ML applications for MRI, ECG, and Plethysmography

Next_AIM General Meeting

Alberto Zucchetta

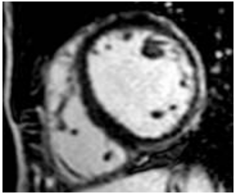
Istituto Nazionale di Fisica Nucleare - Sezione di Padova

February 14, 2023

Contents

Myocardial Fibrosis

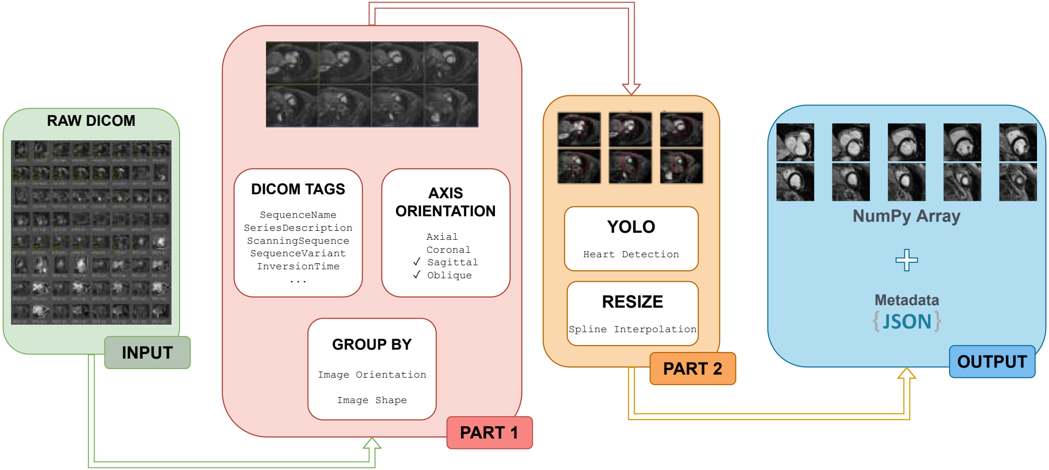
- This work is about the presence of Myocardial Fibrosis using Late Gadolinium Enhancement (LGE) technique from Cardiac Magnetic Resonance (CMR) images
- Goal: binary classification task aimed to identify LGE/Myocardial Fibrosis present/absent
- This project uses Cardiac Magnetic Resonance Images from Padova Hospital (Dott. Cipriani)



Dataset

- Dataset consists of 642 CMR scans ($>$ 200 Gb) plus annotations made by expert cardiologists in the form of an Excel file
- The presence of Myocardial Fibrosis is indicated alongside its location in the bullseye diagram of the heart
- Subjects are equally divided into LGE/Myocardial Fibrosis YES/NO, according to the presence/absence of scars.
- Weakly supervised problem:
  - no segmentation available
  - binary label per subject + binary label per each segment in bullseye diagram

- Raw DICOM files are preprocessed through an automated pipeline:



Federico Agostini - *Presented during EACVI/SCMR Joint Summit on AI in CMR (London, May 2022)*

LGE Data Extraction Pipeline

Myocardial Fibrosis

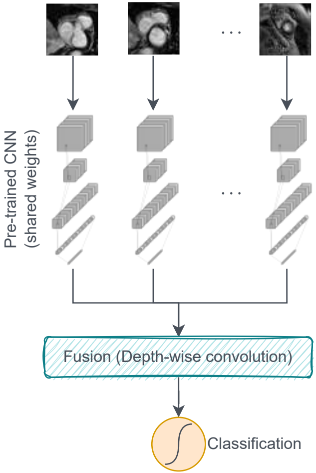
Part 1

- Scans are grouped together by image orientation (requesting a min and max number of elements per group) and only the group with the largest number of files is selected
- DICOMs are grouped by image shape (demanding a min number of elements), and only the series with the highest resolution is retained
- **Output**: 3D-array `(N,H,W)`, with `N` number of slices, and `(H,W)` image resolution (non-homogeneous between subjects)

Part 2

- The 3D-array is reshaped through a spline interpolation to the desired final number of slices and resolution
- The ROI (the heart) is extracted with a YOLO network [1] for object detection
  - the network is applied to all the slices; then the images are cropped by keeping the largest bounding box
- **Output**:
  - Images saved as Numpy arrays
  - Other Dicom metadata (e.g. weight, age, ...) stored in a JSON

