Laboratorio di Informatica Applicata ai Servizi Ospedalieri

Hands-on on medical data import and visualization

Alessandra Retico
alessandra.retico@pi.infn.it
INFN - Pisa

Overview

- General overview of these hands-on sessions

Apr. 9
Apr. 15
Apr. 26
Apr. 29
May 3

- Today's objectives:
  - Standard formats in Medical Imaging: DICOM and NIfTI
  - Download MANGO and visualize medical images
  - Image reading, visualization, anonymization

Medical image file formats

- The file format describes how the image data are organized in the file and how pixel data should be interpreted by a software for correct reading and visualization

- Numerical values of the pixels depend on image modality, acquisition protocol, reconstruction and applied post processing algorithms:
  - ➢ A medical image which is separated from the context information is meaningless

- Medical image file formats belong to two categories:
  - Those intended to standardize images generated by different diagnostic modalities, e.g. the DICOM standard
  - Those aiming to facilitate the post-processing analysis (e.g. NIfTI for neuroimaging)

DICOM file format

https://www.dicomstandard.org

- In response to the increased use of digital images in radiology the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) formed a joint committee in **1983** to create a **standard format for storing and transmitting medical images.**
- The committee published the original ACR-NEMA standard in **1985**.
- This standard has subsequently been revised and in **1993** it was renamed **DICOM**.

- **Digital Imaging and COmmunication in Medicine (DICOM):**
  - It is both a file format and communication protocol
  - It is the standard format used in digital imaging medical devices
  - The header and the image are contained in the same file
  - The header contains many information on the imaging device, the acquisition parameters, the patient and the physician
- DICOM supports most imaging modalities

alessandra.retico@pi.infn.it

DICOM file format

- DICOM is most commonly used for storing and transmitting medical images enabling the integration of medical imaging devices such as scanners, servers, workstations, printers, network hardware, and picture archiving and communication systems (PACS) from multiple manufacturers.

- Pixel data cannot be separated from the description of the image formation procedure:
  - Images should be self-descriptive

- The DICOM format ("anachronistically") stores volume as a sequence of 2D slices.
  - (A 3D DICOM format exists, but it is not widespread).
  - It is not that bad to have images stored slice by slice, as some acquisition parameters may change slice-wise during the acquisition.

- The DICOM format only stores integer numbers as pixel values, thus a *slope* and an *intercept* to linearly transform data in the allowed range are specified

