BIOMEDICAL IMAGING AND GENETIC DATA ANALYTICS WITH AI

TOWARDS PRECISION MEDICINE

Wiro Niessen
Erasmus MC & TU Delft
Quantib BV

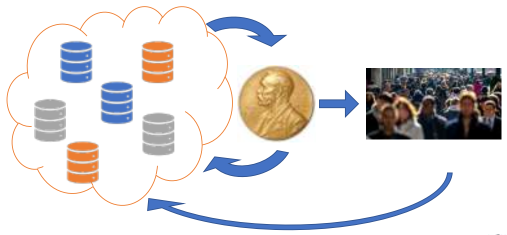
1

Disclosure

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

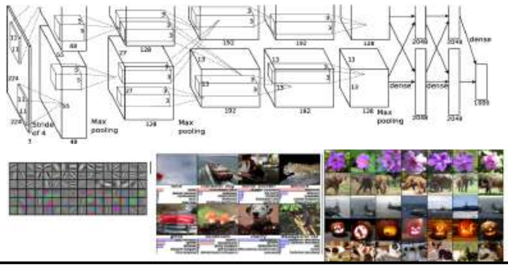
2

Data fuels science for society – but what about the health domain?



3

ImageNet 2012: Image classification breakthrough with convolutional neural nets



4

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

**Source: Wikipedia 2019*

5

Can this success be translated to clinical practice?

Specific health domain challenges:

- We need to do to 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

6

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

* See e.g. blog Julia Powles, OneZero

Erasmus MC

7

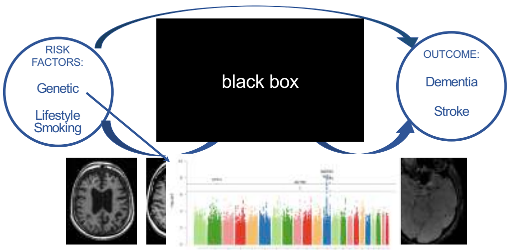
Population imaging: Rotterdam Study

- Population study running over 25 years
- > 15.000 subjects included
- Extensive geno- and phenotyping (imaging) available

Erasmus MC

8

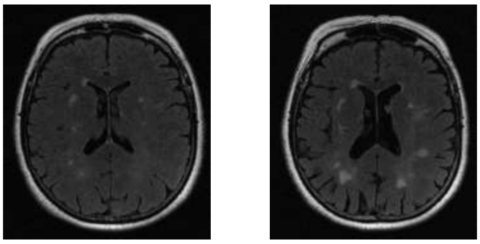
Population imaging: design



Erasmus MC

9

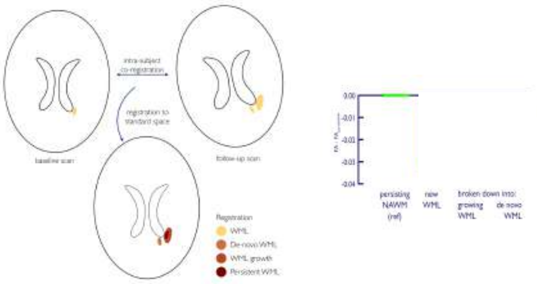
Can this approach succeed?



Erasmus MC

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

10



Erasmus MC

11

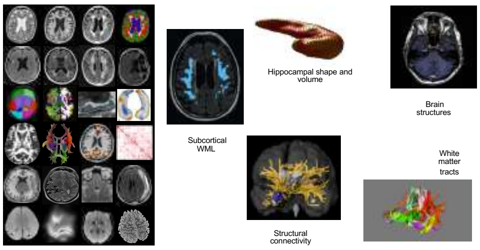
What we currently see visually (as appreciable white matter lesions) is only the tip of the iceberg of white matter pathology: searching for QIBs logical next step

Erasmus MC

12

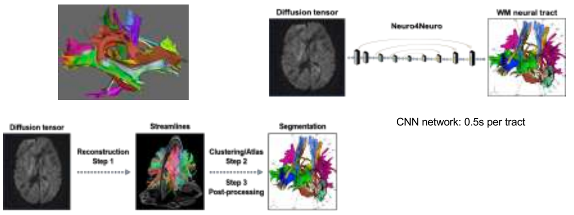
Rotterdam Scan Study (> 15.000 brain MRI) library of quantitative imaging biomarkers

