Topological Data Analysis for Arrhythmia Detection Through Modular Neural Networks

Meryll Dindin

Ecole Centrale Paris, France
meryll_dindin@berkeley.edu

Yuhei Umeda

Fujitsu Laboratories Ltd., Japan
umeda.yuhei@fujitsu.com

Frederic Chazal

INRIA, France
frederic.chazal@inria.fr

Heart diseases are one of the **leading cause** of deaths, there has been an unprecedented need for **monitoring** and dynamic **diagnosis**.

37s

Every **37s**, someone in the US dies of a heart condition. *

12%

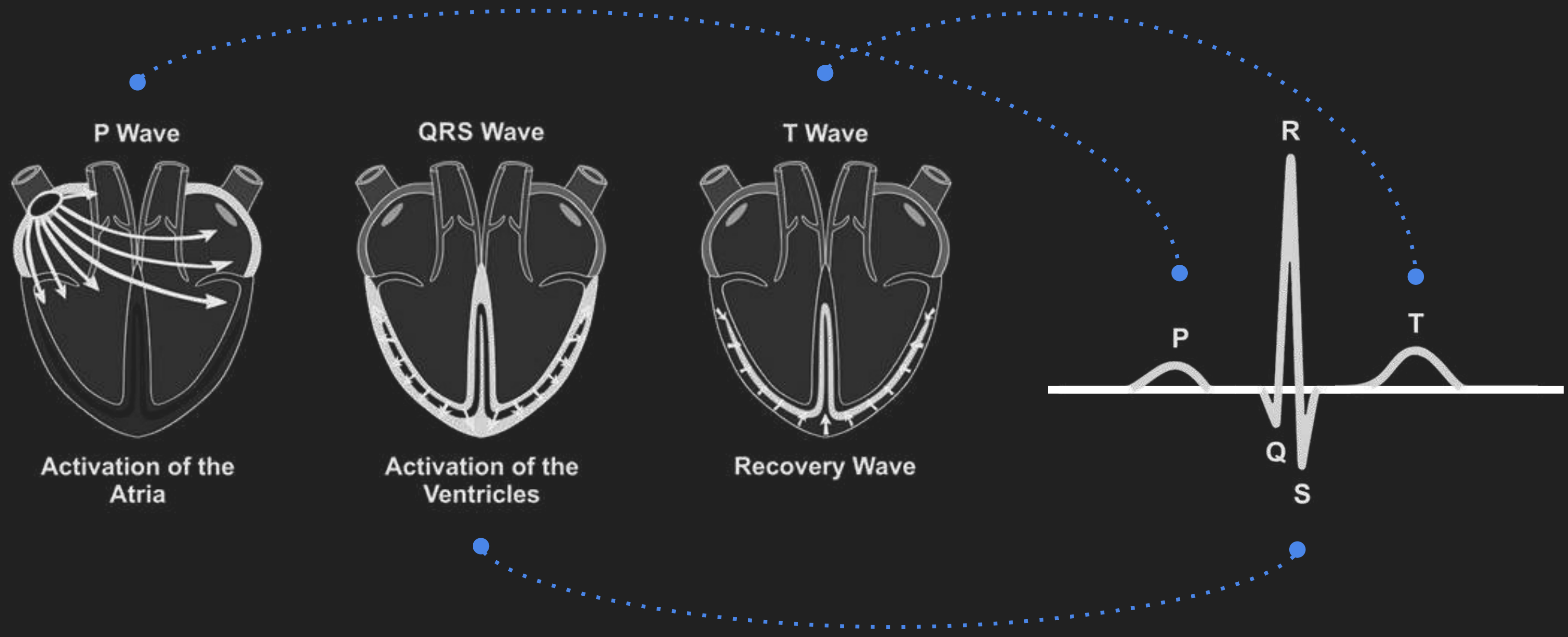
Of the US population is diagnosed with a heart condition. *

