

# Medical Informatics

## Lecture 14: Imaging and text data in Medicine

Dr Areti Manataki



Nanjing Medical University

# Copyright

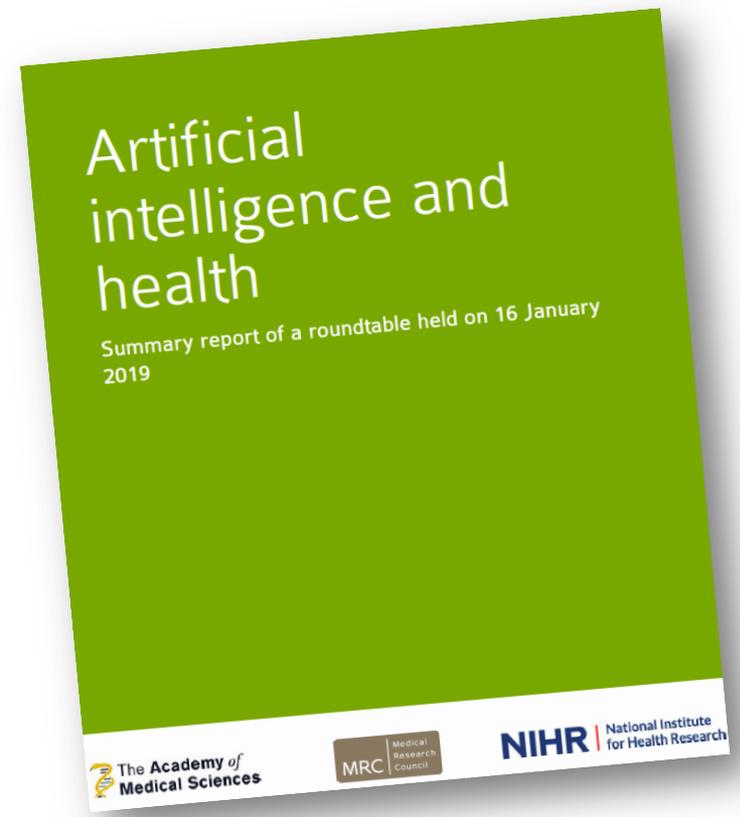
Some of the content of this lecture is based on the Data Science in Stratified Healthcare and Precision Medicine course on Coursera, which was developed by The University of Edinburgh under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license.

These lecture slides are also licensed under a CC BY-SA 4.0 license.



# Beyond tabular data

*“AI can add substantial value to learning systems that have cryptic hard to define features, such as those found in **images**, free **text** or **video** footage.”*



# Medical Imaging

# Medical imaging

- Medical imaging allows us to create visual representations of the interior of the body for clinical analysis and medical intervention.
  - To reveal internal structures hidden by the skin and bones
  - To diagnose and treat disease
- Medical imaging in the data science era:
  - unprecedented growth in the volume of medical imaging data
  - challenges, due to the complexity and the poor quality of imaging data

# Medical imaging

- How do we acquire medical images?
- What is the format of medical images?
- How do we store and access medical images?
- How do we analyse medical images?

# Medical imaging

- How do we acquire medical images?
- What is the format of medical images?
- How do we store and access medical images?
- How do we analyse medical images?

# Medical imaging modalities

- X-ray imaging
- Computed Tomography (CT)
- Magnetic Resonance Imaging (MRI)
- Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT)

# Medical imaging modalities

- X-ray imaging
  - the oldest medical imaging technique
  - Dense parts of the body show up as clear white areas on the image, while softer parts show up as darker areas.



*[Image by Louis Philippe Lessard  
on Wikimedia, licensed as CC BY-SA  
3.0]*

# Medical imaging modalities

- X-ray imaging
- Computed Tomography (CT)
  - Also called Computed Axial Tomography (CAT)
  - A type of slice imaging: imaging by sections or slices through the body
  - It can produce detailed images of many structures in a thin section of the body.

# Medical imaging modalities

- X-ray imaging
- Computed Tomography (CT)
- Magnetic Resonance Imaging (MRI)
  - Can be used to examine almost any part of the body
  - MRI vs. CT:
    - MRI has excellent soft tissue contrast. CT is preferred for lung and bone imaging.
    - CT is faster and it is more widely used than MRI.
    - CT scanning is associated with increased risk of cancer. MRI scanning does not have such health hazards.