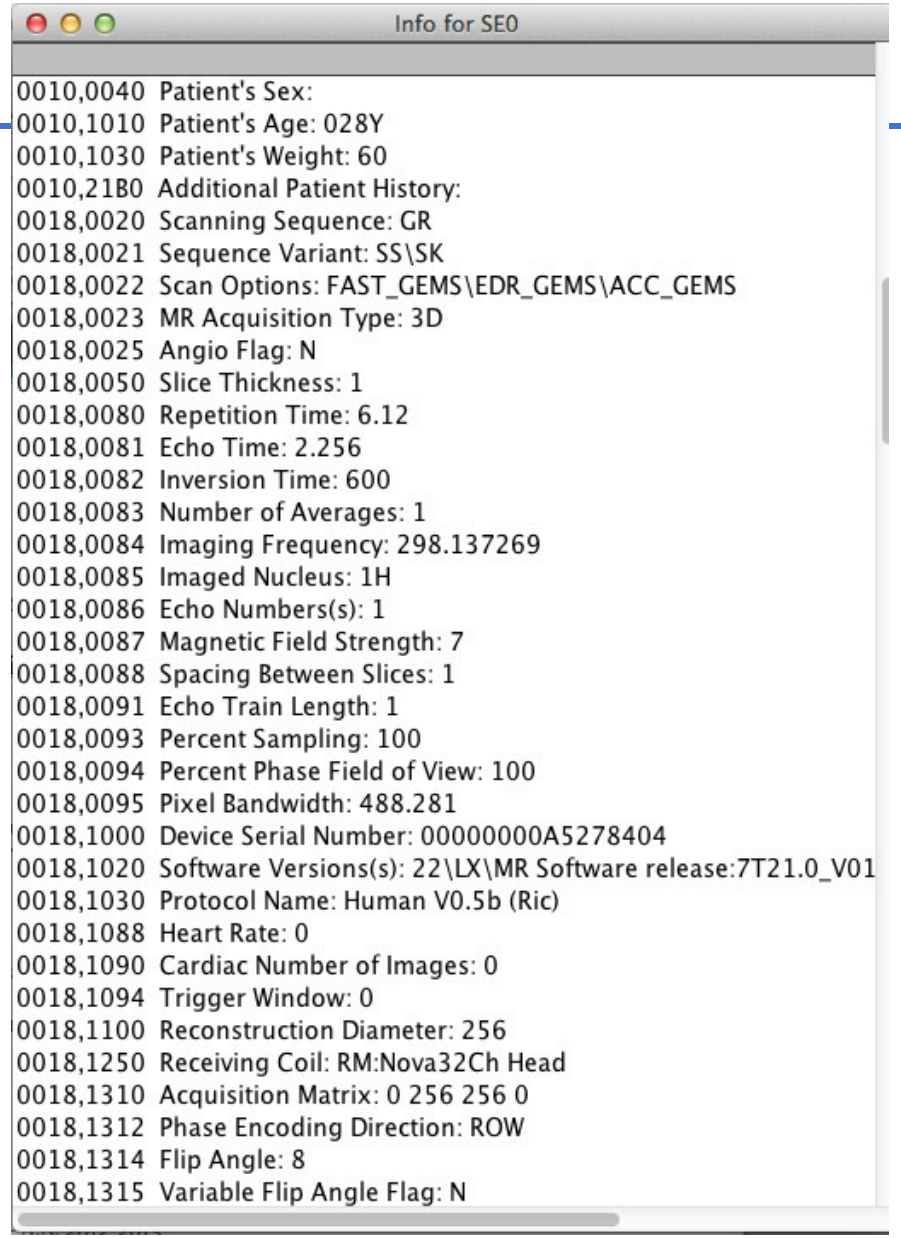
DICOM metadata

A **DICOM data element, or attribute,** is composed of the following most important parts:

- a **tag** that identifies the attribute, usually in the format (XXXX,XXXX) with hexadecimal numbers, and may be divided further into DICOM Group Number and DICOM Element Number;
- a **DICOM Value Representation (VR)** that describes the data type and format of the attribute value (e.g. PN: Person Name; DT: Date Time; SS: Signed Short).

Check **https://www.dicomlibrary.com/dicom/dicom-tags/**

- most fields are optional
- there are also vendors' private keys
- some examples:
  - Session Name and Study Number
    - (0008, 0090) ID Referring Physician
    - (0010,0010) Patient's Name
  - Image "Shape"
    - (0028, 0010) IMG Rows
    - (0028, 0011) IMG Columns
    - (0028, 0030) IMG Pixel Spacing
    - (0018, 0050) ACQ Slice Thickness
  - How and where the image data is stored
    - (0028, 0100) IMG Bits Allocated
    - (0028, 0101) IMG Bits Stored
    - (0028, 0102) IMG High Bit



The Neuroimaging Informatics Technology Initiative (NIfTI) file format

Another file format commonly used to store brain imaging data obtained using Magnetic Resonance Imaging methods is the Neuroimaging Informatics Technology Initiative (**NIfTI**)

- The NIfTI stored volumes can be:
  - in dual file format: file.hdr, file.img
  - in a single file: file.nii

- NIfTI is the default file format of most software packages for neuroimaging post-processing:
  - FSL, SPM, itk-SNAP, 3D Slicer, ITK & VTK, nipy, etc.

- The NIfTI format allows a double way to store the orientation of the image volume in the space:
  1. rotation + translation to be used to map voxel coordinates to the scanner reference frame
  2. 12-parameter or more general transformation adopted to realign the volume to a standard template coordinate system.

Medical image viewers

- ImageJ
  - https://imagej.nih.gov/ij/
- OsiriX (only for Mac, iPhone, iPad)
  - https://www.osirix-viewer.com
- 3DSlicer
  - https://www.slicer.org
- itk-SNAP
  - http://www.itksnap.org/pmwiki/pmwiki.php
- MANGO:
  - http://ric.uthscsa.edu/mango/index.html
- …

Images viewers

- MANGO: http://ric.uthscsa.edu/mango/index.html

Mango for the Desktop

Mango supports Mac, Windows, and Linux operating systems.

