Diagnostic imaging

MTAT.03.242 Bioinformatics Seminar
March 27th, 2017
by Lauri Listak
Supervisor: Dmytro Fishman
Topics covered

1. Where does the data come from?
2. What can we do with it?
3. What does the data look like?
4. How to gain access to it?
5. Homework!
Why would this even be useful?
Why computers could help?

The major assumption in medical imaging is that you can derive high-order information (phenotypical) from studying pixel images of an object (cell, tissue, organ, molecule etc.).

Also research attempts to cluster these objects into distinct phenotypical groups (cancer/non-cancer) by studying their appearances.
What can automation do?

- CADe - Computer Aided Detection - raise alarm
- CADx - Computer Aided Diagnosis - predict likelihood
- Content-based image retrieval - show me similar cases
- Assess image quality - what should actually be shown?
- Brain mapping - find out the structure of the brain

For all that, you need to go through the usual data science pipeline.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4220564/
What kind of technology is used?
X-ray imaging

2 types:

- “Usual” X-ray
- CT-scan

Lots of advancements: helical scanning, multiple slices, etc

https://www.fda.gov/Radiation-EmittingProducts/RadiationEmittingProductsandProcedures/MedicalImaging/MedicalX-Rays/ucm115318.htm
Ultrasound

Actually work in the MHz range

Can also be done inside the body

Diagnostic vs functional vs therapeutic

MRI

No x-rays, but a strong magnetic field is used

Used for non-bony parts, soft tissues. Therefore, brain, nerves, muscles, ligaments, etc can be imaged

PET

Positron emission tomography

Look at blood flow, metabolism, neurotransmitters, and the effect of marked drugs

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1126321/
Methods of analysis

A few mentioned things were:

- SVMs
- Deep neural networks

However, problems abound:

- New data will not be from the same distribution as training+test+validation
- Diagnostics itself is actually a hard problem to solve
The data itself!
DICOM

- Used for lots of purposes, started in radiology
- Metadata is included in the file
- Various standards for encoding pixel data - e.g., JPEG variations

Watch out for the real world stuff, like never using SliceThickness, but instead computing it from other values.

The data dictionary at [http://dicom.nema.org/dicom/2006/06_06pu.pdf](http://dicom.nema.org/dicom/2006/06_06pu.pdf) for fields is 106 pages long.
Nifti

Analyze -> Nifti-1 -> Nifti-2

Used for neuroimaging

Allows for double orientation of the image volume: rotation+translation to scanner frame of reference and 12 parameters for a mapping to standard coordinate system