- Brain tissue
- White matter lesions
- Brain structures
- Microstructure
- Incidental findings
- Micro bleedings



13

White matter tract segmentation



CNN network: 0.5s per tract

Tractography and atlas-based segmentation
Minutes to multiple hours

14

Clinical decision support

Quantib® ND*

Reference imaging biomarker curves from 5.000 individuals of the population-based Rotterdam Scan Study

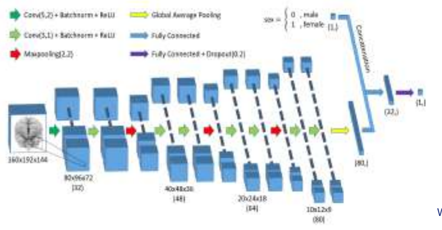
*FDA cleared and CE marked

Courtesy: Quantib

15

Totally new imaging biomarkers

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



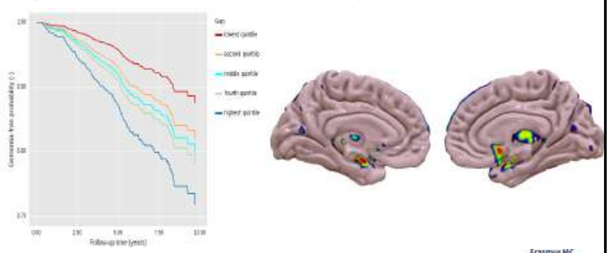
"biological" brain age

Wang et al. PNAS 2019

Erasmus MC

16

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

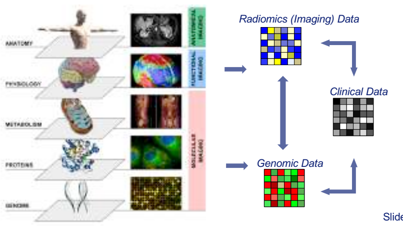


Erasmus MC

17

Oncology: radiomics hypothesis:

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

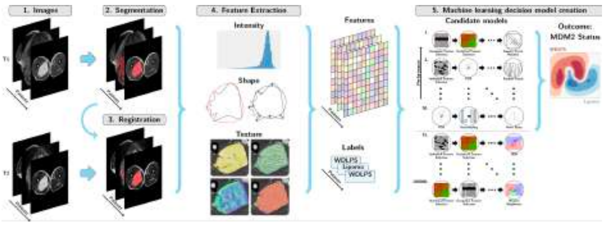


Slide courtesy: Philippe Lambin

Erasmus MC

18

Oncology: radiomics pipelines

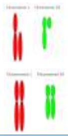


Starmans et al. British Journal of Surgery 2019

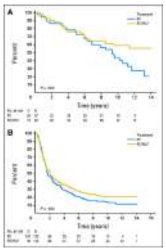
Erasmus MC

19

1p/19q Mutation in Low Grade Glioma



Goal	Predict 1p/19q genetic mutations in 81 patients
Ground Truth	PA
Modality	MR (T2, DWI)
Discovery	Two-class: co-deleted vs non co-deleted



Van der Voort et al. Clinical Cancer Research 2019

Erasmus MC

20

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

Table 3. Predictive performance of four clinical experts compared with the algorithm on the TCIA Validation Dataset.

	Neuro-surgeon 1	Neuro-surgeon 2	Average of surgeons	Neuro-radiologist 1	Neuro-radiologist 2	Average of radiologists	Algorithm
Accuracy, with p value*	0.520, 0.073	0.457, 0.002	0.489	0.690, 0.720	0.574, 0.266	0.632	0.693
AUC	0.580	0.449	0.515	0.830	0.792	0.811	0.723
Sensitivity	0.370	0.459	0.415	0.610	0.459	0.535	0.732
Specificity	0.820	0.455	0.638	0.840	0.795	0.818	0.617