Download for **Mac** (v4.1)

Download for **Windows** (v4.1)

Download for **Linux** (v4.1)

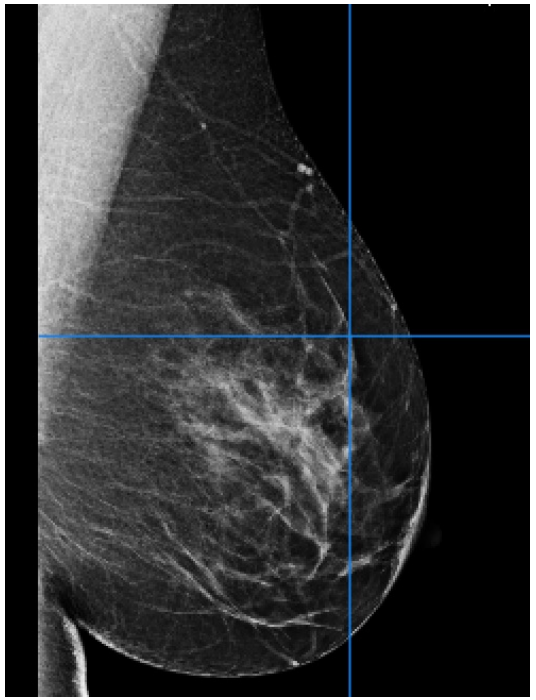
Download Previous Versions

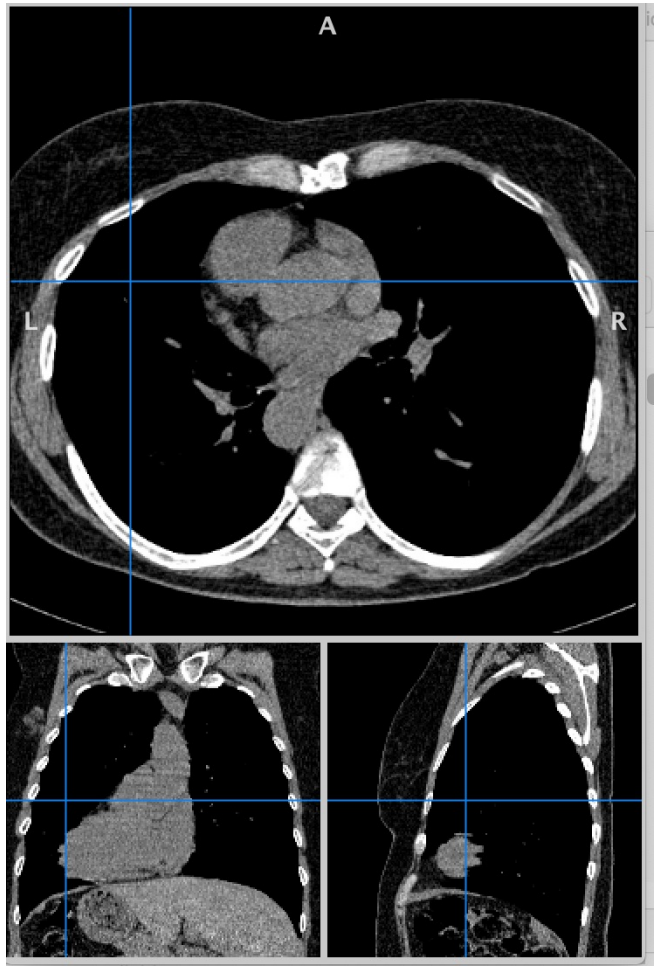
License

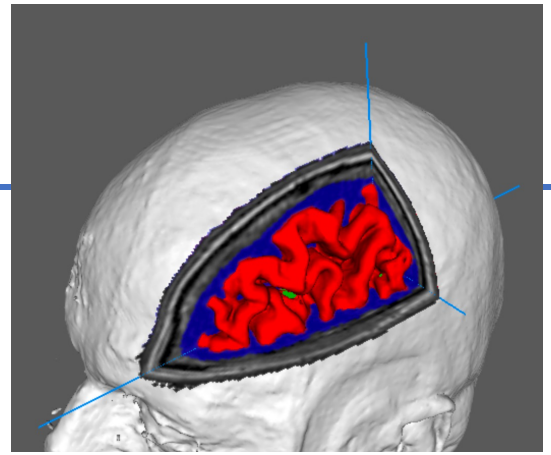
Mango software is copyrighted by the University of Texas. Mango is available for use, as is, free of charge, for educational and scientific, non-commercial purposes. The software and data derived from Mango software may be used only for research and may not be used for clinical purposes. If Mango software or data derived from Mango software is used in scientific publications, we request that the 'Research Imaging Institute, UTHSCSA' be cited as a reference. Use or incorporation of this software or data derived from this software in commercial applications or commercial publications requires a written consent from the RII-UTHSCSA.