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3948928/
<table>
<thead>
<tr>
<th>Tag Name</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SliceThickness</td>
<td>0018,0050</td>
<td>8.836</td>
</tr>
<tr>
<td>CountsAccumulated</td>
<td>0018,0070</td>
<td>0</td>
</tr>
<tr>
<td>SpacingBetweenSlices</td>
<td>0018,0088</td>
<td>10.000</td>
</tr>
<tr>
<td>SoftwareVersions</td>
<td>0018,1020</td>
<td>Y3.9</td>
</tr>
<tr>
<td>StudyInstanceUID</td>
<td>0020,000D</td>
<td>1.2.752.37.1.1.291650441.4.4833</td>
</tr>
<tr>
<td>SeriesInstanceUID</td>
<td>0020,000E</td>
<td>1.2.752.37.1.1.291650441.5.7962</td>
</tr>
<tr>
<td>StudyID</td>
<td>0020,0010</td>
<td>06036567NUTIWB</td>
</tr>
<tr>
<td>SeriesNumber</td>
<td>0020,0011</td>
<td>1</td>
</tr>
<tr>
<td>AcquisitionNumber</td>
<td>0020,0012</td>
<td>1</td>
</tr>
<tr>
<td>InstanceNumber</td>
<td>0020,0013</td>
<td>1</td>
</tr>
<tr>
<td>ImagesInAcquisition</td>
<td>0020,1002</td>
<td>1</td>
</tr>
<tr>
<td>ImageComments</td>
<td>0020,4000</td>
<td>WHOLE BODY TL201</td>
</tr>
<tr>
<td>SamplesPerPixel</td>
<td>0028,0002</td>
<td>1</td>
</tr>
<tr>
<td>PhotometricInterpretation</td>
<td>0028,0004</td>
<td>MONOCHROME2</td>
</tr>
<tr>
<td>NumberofFrames</td>
<td>0028,0008</td>
<td>16</td>
</tr>
<tr>
<td>FrameIncrementPointer</td>
<td>0028,0009</td>
<td>8388692</td>
</tr>
<tr>
<td>Rows</td>
<td>0028,0010</td>
<td>64</td>
</tr>
<tr>
<td>Columns</td>
<td>0028,0011</td>
<td>64</td>
</tr>
<tr>
<td>PixelSpacing</td>
<td>0028,0030</td>
<td>8.836</td>
</tr>
<tr>
<td>value</td>
<td>8.836</td>
<td></td>
</tr>
<tr>
<td>PixelAspectRatio</td>
<td>0028,0034</td>
<td>1</td>
</tr>
<tr>
<td>value</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>BitsAllocated</td>
<td>0028,0100</td>
<td>16</td>
</tr>
<tr>
<td>BitsStored</td>
<td>0028,0101</td>
<td>16</td>
</tr>
<tr>
<td>HighBit</td>
<td>0028,0102</td>
<td>15</td>
</tr>
<tr>
<td>PixelRepresentation</td>
<td>0028,0103</td>
<td>0</td>
</tr>
<tr>
<td>SmallestImagePixelValue</td>
<td>0028,0106</td>
<td>0</td>
</tr>
<tr>
<td>LargestImagePixelValue</td>
<td>0028,0107</td>
<td>245</td>
</tr>
<tr>
<td>WindowCenter</td>
<td>0028,1050</td>
<td>122</td>
</tr>
<tr>
<td>WindowWidth</td>
<td>0028,1051</td>
<td>245</td>
</tr>
<tr>
<td>LossyImageCompression</td>
<td>0028,2110</td>
<td>00</td>
</tr>
<tr>
<td>NumberofEnergyWindows</td>
<td>0054,0011</td>
<td>1</td>
</tr>
<tr>
<td>EnergyWindowInformationSequence</td>
<td>0054,0014</td>
<td>59.580.5</td>
</tr>
<tr>
<td>Item</td>
<td>59.580.5</td>
<td></td>
</tr>
<tr>
<td>EnergyWindowRangeSequence</td>
<td>59.580.5</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>59.580.5</td>
<td></td>
</tr>
<tr>
<td>EnergyWindowLowerLimit</td>
<td>0054,0015</td>
<td>59.5</td>
</tr>
<tr>
<td>EnergyWindowUpperLimit</td>
<td>0054,0015</td>
<td>80.5</td>
</tr>
<tr>
<td>RadiopharmaceuticalInformationSequence</td>
<td>0054,0015</td>
<td>59.580.5</td>
</tr>
</tbody>
</table>
Recoding to various formats for use

- JPEG - lossy
- TIFF - lossless
- PNG - can be both (set compression level)
How would I get my hands on this kind of data?
Public sources

UK Biobank seems to have released a few thousand MRI images: http://biobank.ctsu.ox.ac.uk/crystal/label.cgi

54000 searchable images available at https://medpix.nlm.nih.gov/home

Detection competitions for your interest are an option: for example, LUNA16 and Kaggle Data Science Bowl 2017 for lung cancer
Homework

Choose 1 of 2 alternatives.

Send the results to lauri@listak.net by the deadline specified in the seminar guidelines
Homework variant I

Find a type of diagnostic imaging not mentioned in this presentation and describe it in about 100 words: what’s the physical/chemical background, what is it used for, what are the challenges?
Homework variant II

Find a suitable package for your favorite programming language and write out the command(s) for reading in a DICOM file and displaying it. Send me:

1. The DICOM file you used
2. The source code you used
3. Screenshot of the final result
Thank you!