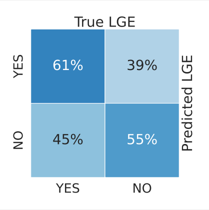
- Resut: a consistent dataset for subsequent analysis of 10 slices for each subject
- All images resized to 128 by 128 pixels
- Dataset is divided into training and test sets, with proportions 80%-20%
- Then, 3 different ML methods are applied (next slides)



Federico Agostini - *Presented during EACVI/SCMR Joint Summit on AI in CMR (London, May 2022)*

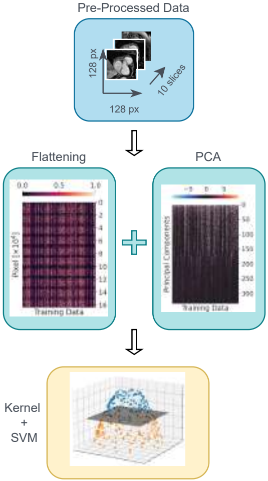
- The first analysis is based on CNN models, **pre-trained** on the ImageNet dataset.
- The training is done with **shared weights**
- Optimized the learning rate
- Early stopping and standard **data augmentation techniques**

Results



- Results show an **Accuracy of 58%** and a **Sensitivity of 58%**.

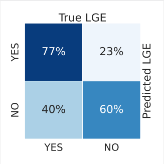
PCA + SVM
Myocardial Fibrosis



Federico Agostini - *Presented during EACVI/SCMR Joint Summit on AI in CMR*

- The second attempt is based on **Kernel Methods** and **SVM**
- Dimensionality reduction is implemented using a **Principal Component Analysis** (PCA) retaining 99% of the variance
  - the resulting 335 features are passed as input to a **SVM**
- Different Kernels (e.g. Linear, Gaussian, Cossim) are tested and models are trained and optimized using Grid Search with Cross-Validation

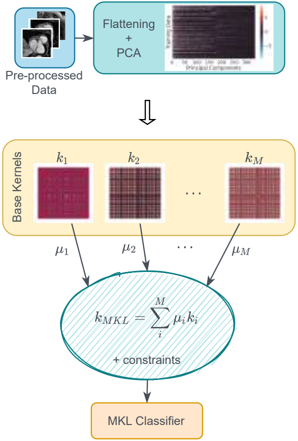
Results



- The best model is obtained with a **Gaussian Kernel**, and it displays **68% of Accuracy** and **60% of Sensitivity**.

PCA + MKL

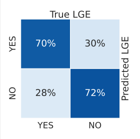
Myocardial Fibrosis



- Improved results could be obtained using state-of-the-art Multiple Kernel Learning (MKL) algorithms
- First, the dimensionality is reduced through PCA and then MKL is applied
- With this approach, the final Kernel is given by an optimal combination of base Kernels
  - top results are obtaining through a combination of multiple Gaussian Kernels
- Training is performed using Cross-Validation

Federico Agostini - *Presented during EACVI/SCMR Joint Summit on AI in CMR (London, May 2022)*

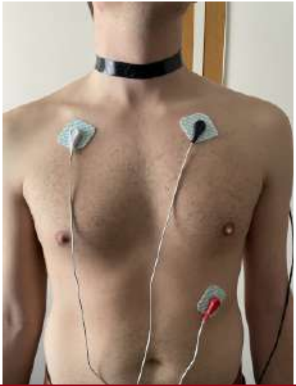
Results



- multiple Gaussian Kernels feature 71% Accuracy and 72% of Sensitivity

Plethysmography

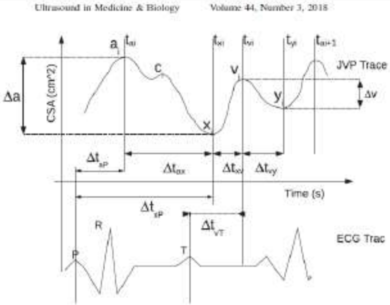
- A Plethysmograph is a device that measures changes of volume in different parts of the body, such as to arms, legs or other extremities to determine circulatory capacity
- Long time expertise in the Ferrara group [1], [2], [3]
  - Gianfranco Paternò, Antonino Proto, Daniele Conti, Erica Menegatti, Angelo Taibi, Giacomo Gadda
- *Drain Brain project:* monitor the cerebral venous outflow of a crew member during an experiment on the International Space Station (ISS)
- *Wise project:* develop a portable, non-invasive, and non-operator dependent device to measure venous blood volume on Earth
  - composed by a strain-gauge sensor, to be wrapped around the neck, that measures the variation in the neck circumference
  - the device can monitor the jugular venous pulse and the electrocardiogram synchronously



