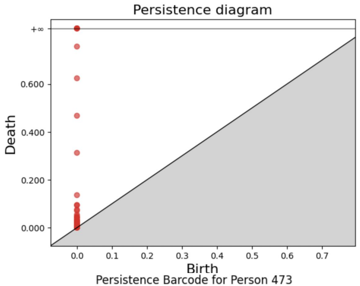
1. Persistence Diagrams
2. Persistence Images
3. Classification
   a. Regression
   b. LDA
   c. DTC

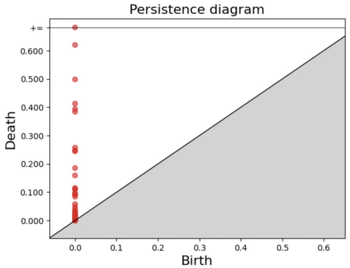


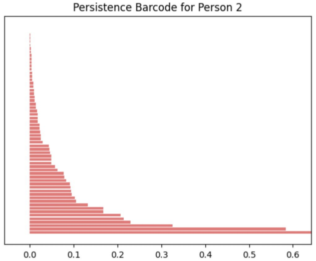
PDs and PIs

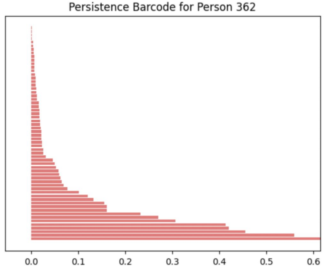

Persistence diagram
Death
Birth

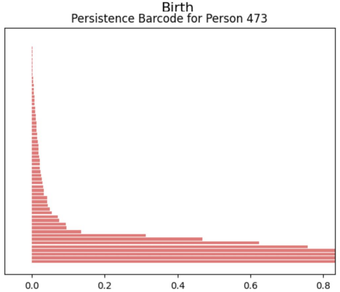

Persistence diagram
Death
Birth

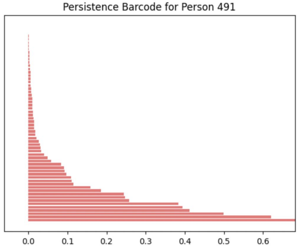

Persistence diagram
Death
Birth

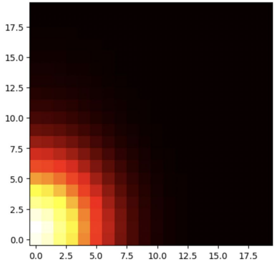

Persistence diagram
Death
Birth

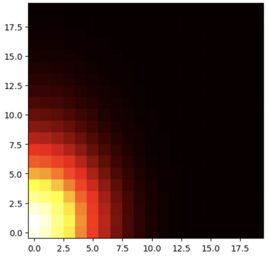

Persistence Barcode for Person 2

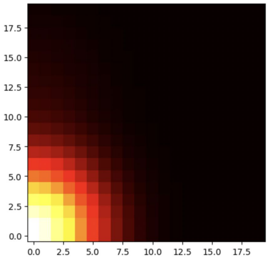

Persistence Barcode for Person 362

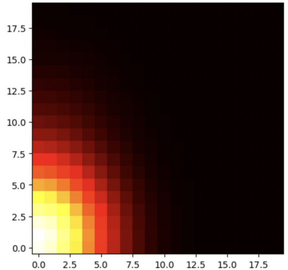

Persistence Barcode for Person 473


Persistence Barcode for Person 491









Classification Results

	TDA: Binary (full set)	TDA: Binary (modified set)	TDA: Multi Class (full set)	TDA: Multi Class (modified set)	Non TDA (full dataset only)
Logistic regression	89.7%	97.4%	80.0%	88.8%	94.0%
DTC		97.0%		88.0%	94.0%
LDA		100%		88.8%	92.7%

Future Research

- Non-aligned or multi-beat arrhythmias
- Tuning the persistence images
  - Weight functions
  - Resolution
- Multi-dimensional analysis