1 in 5

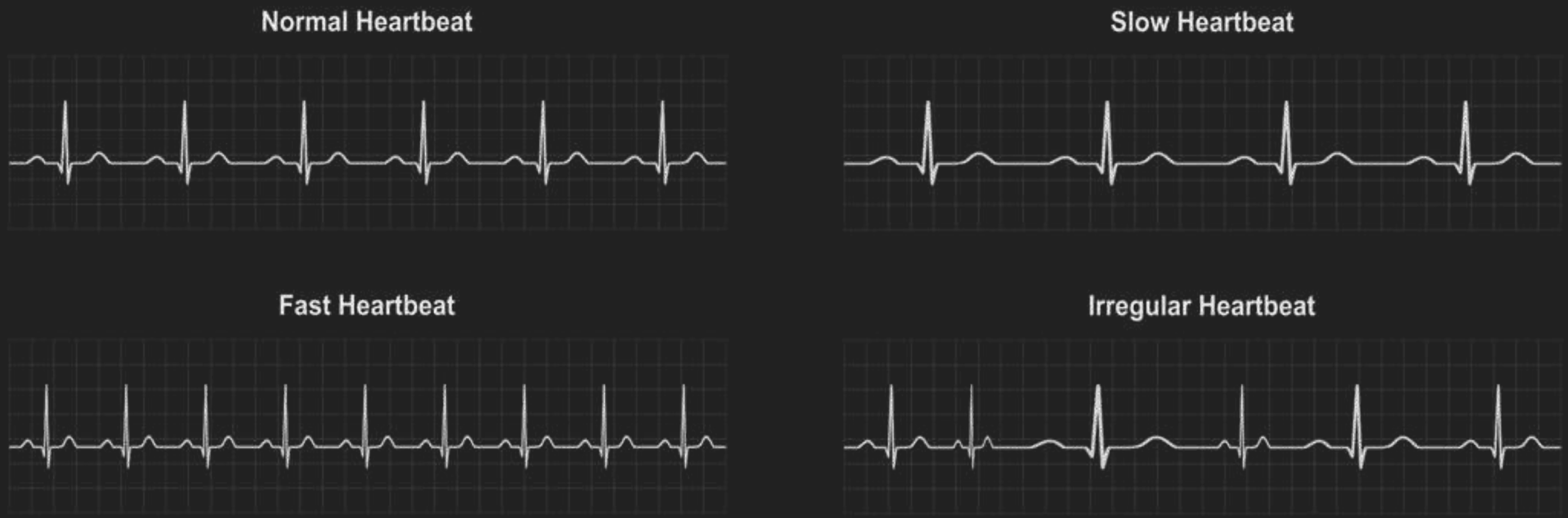
Of all heart attacks are silent, harming without telling. *

(*) Center for Disease Control and Prevention

A bit larger than the size of your fist, the heart is the **strongest organ** in your whole body. It also follows a very **recognizable pattern**.



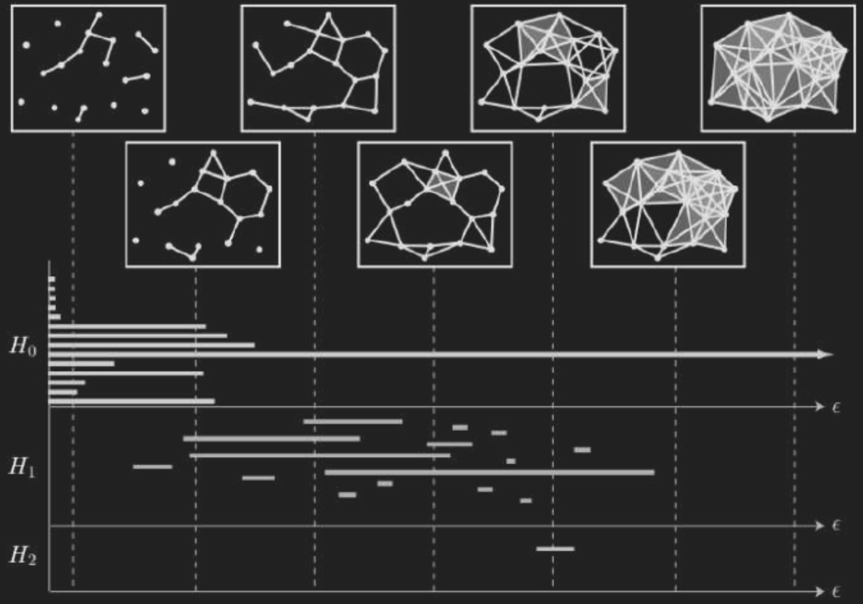
Arrhythmias are an umbrella term for a group of conditions describing **irregular heartbeats**, both in terms of **shape** or **frequency**.



