Big Data & Big Science: from the LHC to personalized cancer treatment

Mireia Crispin-Ortuzar
Research Fellow, Trinity College, University of Cambridge Cancer Research UK, Cambridge Institute

Geneva, 25 October 2017
8.2 million people die from cancer worldwide every year

New cancer cases are expected to rise 70% within the next 2 decades

Over 100 cancer types exist and each require unique diagnosis and treatment

Probability of developing invasive cancer

41% (Men) 38% (Women)
Figure 1. Trends in Age-adjusted Cancer Death Rates* by Site, Males, US, 1930-2014

Per 100,000, age adjusted to the 2000 US standard population. *Mortality rates for pancreatic and liver cancers are increasing.

Note: Due to changes in ICD coding, numerator information has changed over time. Rates for cancers of the liver, lung and bronchus, uterus, and colon and rectum are affected by these coding changes.


©2017, American Cancer Society, Inc., Surveillance Research
What is cancer?
Unfixable DNA damage → Apoptosis

Cell continues dividing
HYPOTHETICAL SERIES OF MUTATIONS LEADING TO CANCER:

1. Initial mutation inactivates a negative cell cycle regulator.

2. Next mutation overactivates a positive cell cycle regulator.

3. Third mutation inactivates a genome stability factor.

Additional mutations accumulate rapidly.

Cancer cell.
Traditional cancer treatment

RADIATION
- DNA strand break

SURGERY
- Tumour removal

CHEMOTHERAPY
- Non-specific mitosis inhibition

TARGETED DRUG THERAPIES
- Target molecules to block growth
Inter-patient population subtypes

Intra-patient spatial, temporal

Intra-tumor tissue

Intra-tumor genetic

Likely to develop cancer?
What type of cancer is it?
What is the optimal treatment?
What’s the optimal dose?
Will the cancer return?
Which secondary complications will occur?

Biomarker

Adapted from F. Markowetz
How can physicists help?
FUN FACT: Ex-particle-physicists make the worst biologists.
Computed Tomography (CT)

- Rotating x-ray generator
- Detectors positioned opposite to the generator
- Measure the attenuation of x-rays due to different tissue densities
- 2+1D reconstruction from multiple scans
Positron Emission Tomography (PET)

- Positron-emitting radiotracer binds to cells of interest (tumor cells, hypoxic cells...)
- Back-to-back photon emission
- Tomographic reconstruction
Imaging meets machine learning

1) Learn from training data

2) Map unseen (new) data
Imaging meets machine learning

Text Recognition

Self-driving cars

Biology

Recommendation systems

Photo search

and many, many more ...
Imaging meets machine learning

Classification

Regression

Clustering

Labeled data "Supervised"

Labeled data "Supervised"

Unlabeled data "Unsupervised"

§ http://ipython-books.github.io/featured-04/
§ https://www.codeproject.com/Articles/439890/Text-Documents-Clustering-using-K-Means-Algorithm
Machine learning in particle physics

ATLAS b-taggers:

- IP3D: Naïve Bayes
- RNNIP: Recurrent Neural Network
- MV2: Boosted Decision Tree
Machine learning in particle physics

Rejection x2

Relative Efficiency +10%

non b-jet rejection

IP3D
RNN-1HiddenLSTM
RNN-2HiddenLSTM

b-jet efficiency
Radiomics

Source CT Image → Segmented Image

Energy, Entropy, Homogeneity, Contrast

Haralick Texture Features

Machine Learning Classifier

Classifications/Image-Based Outcomes

Applications
Hypoxic cells have much higher survival per unit of dose than oxic cells. Well oxygenated cells show a much steeper cell survival curve.

The ratio in cell kill between oxygenated and deoxygenated cells is the Oxygen Enhancement Ratio.
Oropharyngeal cancer
Chemo-radiotherapy
Setton et al, IJROBP, 2012 Jan 1;82(1):291-8

PET Scan

70 Gy
30 Gy

O₂

O₂
(a) TBR_{max} = 1.5
(b) TBR_{max} = 1.8
(c) TBR_{max} = 1.0

**Top PET features**

<table>
<thead>
<tr>
<th>FDG SUV 90th pct.</th>
<th>FDG SUV skewness</th>
<th>CT^{high-SUV}</th>
<th>CT^{low-SUV}</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Mid</td>
<td>Low</td>
<td>High</td>
<td>Mid</td>
</tr>
<tr>
<td>Low</td>
<td>Mid</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

- Hypoxia, perfusion
- Hypoxic fraction
- Vascularity
- Hypodensity (necrosis)
2) Detecting recurrence

(1) Local recurrence

A

B

C

D

(2) Benign radiation-induced lung injury

E

F

G

H

S. Mattonen et al, IJROBP, 95, 5 (2016)
Table 3.
Leave-one-out cross-validation results of the radiomic signature compared with physician assessment at 2 to 5 months after SABR

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Radiomic signature</th>
<th>Radiation oncologists</th>
<th>Radiologists</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Observer 1</td>
<td>Observer 2</td>
</tr>
<tr>
<td>Error</td>
<td>23.7%</td>
<td>34.2%</td>
<td>34.2%</td>
</tr>
<tr>
<td>FPR</td>
<td>24.0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>FNR</td>
<td>23.1%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Percent very or completely certain</td>
<td>-</td>
<td>50%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Abbreviations: FNR = false negative rate; FPR = false positive rate; SABR = stereotactic ablative radiation therapy.
3) Outcome prediction
Consensus value heat map

☐ Never clustered together

☐ Always clustered together

Cluster 1

Cluster 2

Cluster 3
Machine Learning

Input ➔ Feature extraction ➔ Classification ➔ Output

Deep Learning

Input ➔ Feature extraction + Classification ➔ Output
Classical lymphocytic leukemia
Follicular lymphoma
Mantle cell lymphoma

Classification accuracy: 96.58% ± 0.01%

In summary

☑ Treatment personalization
☑ Patient follow-up
☑ Outcome prediction & data integration
☑ End-to-end classification
Identifying genetic subtypes using deep learning on H&E slides.

Tumour heterogeneity
Finding statistical cluster correlations between modalities.

Luminal A vs. Luminal B
Identifying genetic subtypes using deep learning on H&E slides.

Treatment response
Predicting which patients respond to neoadjuvant chemotherapy.
Where is the data?

Where is the data?

Grand Unification of Sciences
Thank you!

@mireiacrispin

mc973@cam.ac.uk
Additional sources

- http://vision.cse.psu.edu/research/tumorDetect/index.shtml
- https://pct.mdanderson.org
- http://www.nature.com/news/big-science-has-a-buzzword-problem-1.21354