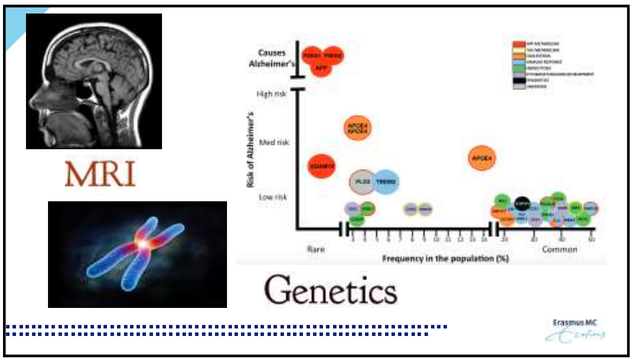
* statistical comparison (McNemar) of accuracy between clinical experts and algorithm

AUC = area under the curve

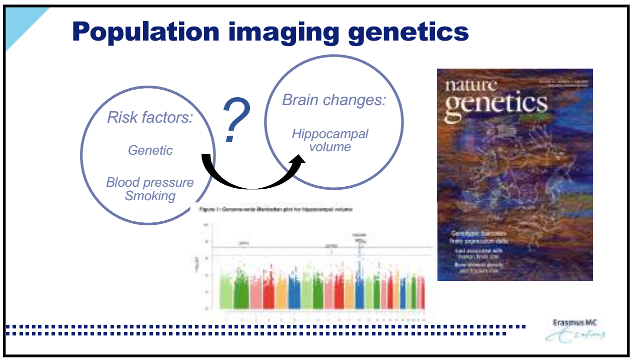
21


BEYOND IMAGING
Erasmus MC

22

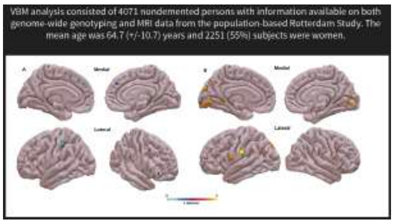

MRI
Causes Alzheimer's
High risk
Med risk
Low risk
Risk of Alzheimer's
APOE4
Rare
Frequency in the population (%)
Common
Genetics
Erasmus MC

23


Population imaging genetics
Risk factors:
Genetic
Blood pressure
Smoking
?
Brain changes:
Hippocampal volume
nature genetics
Erasmus MC

24

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

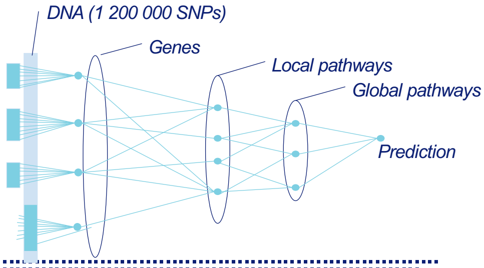


Roshchupkin et al. Neurobiology of ageing

Erasmus MC

25

Neural Network - KEGG Pathway



1,200,000 SNPs
to 21,390 genes
to 390 local pathways
to 43 pathways
to 6 global pathways
to 1 outcome

Erasmus MC

26

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

Erasmus MC

27

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

Erasmus MC

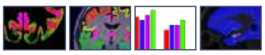
28

Data driven precision health requires health data infrastructure

health RI research infrastructure

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

Utilizing our rich data resources and AI



Erasmus MC

30

FAIR data and distributed learning

health RI research infrastructure

Erasmus MC

31

Next generation validation strategies

COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 36, 387–391 (1986)

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

KEITH PRICE

Powell Hall MC-0273, Intelligent Systems Group, University of Southern California, Los Angeles, California 90089-0273

Received February 3, 1986; revised March 5, 1986

Don't assume, assess!

Erasmus MC

32

MICCAI – ACR / RSNA / ESR collaboration

- AI use cases

Clinically relevant, ethical and effective

- Challenges

Good quality training data & objective performance evaluation

33

Joint effort required!