(*) Credits: *https://www.carolinaheartandleg.com/arrhythmia*

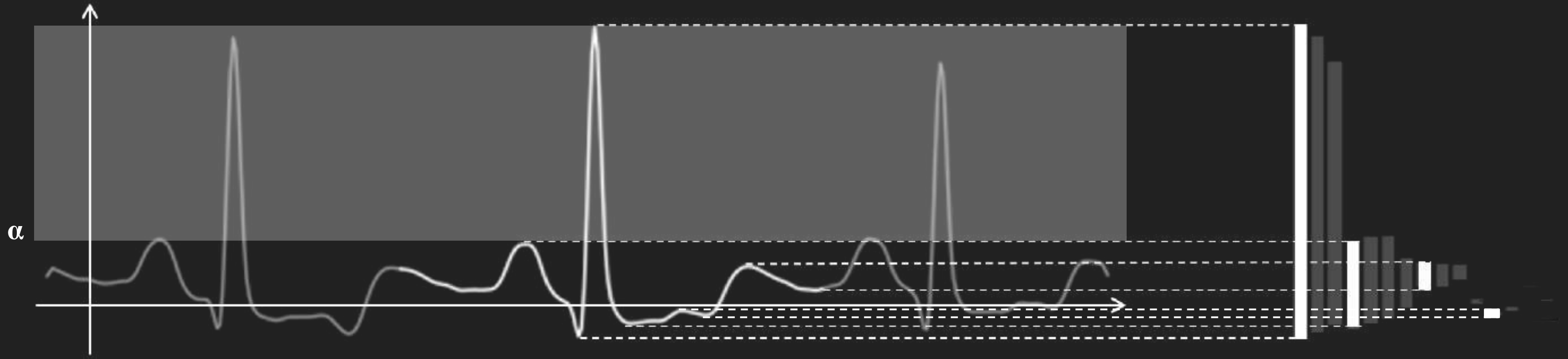
Individual differences make the problem hard to generalize on both **shape** and **frequency**. We introduce **Topological Data Analysis** to reduce the biases.

Topological Data Analysis is a recent and growing field providing mathematically well-founded methods to exhibit **topological patterns** in data and to encode them into **quantitative** and **qualitative** features. One of its foundational theories is known as **persistent homology**.



(*) Credits: Ghrist Robert, Barcodes: The Persistent Topology of Data, 2008

Using **Persistent Homology** enables us to robustly **encode** electrocardiograms, while **minimizing temporal differences** and **emphasizing spatial information**.



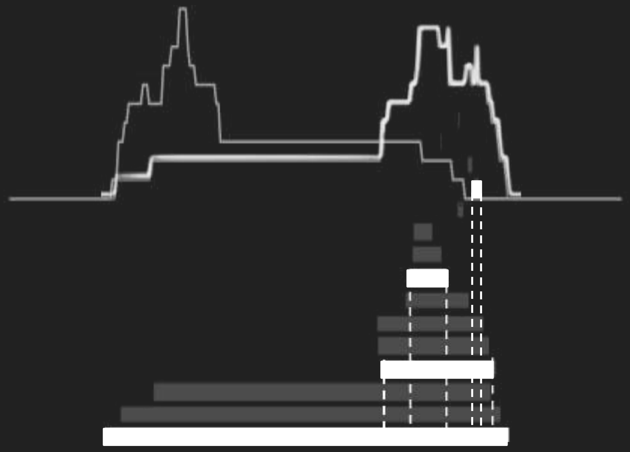
Persistent homology of the **upper-level sets filtration** of time series through the study of the **connected components** of:

$$F^{\alpha} = \{t \in I : f(t) \geq \alpha\}$$

Barcode Diagrams

Practically, **barcode diagrams** (aka ***sets***) do not interface well with **standard machine learning** techniques, bringing us to the construction of the **Betti curves**.