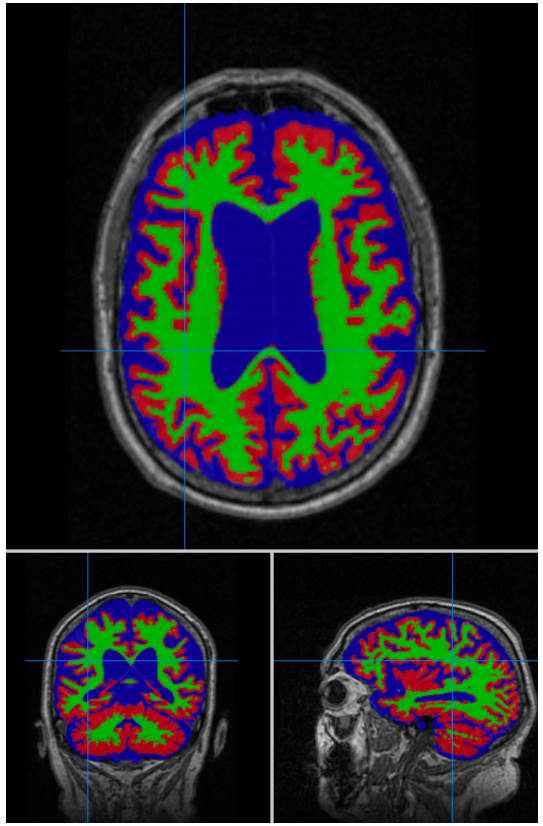
Let’s read real images with Mango …

- Explore 2D, 3D and 4D data sample
- Make image overlays









alessandra.retico@pi.infn.it

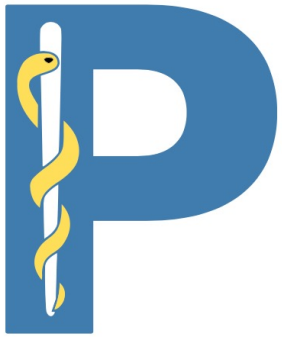
Sample DICOM/NIfTI images

- You can download the DATASETS directory, which contains a sample of publicly available medical images and data, from:
  - https://drive.google.com/open?id=1YqK7ZkM-P2IrqfD7Pj-SCmjz-GWd_1-Y

- Example of open repositories of medical data
  - https://openneuro.org
  - http://fcon_1000.projects.nitrc.org
  - https://nihcc.app.box.com/v/ChestXray-NIHCC
  - https://www.cancerimagingarchive.net
  - http://www.oasis-brains.org
  - http://adni.loni.usc.edu

- Additional data samples (medical images/image features) for hands-on sessions will be distributed later during the course

The pydicom library



- Pydicom is a pure Python package for working with DICOM files such as medical images, reports, and radiotherapy objects.
- Pydicom makes it easy to read these complex files into natural pythonic structures for easy manipulation.

https://pydicom.github.io/pydicom/

- Requirements:
  - numpy library is recommended, (it is only required if manipulating pixel data)
  - matplotlib is necessary to visualize data

See demo code at https://github.com/retico/SSFM_2021

Jupyter notebooks:
- Demo1_read_DICOM_file.ipynb (read and visualize dicom files, e.g. a 2D sclice)
- Demo2_read_DICOM_dir.ipynb (read and visualize dicom dirs, e.g. a 3D volume)

The General Data Protection Regulation (GDPR)

- The EU General Data Protection Regulation (GDPR) was approved by the EU Parliament on 14 April 2016 https://eugdpr.org
- It is designed to:
  - Harmonize data privacy laws across Europe
  - Protect and empower all EU citizens data privacy
  - Reshape the way organizations across the region approach data privacy

- GDPR and Data Science
  - De-identifying medical imaging is a fundamental prerequisite for data storing, processing and sharing within research projects in order to be compliant with GDPR:
    - Anonymization: using the Hash function (non-reversible)
    - Pseudo-anonymization: data is tokenized, a separate lookup file (with the original entry and the token) is generated and stored in a restricted database.

See demo code (Jupyter notebooks):

- Demo3_anonymize.ipynb (how to anonymize dicom files)