Ingredient for Success	AI Startups	Established Companies	Healthcare Delivery Systems	Professional Societies	Academic Health Systems
Deep technical knowledge	available	available	difficult to acquire	available	available
High performance computing	available	available	can acquire	can acquire	available
Interdisciplinary teams	can acquire	can acquire	can acquire	available	available
Ongoing source of labeled images	difficult to acquire	difficult to acquire	can acquire	can acquire	available
Infrastructure for prospective evaluation	can acquire	can acquire	can acquire	can acquire	available
Market dissemination channel	available	available	difficult to acquire	difficult to acquire	difficult to acquire

=available =can acquire =difficult to acquire

Recht et al. European Radiology 2019

34

SPEED-UP BY COVID-19

35

Dutch COVID-19 Data Support Programme

Health-RI launched the Dutch COVID-19 Data Support Programme. To support investigators and health care professionals with tools and services in their search for ways to overcome the pandemic and its health consequences.

- 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
- Health-RI 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

36

GO FAIR

health RI research infrastructure

Virus Outbreak Data Network (VODAN)

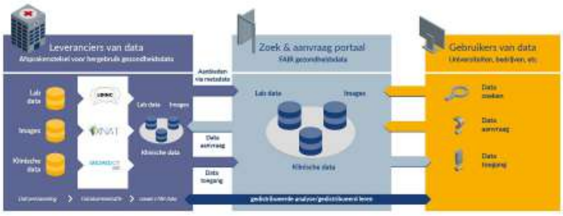
"During this epidemic and in earlier occasions, we have seen severely suboptimal data management and data reuse."

Ensure that the WHO-CRF(s) and other input forms for Corona data (and later viral 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 objects.

Erasmus MC

37

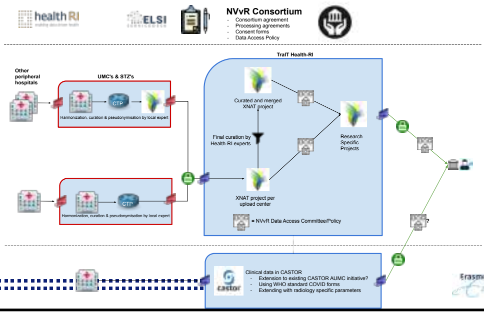
Towards a national COVID-19



Erasmus MC

38

NvvR-Health RI Covid-19 imaging database



Erasmus MC

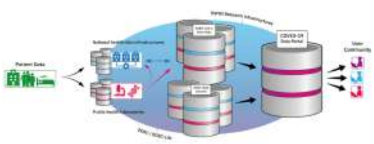
39

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*

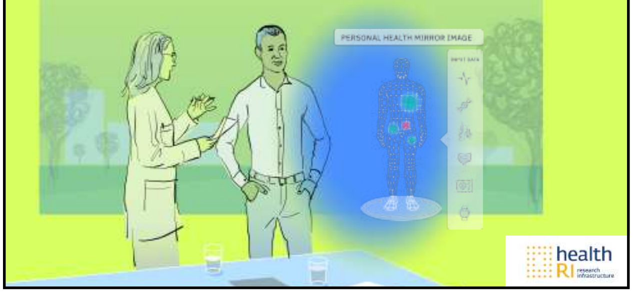


- *FAIR metadata*

Erasmus MC

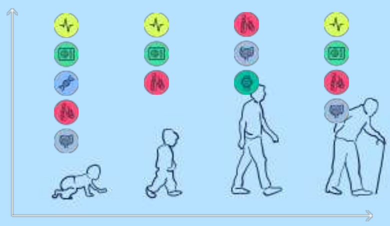
40

Future of data-driven health



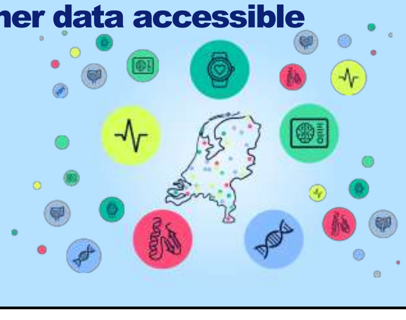
41

Data along full life cycle



42

Infrastructure to make more and richer data accessible



43

Distributed analysis for deeper insights

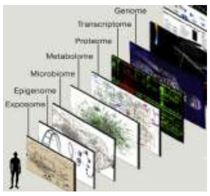
State of the art machine learning on rich data (AI) to support prevention, early disease detection, improved diagnostics & prognostics

44

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



45

Thank you

46