# Medical imaging modalities

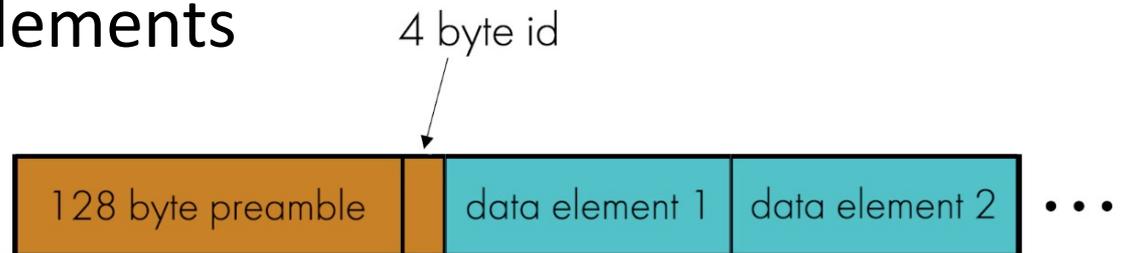
- X-ray imaging
- Computed Tomography (CT)
- Magnetic Resonance Imaging (MRI)
- Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT)
  - nuclear imaging: assessment of physiology rather than anatomy
  - Modern scanners may integrate PET or SPECT with other modalities.

# Medical imaging

- How do we acquire medical images?
- **What is the format of medical images?**
- How do we store and access medical images?
- How do we analyse medical images?

# DICOM format

- DICOM (Digital Imaging and Communications in Medicine) is an international standard for storing, exchanging and transmitting medical images.
- DICOM file content:
  - image pixel data
  - header of fixed length, followed by a sequence of tagged data elements



*[Image from  
DataSciMed MOOC]*

# DICOM format

- DICOM (Digital Imaging and Communications in Medicine) is an international standard for storing, exchanging and transmitting medical images.
- DICOM file content:
  - image pixel data
  - header of fixed length, followed by a sequence of tagged data elements
    - The header contains information about the file, the imaging equipment used, the study and the patient that it belongs to.

# Medical imaging

- How do we acquire medical images?
- What is the format of medical images?
- **How do we store and access medical images?**
- How do we analyse medical images?

# PACS

- Medical images are created, stored, accessed and processed in the restricted environment of a hospital.
- They are typically kept in a **Picture Archiving and Communication System (PACS)**.
- A PACS archives medical images (and the associated meta-information) within radiology and distributes them to departments that ordered the images.

# Medical imaging

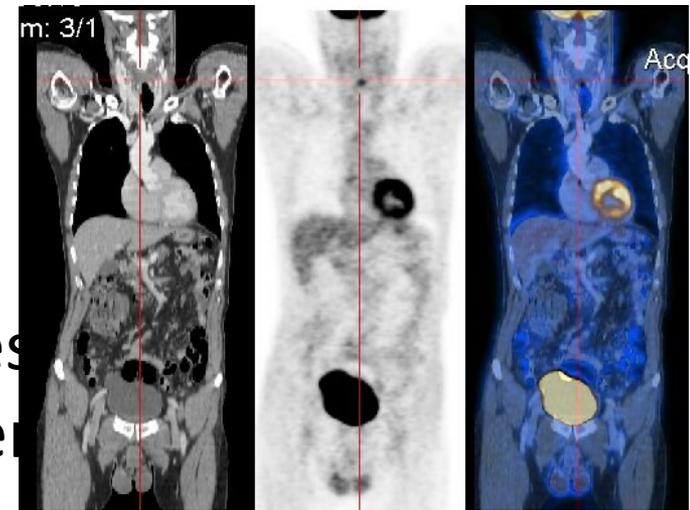
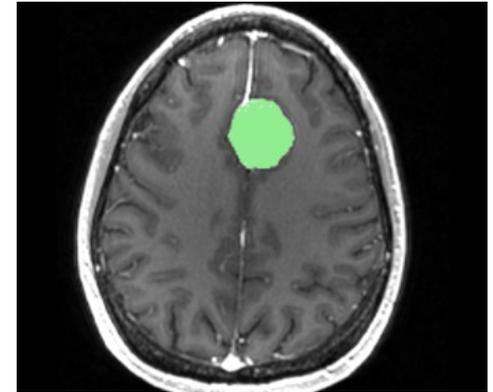
- How do we acquire medical images?
- What is the format of medical images?
- How do we store and access medical images?
- How do we analyse medical images?

