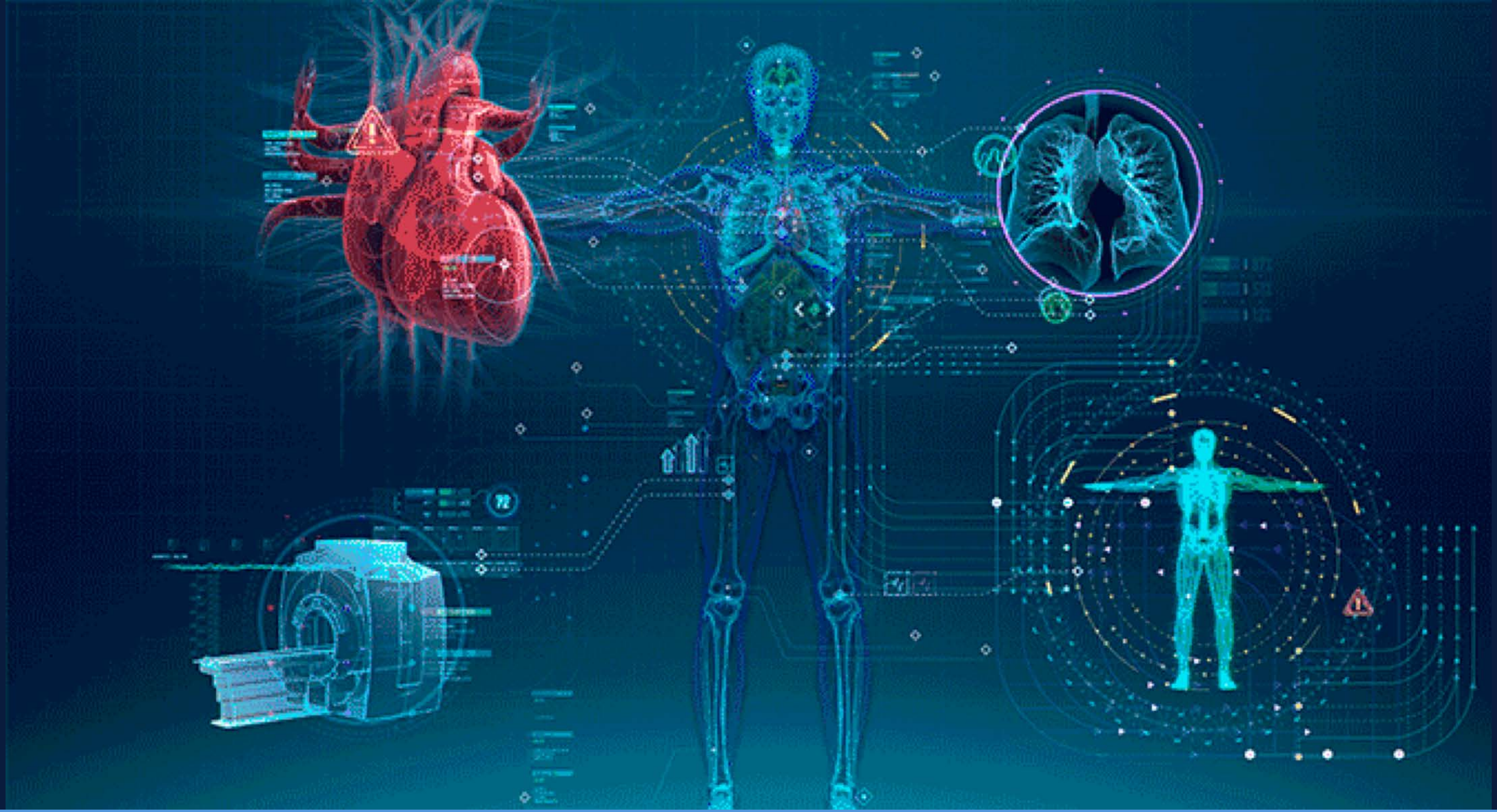


Demystification of AI-driven medical image interpretation

Benoit GALLIX

JAMA December 13, 2016 Volume 316, Number 22

JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD; Lily Peng, MD, PhD; Marc Coram, PhD; Martin C. Stumpe, PhD; Derek Wu, BS; Arunachalam Narayanaswamy, PhD; Subhashini Venugopalan, MS; Kasumi Widner, MS; Tom Madams, MEng; Jorge Cuadros, OD, PhD; Ramasamy Kim, OD, DNB; Rajiv Raman, MS, DNB; Philip C. Nelson, BS; Jessica L. Mega, MD, MPH; Dale R. Webster, PhD