Betti Curves



$$F^\alpha = \{t \in I : f(t) \geq \alpha\}$$

Towards a better generalization

Theorem: Time Independence of Betti Curves

Given a function $f : I \to \mathbb{R}$ and a real number $a > 0$ the Betti curves of $t \to f(t)$ and $t \to f(at)$ are the same.

Moreover, if $g(t) = bf(t)$ for some $b > 0$, then the Betti curves of f and g are related by $BC_g(\alpha) = BC_f(\frac{\alpha}{b})$.

We gathered all the **open-sourced datasets** extensively studied in the literature to confront our **generalization** claims: the **MIT-BIH databases**.

Database	# Patients	# Labels
Arrhythmia	48	109494
Long Term	7	668735
Normal Sinus	18	1729629
Supraventricular Arrhythmia	78	182165
Malignant Ventricular	90	790565

Main Characteristics:

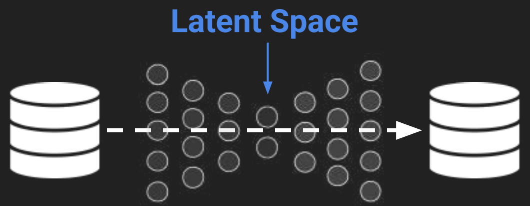
- ➔ **Two or more annotators**
- ➔ **ECGs sampled at 360 Hz**
- ➔ **ECGs have a 10 mV range**
- ➔ **Peaks are annotated**

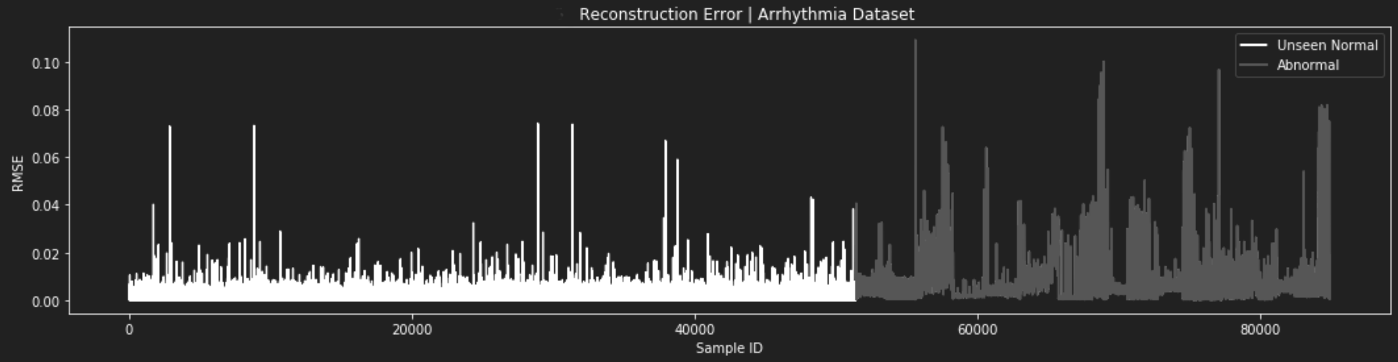
Preprocessing:

- **Resampling at 200 Hz**
- **Baseline Drift Removal**
- **FIR and Kalman Filters**
- **MinMax and Standard Scalings**
- **Heartbeats Dissociation**
- **Feature Engineering**

To help with **generalization**, we took advantage of the **heavy imbalance** between normal and abnormal heartbeats through a **fully-connected auto-encoder**.

Training of a **FC Auto-Encoder** on a CV patient-split of the heartbeats annotated as normal, to build a **new feature** through the **reconstruction error**.