Plethysmography

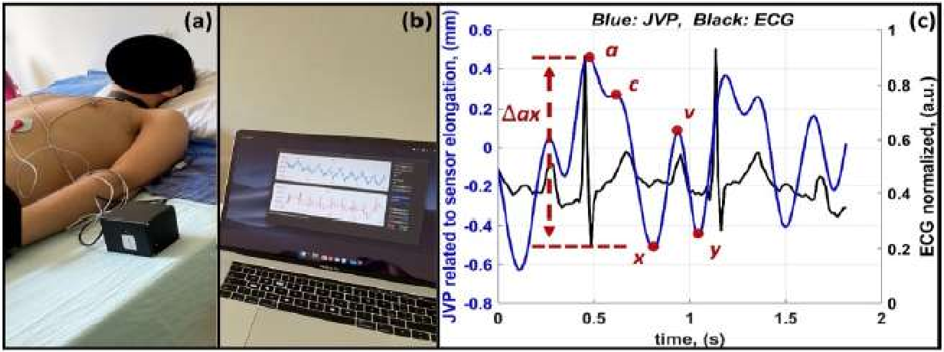
The relevance of plethysmography and its applications

- Plethysmography is used to investigate the jugular venous pulse (JVP) as an indicator of the vein cross-sectional area, and it is an index of the time variation of the blood volume in the IJV
- The JVP is one of the main parameters of cardiac status → used by cardiologists as a heart-failure indicator
- The JVP is commonly extrapolated via the ultrasound technique, but it requires a qualified healthcare operator to perform the real-time examination
- Instead, the device is wearable and can perform a continuous, non-invasive measurement of the JVP



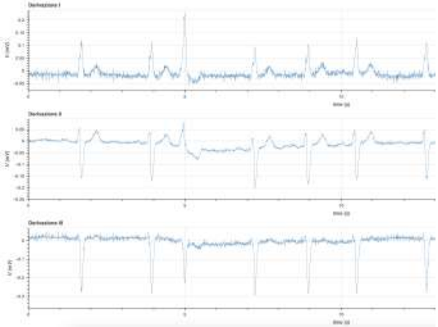
- The analysis of the signal aims to identify peaks *a* and *x*, which correspond to the atrial and ventricular contraction, respecively
- The pressure gradient between these two phases is correlated to Δax
- We also want to correlate Δax with the physical exercise of the patient: at rest (supine or sitting), or during walking or leg-pressing

- Current dataset consist of a 30-second symultaneous plethysmography + ECG measurements for 20 patients
- Analysis consists in artifact removal + signal filtering + manual labeling of the relevant features (*a* and *x*)
- Currently collaborating to apply ML techniques to automate the analysis (LSTM network) to:
  1. correlation with the ECG signal
  2. classification of the type of measurement (supine, sitting, moving)
  3. determination of the features of the signal (*a*, *x*, but possibly also *c*, *v* and *y*)
  4. automatically locate and remove artifacts (large difference between prediction and signal) from neck muscles, external causes



ECG analysis

- The (ambitious) goal of the study is to test if ML can reliably assess the risk of sudden cardiac arrest (SCA) in young athletes
- The project is based on 17k ECGs under stress recorded in the past years by the Medicina dello Sport in Treviso and Asolo (Medical coordinator: dott. Sarto)