Editorial by Wong and Bressler

- limits of the study
- need for further validation of the algorithm in different populations
- unresolved challenges

FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems

FDA News Release

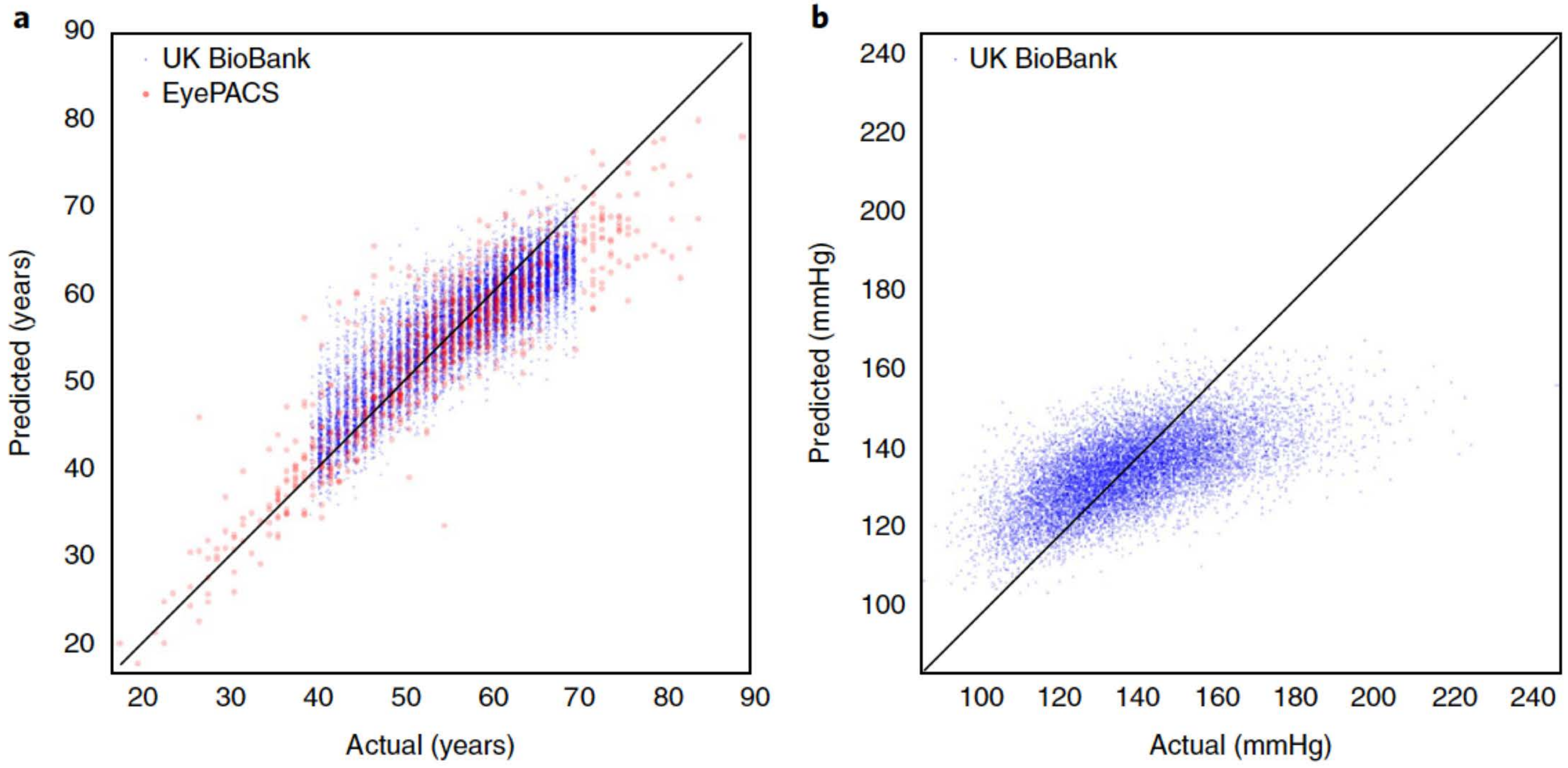
For Immediate Release

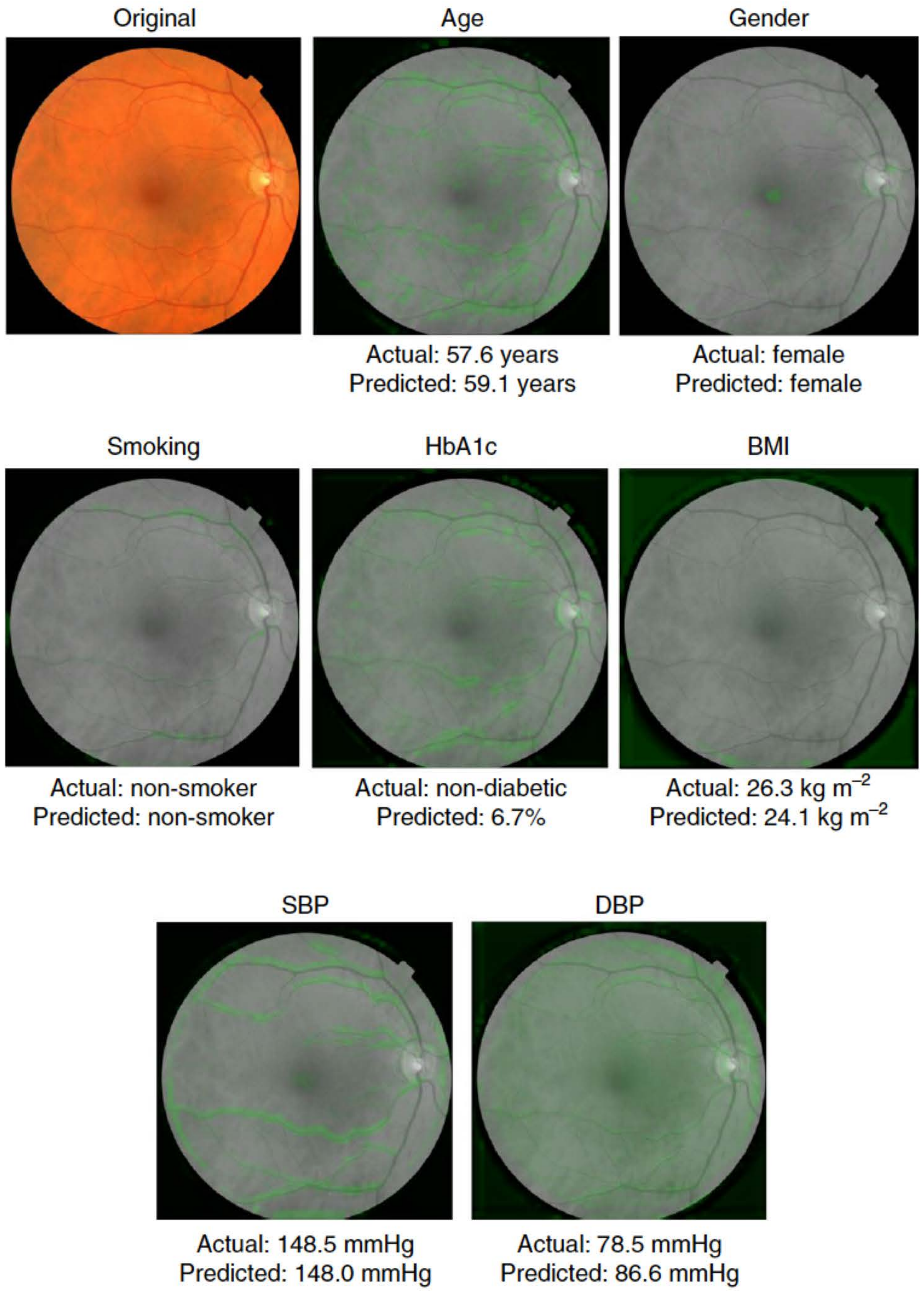
May, 2018

ARTICLES
https://doi.org/10.1038/s41551-018-0195-0
nature biomedical engineering

Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning