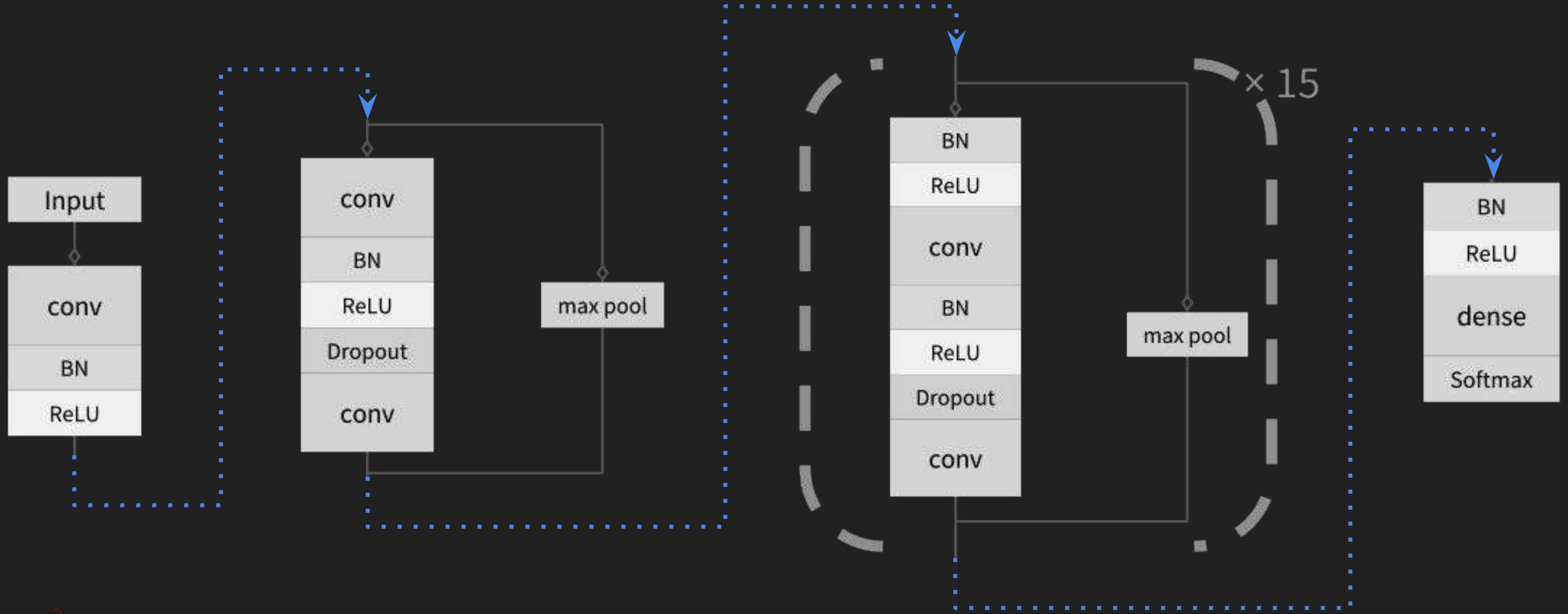




Useful as a feature, not powerful enough for a classification.

