BIOMEDICAL IMAGING AND GENETIC DATA ANALYTICS WITH AI
TOWARDS PRECISION MEDICINE

Wiro Niessen
Erasmus MC & TU Delft
Quantib BV

Disclosure
Wiro Niessen is founder, scientific lead (0.2 fte) and shareholder of Quantib BV

Data fuels science for society – but what about the health domain?
ImageNet 2012: Image classification breakthrough with convolutional neural nets

ImageNet statistics

- More than 14 million images have been hand-annotated by the project to indicate what objects are pictured and in at least one million of the images, bounding boxes are also provided.
- ImageNet contains more than 20,000 categories with a typical category, such as "balloon" or "strawberry", consisting of several hundred images.

Key to its success: large open data resource & challenge aspect

Can this success be translated to clinical practice?

Specific health domain challenges:
- We need to do more than image perception.
- We need to collect more than images alone (genetics, omics, clinical information, exposome).
- Human biology and pathology is highly variable.
- Data bias is a challenge.
Promises & challenges in health domain

- DeepMind: “predicting acute kidney injury up to 2 days before it happens” (Nature, July 2019).
- 703,000 patients.
- 620,000 data points / 3,600 predictive.
- Blood tests, vital signs, past procedures, prescription, intensive care unit admission)
- No actual prediction has been made (retrospective study); accuracy is 55.8% and depends on time to event; prospective validation needed.
- Dataset obtained via US Department of Veterans Affairs: 94% male, and biased population.*

Some features may be very much dependent on health care system/setting

---

Population imaging: Rotterdam Study

- Population study running over 25 years
- > 15,000 subjects included
- Extensive geno- and phenotyping (imaging) available

---

Population imaging: design

- RISK FACTORS
  - Genetic
  - Lifestyle
  - Smoking
- BLACK BOX
  - Dementia
  - Stroke
- OUTCOMES
  - Brain atrophy
  - Lesions
  - Black box
Can this approach succeed?

Source: De Groot et al, "Stroke 2013 Progress and Innovation award"

What we currently see visually (as appreciable white matter lesions) is only the tip of the iceberg of white matter pathology: searching for QIBs logical next step.
Tractography and atlas
Minutes to multiple hours
Microstructure
Brain structures

Rotterdam Scan Study (> 15,000 brain MRI)
library of quantitative imaging biomarkers

White matter tract segmentation

Clinical decision support

Quantib® ND®
Reference imaging biomarker curves from 5,000 individuals of the population-based Rotterdam Scan Study
VDA cleared and CE marked
Totally new imaging biomarkers

Convolutional Neural Network architecture for brain age prediction (trained on 5865 images, tested on 2353)

”biological” brain age

Wang et al. PNAS 2019

Kaplan-Meier curves for new biomarker (delta brain / calendar age)

Oncology: radiomics hypothesis:

There exists a correlation between medical image features and underlying biological information

Slide courtesy: Philippe Lambin
Oncology: radiomics pipelines

1p/19q Mutation in Low Grade Glioma

Radiogenomics: predicting genetic mutation status from non-invasive imaging data

Table 3. Predictive performance of four clinical experts compared with the algorithm on the TCGA validation dataset.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Neurosurgeon 1</th>
<th>Neurosurgeon 2</th>
<th>Neurologist 1</th>
<th>Neurologist 2</th>
<th>Average of algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.205</td>
<td>0.317</td>
<td>0.480</td>
<td>0.600</td>
<td>0.574</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.205</td>
<td>0.350</td>
<td>0.550</td>
<td>0.700</td>
<td>0.611</td>
</tr>
<tr>
<td>AUROC</td>
<td>0.599</td>
<td>0.888</td>
<td>0.715</td>
<td>0.944</td>
<td>0.815</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.372</td>
<td>0.499</td>
<td>0.618</td>
<td>0.468</td>
<td>0.505</td>
</tr>
</tbody>
</table>

* p values for significance of difference between clinical experts and algorithm.
Risk factors:
- Genetic
- Blood pressure
- Smoking

Brain changes:
- Hippocampal volume

Population imaging genetics
Imaging genetics: gaining insight in relation genetic liability, environmental factors and imaging phenotype

Neural Network - KEGG Pathway

Future of imaging genetics

Holy grail: find phenotype = f (genotype, environmental factors)

Current approaches: mostly massive number of linear regressions

Promises in:
- Larger datasets
- Machine and deep learning for learning more complex relations

Challenges:
- DL/ML cannot straightforwardly be applied (heterogeneous data, biological variability)
- Modular approach, integrating prior knowledge with DL
Requirements successful introduction AI

- High quality data to train algorithms using state of the art algorithm optimization methods
- Clear definition of tasks and seamless integration into the workflow
- Proper validation strategies:
  - Many promising algorithms may not function as well in clinical practice as reported in literature
  - Evaluation has been performed on retrospective data, often one or limited number of centers
  - Issues: data bias, lack of generalizability

Data driven precision health requires health data infrastructure

Taking individual variability into account to promote health, prevent & optimize diagnosis, prognosis and treatment

Utilizing our rich data resources and AI

FAIR data and distributed learning

- Findable
- Accessible
- Interoperable
- Rewritable
Next generation validation strategies

Anything You Can Do, I Can Do Better
(No You Can’t) . . . *

Keith Poetz
Pavilions MCADL, Prodigy Imaging Group, University of Southern California,
Los Angeles, California, 90009-9271
Received February 5, 1986, revised March 3, 1986

Don't assume, assess!

MICCAI – ACR / RSNA / ESR collaboration

- AI use cases
  - Clinically relevant, ethical and effective
- Challenges
  - Good quality training data & objective performance evaluation

Joint effort required!

[Diagram showing ingredients for success and their availability]

[Table with categories: AI Startups, Established Companies, Healthcare Delivery Systems, Professional Societies, Academic Health Systems]

[Legend indicating availability: available, can acquire, need to acquire]
The outcome of tomorrow’s COVID-19 patient is strongly determined by having access to the data of today’s patient.

To collect and find the right data, to make them accessible, and to reuse them is non-trivial.

Currently many data collection initiatives worldwide.

HealthRI provides overview of initiatives and provides services & tools and with clinical partners builds Dutch covid database of imaging and clinical data for development and objective validation AI algorithms.

“During this epidemic and in earlier occasions, we have seen severely suboptimal data management and data reuse.”

Ensure that the WHO-CRFs and other input forms for Corona data (and later and outbreaks in general) are properly mapped to a machine-readable (RDF) format, so that any stakeholder can create input forms that lead to the resulting data being a machine actionable (FAIR) digital object.
Towards a national COVID-19

NvvR-Health RI Covid-19 imaging database

Linking to international Covid data initiatives

Linking to European Covid-19 data portal: https://www.covid19dataportal.org
- Coordinated by ELIXIR and EMBL-EBI
- Health-RI connected to Covid-19 data portal initiative via ELIXIR-NL and EOSC-LIFE

- Federated national EGA to ensure rapid sharing of Covid-19 host omics - phenotypic data across Europe
- FAIR metadata
State of the art machine learning on rich data (AI) to support prevention, early disease detection, improved diagnostics & prognostics

Distributed analysis for deeper insights

Future of healthcare is a learning healthcare system

What is needed?

- Work on higher quality and better accessible (image) data for science and innovation
- Implement FAIR data, distributed access and Open Science
- Create ML/DL challenges for important tasks
- Prospective validation for responsible introduction AI

Thank you