# Medical imaging pre-processing

- We typically **enhance** the image before analysing it
  - to enhance contrast by reducing any artefacts or noise in the image or by emphasising differences between objects (e.g. enhancing edges)
- **Feature detection**: may involve edge tracking, corner detection and template matching, e.g. blob detection

# Medical imaging analysis

- **Segmentation**: to partition an image into different meaningful segments, such as particular organs, different tissue classes or pathologies
- **Registration**: to compare or combine different images by determining a one-to-one mapping between the coordinates in one image and those in another



*[Images by Rkikinis and MBq on Wikimedia, licensed as CC BY-SA 3.0]*

# Segmentation techniques

- Thresholding:
  - the simplest segmentation method
  - it works particularly well when there is a good level of contrast between different structures
- Region growing
  - domain knowledge: the user specifies a seed point in an object of interest that needs to be separated
- Region merging
- Watershed transform
- Live wire

# Natural Language Processing



# Natural Language Processing

- **Natural Language Processing** is a set of computational methods to analyse, understand, and derive meaning from human language.
- Typical tasks:
  - sentence boundary detection
  - tokenisation
  - part-of-speech tagging
  - named entity recognition

# Natural Language Processing

- **Natural Language Processing** is a set of computational methods to analyse, understand, and derive meaning from human language.
  - Typical tasks:
    - sentence boundary detection
    - tokenisation
    - part-of-speech tagging
    - named entity recognition
- The patient seemed concerned. I explained the different treatment options.

# Natural Language Processing

- **Natural Language Processing** is a set of computational methods to analyse, understand, and derive meaning from human language.

- Typical tasks:

- sentence boundary detection
- tokenisation
- part-of-speech tagging
- named entity recognition

*S1*  
The patient seemed concerned. I *S2*  
explained the different treatment  
options.

# Natural Language Processing

- **Natural Language Processing** is a set of computational methods to analyse, understand, and derive meaning from human language.
- Typical tasks:
  - sentence boundary detection
  - tokenisation
  - part-of-speech tagging
  - named entity recognition

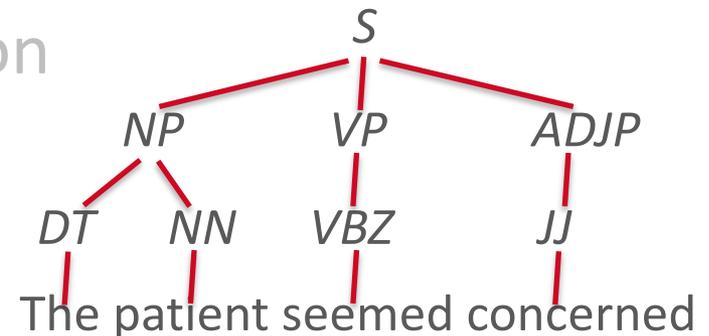
The patient seemed concerned.

# Natural Language Processing

- **Natural Language Processing** is a set of computational methods to analyse, understand, and derive meaning from human language.
  - Typical tasks:
    - sentence boundary detection
    - tokenisation
    - part-of-speech tagging
    - named entity recognition
- The patient seemed concerned.
- Tokens: ['The', 'patient', 'seemed', 'concerned', ':']*

# Natural Language Processing

- **Natural Language Processing** is a set of computational methods to analyse, understand, and derive meaning from human language.
- Typical tasks:
  - sentence boundary detection
  - tokenisation
  - part-of-speech tagging
  - named entity recognition



# Natural Language Processing

- **Natural Language Processing** is a set of computational methods to analyse, understand, and derive meaning from human language.
- Typical tasks:
  - sentence boundary detection
  - tokenisation
  - part-of-speech tagging
  - named entity recognition

The **patient** seemed concerned.

|  
**PERSON**

# Natural Language Processing

- **Natural Language Processing** is a set of computational methods to analyse, understand, and derive meaning from human language.
- Other high-level tasks:
  - word sense disambiguation
  - negation and uncertainty identification
  - relationship extraction
  - temporal inferences

# Computational methods in NLP

Two main approaches:

- rule-based: using linguistic rules and patterns
- machine learning:
  - neural networks and deep learning
  - support vector machines

# Ontologies in NLP

Suppose that we want to analyse free text in electronic health records to answer two questions:

- Does the patient have disease X?
  - Use ontologies to make use of definitions and synonyms of diseases
- Which patients are on mood stabilisers?
  - Use ontologies to make use of terminological information, e.g. subclass

*Recommended reading at <https://towardsdatascience.com/derive-insights-from-health-data-using-knowledge-graph-technologies-b6cf2b742cd6>*

# Conclusions

- Imaging and natural language data are unstructured data. Their analysis is challenging.
- Some success stories in medical image analysis using artificial intelligence.
- Growing interest in the automated analysis of free text in electronic health records.
- Watch this space!