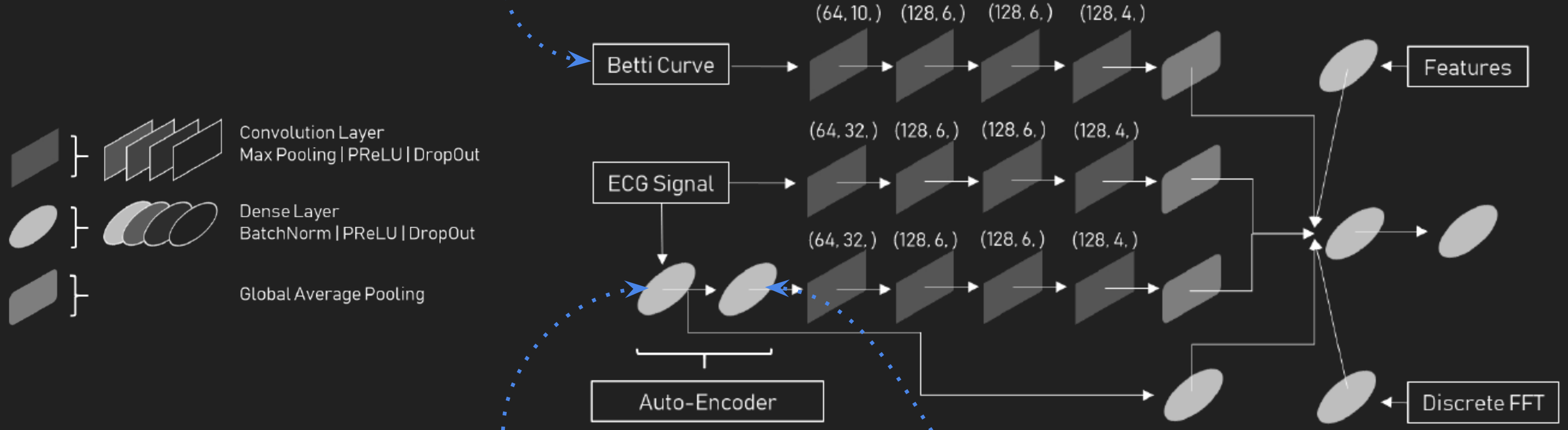
As traditional methods exploiting **Convolution Neural Networks** are increasingly getting **deeper**, we sought **multi-modality** and **shallow networks**.

Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks,
Pranav Rajpurkar*, Awni Hannun*, Masoumeh Haghpanahi, Codie Bourn, and Andrew Ng
2019



Literature and numerous experiments made us converge to a specific **multi-modal neural architecture**, exploiting both the **TDA** and the **auto-encoder latent space**.

Handle **Betti Curves** as 1D Time Series (controlled distributions)

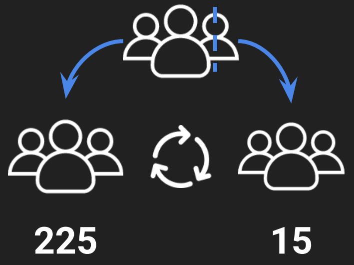


Use the **auto-encoder embedding** as a feature map

Use the **reconstruction error** as a 1D Time Serie

We first assessed the impact of using our **topological features** by turning on and off their respective channel, to better quantify their impact on **generalization**.

ID	Arrhythmia detection		Arrhythmia classification	
	With TDA	Without TDA	With TDA	Without TDA
0	**0.99**	0.98	**0.73**	0.68
1	**0.96**	0.90	**0.75**	0.69
2	0.85	**0.86**	**0.68**	0.65
3	0.94	**0.95**	0.95	**0.96**
4	**0.85**	0.80	**0.97**	0.97
5	**0.87**	0.77	**0.96**	0.93
6	0.78	**0.80**	**0.94**	0.93
7	**0.81**	0.63	**0.90**	0.80
8	**0.79**	0.65	**0.85**	0.78
9	0.84	**0.86**	**0.68**	0.47

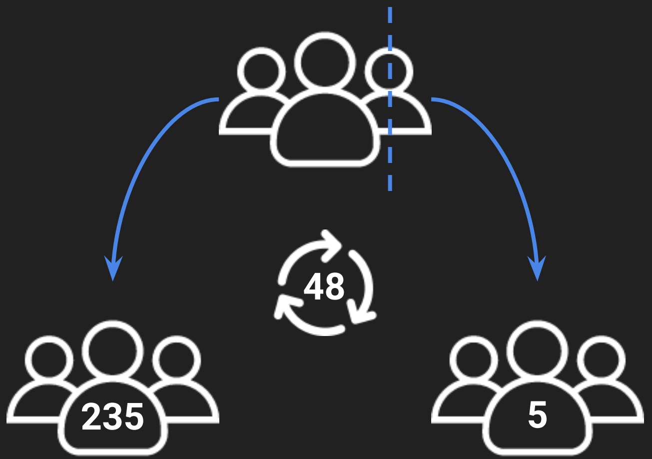


Experiment results:

Shows **generalization improvements** on both tasks of arrhythmia **detection** and arrhythmia **classification**.