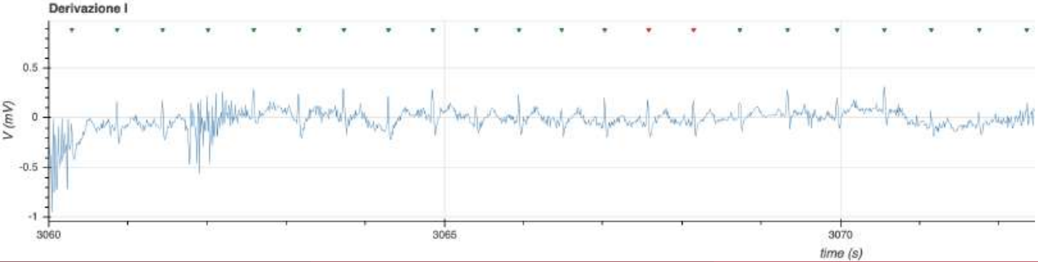


- The athletes have been monitored for several years, until the final dignosis (positive or negative)
- Possibility to track the evolution over time (years) of the ECG features
- The data sample is consistent and of good quality
- The procedure to extract the data has to be outsourced to MedRxiv and is work in progress

- link to the interactive plot

Holter analysis

- Assist cardiologists scrutinize the sizable amount of data (24 hours of continuous recording $\times$ 12 derivations)
- The dataset available to the University of Padova consists of approximately 450 Holter ECGs collected in the past decade by the Cardiology Department
- Use ML to isolate potentially anomalous and dangerous heartbeats, discriminating against normal beats and artifacts due to noise and patient movements
- Proper data and metadata extraction is not possible due to by poorly documented (if any) proprietary formats and softwares
  - link to the interactive plot



Future Outlook

- We have few projects that are work in progress, next steps:
  - Myocardial Fibrosis:
    - Document the studies and write a paper to be submitted to ArXiV
  - Plethysmography:
    - Test ML techniques for data analysis
    - Start with the simplest ones (classification)
    - Check if the current sample can be sufficient
    - Test data augmentation techniques
  - ECG:
    - If not possible to get the correct annotations for the Holter ECG, find a workaround, or try unsupervised learning
    - Conclude the export procedure to access the full Sports Medicine dataset
- The data accessibility is of paramount importance for a positive outcome of any project
  - and metadata is equally important

Backup slides

- Every time an Holter ECG is performed by a patient, the cardiologists need to inspect a large amount of data
  - 24 hours of continuous recording $\times$ up to 12 leads
  - the interesting part of the recording (the one that may show “real” anomalous beats) is only a small subset of the sample
  - plenty of artifacts due to noise and patient movements

- Commercial softwares can perform a preliminary analysis of anormal beats, but a very large fraction of them are false positive anomalies
  - machine learning could provide a better identification and classification of the anomalies, significantly reducing the false-positive ratio

Holter analysis

the challenges of the data format

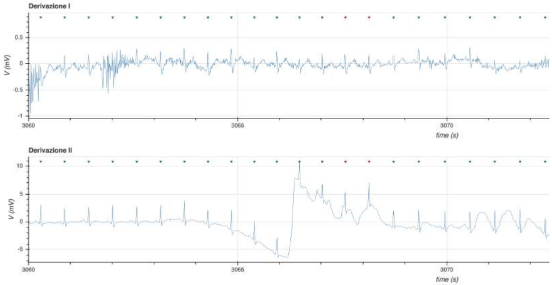
Data acquisition

- The dataset available to the University of Padova consists of approximately 450 Holter ECGs (anonymized) collected in the past decade by the Cardiology Department
- For each patient, the 24 hour raw data recording is exported directly from the Holter SD card (in total, $\approx$ half TB of data)
- Each heartbeat has been classified by the cardiologists by labelling by hand the real anomalous beats

The Mortara proprietary format

- the Holter instruments are provided by Mortara, which does not provide documentation of their (proprietary) formats
- the analysis software (HScribe), used to visually inspect the heartbeats and export data and metadata, is also provided by Mortara and closed-source
- most of our work so far has been dedicated to the data extraction, only with partial success

1. We were able to obtain from Mortara, under specific conditions, another of their experimental proprietary software (SuperECG) which converted the Mortara raw data to another open raw format (PhysioNet)
2. the PhysioNet raw data can be unpacked with the command line WFDB tool [1] in the SWIG toolbox, or the equivalent Python library [2]
3. the output .csv files can be explored with custom-made scripts [3]



link to the interactive plot

We documented our work in a public document [4] for future reference.

- The raw data extraction is successful
- However, without the metadata (annotations), the applications are limited
- Apparently, only the HScribe internal annotations can be exported, but the ones validated by the cardiologists are overwritten in the export process
- Highly dependent on the closed-source tools from Mortara, which lack of documentation and support