To prove our performances in regards to both **imbalance** and **individual differences**, we had to also converge on a **rigorous methodology of tests**.

Initial 240 Patients



Train and **Test sets** as percentage splits of the same cohort

Validation sets as cohort of patients not selected for training

What it shows:

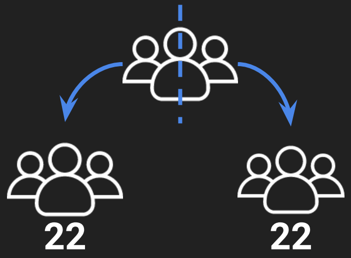
- → **Test** performances emphasize the ability to classify heartbeats on a **known population**
- → **Validation** performances highlight the generalization abilities on an **unknown population**

Experiment results on detection:

- → Mean **test** accuracy: **98%**
- → Mean **validation** accuracy: **90%**

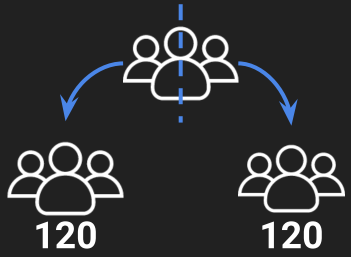
Our first **benchmarking** against literature was on the task of **detecting premature ventricular heartbeats** available in the **MIT-BIH databases**.

MIT-BIH Arrhythmia Database



Paper	Acc	PPV	Sen
Proposed	99.2%	95.1%	95.5%
Li et al.	99.4%	93.9%	98.2%

All Five MIT-BIH Databases



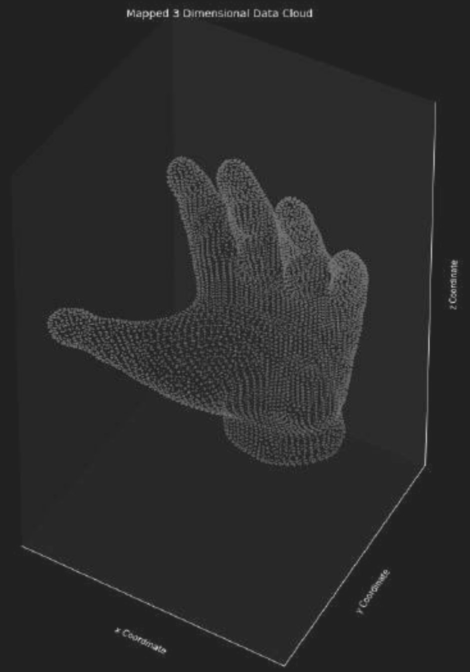
Paper	Acc	PPV	Sen
Proposed	96.1%	97.0%	96.0%
Li et al.	95.6%	94.1%	92.7%

The second **benchmark** comparison was done on a **8-classes classification** task, once again based on the **MIT-BIH Arrhythmia database**.

Paper	Classes	Acc	PPV	Sen
Proposed	8	99.0%	99.0%	98.5%
Jun et al. (2017)	8	99.0%	98.5%	97.8%
Kiranyaz et al. (2017)	5	96.4%	79.2%	68.8%
Güler et al. (2018)	4	96.9%	–	96.3%
Melgani et al. (2008)	6	91.7%	–	93.8%

Topological Data Analysis proved to be improving **generalization abilities** of our models. Its **modularity** characteristic opens up the path for **new problems**.

Aside from providing **high quality inputs** that are less likely to represent individual differences, **Topological Data Analysis** opens up a new interface for machine learning problems where data can be translated to **persistent diagrams** or **barcodes** (equivalent). **3D point clouds**, **graph structures**, **multi-modal time series** all enter this domain.



Topological Data Analysis for Arrhythmia Detection Through Modular Neural Networks

Meryll Dindin

Ecole Centrale Paris, France
meryll_dindin@berkeley.edu

Yuhei Umeda

Fujitsu Laboratories Ltd., Japan
umeda.yuhei@fujitsu.com

Frederic Chazal

INRIA, France
frederic.chazal@inria.fr