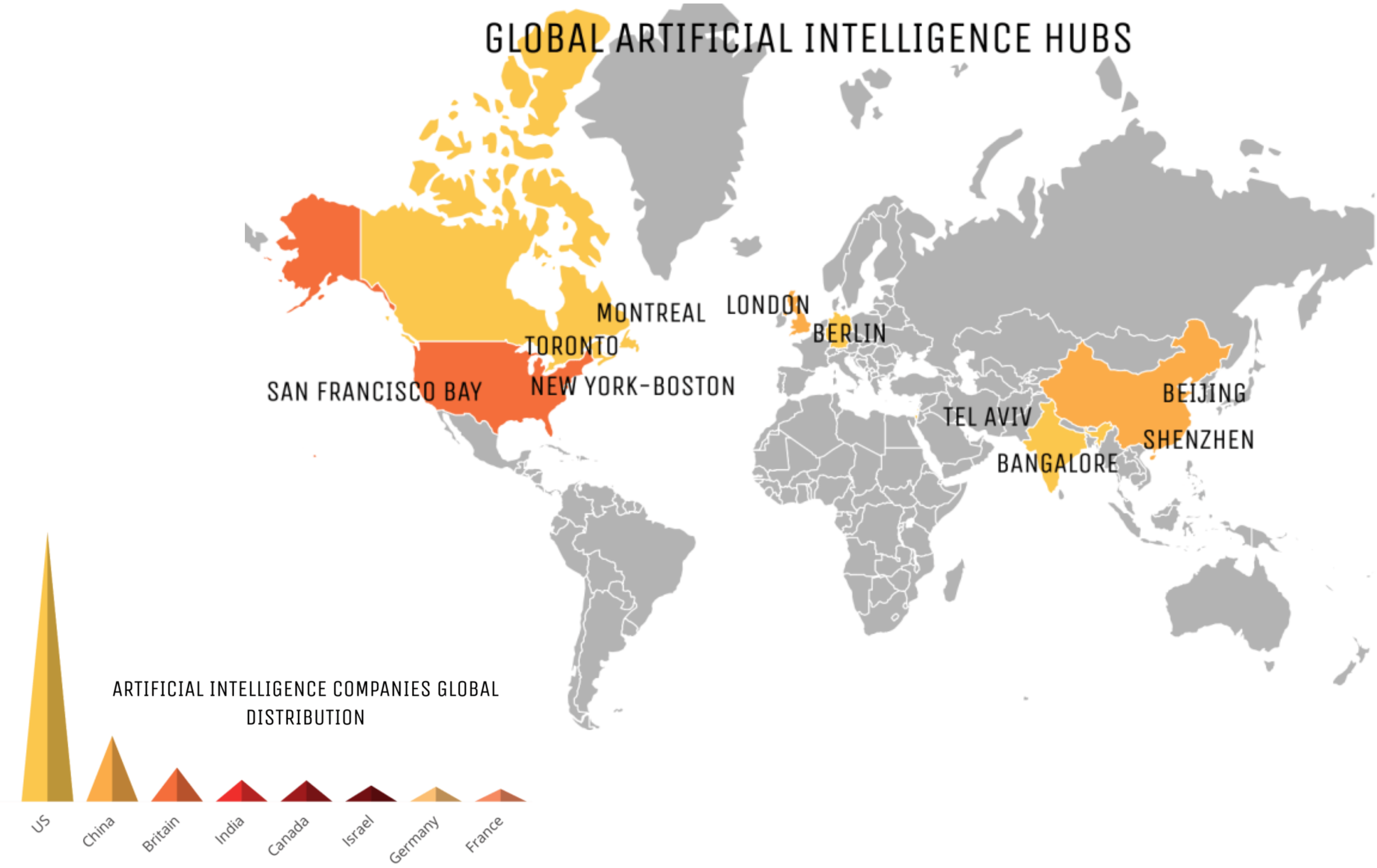
Ryan Poplin[1,4], Avinash V. Varadarajan[1,4], Katy Blumer[1], Yun Liu[1], Michael V. McConnell[2,3], Greg S. Corrado[1], Lily Peng[1,4]* and Dale R. Webster[1,4]



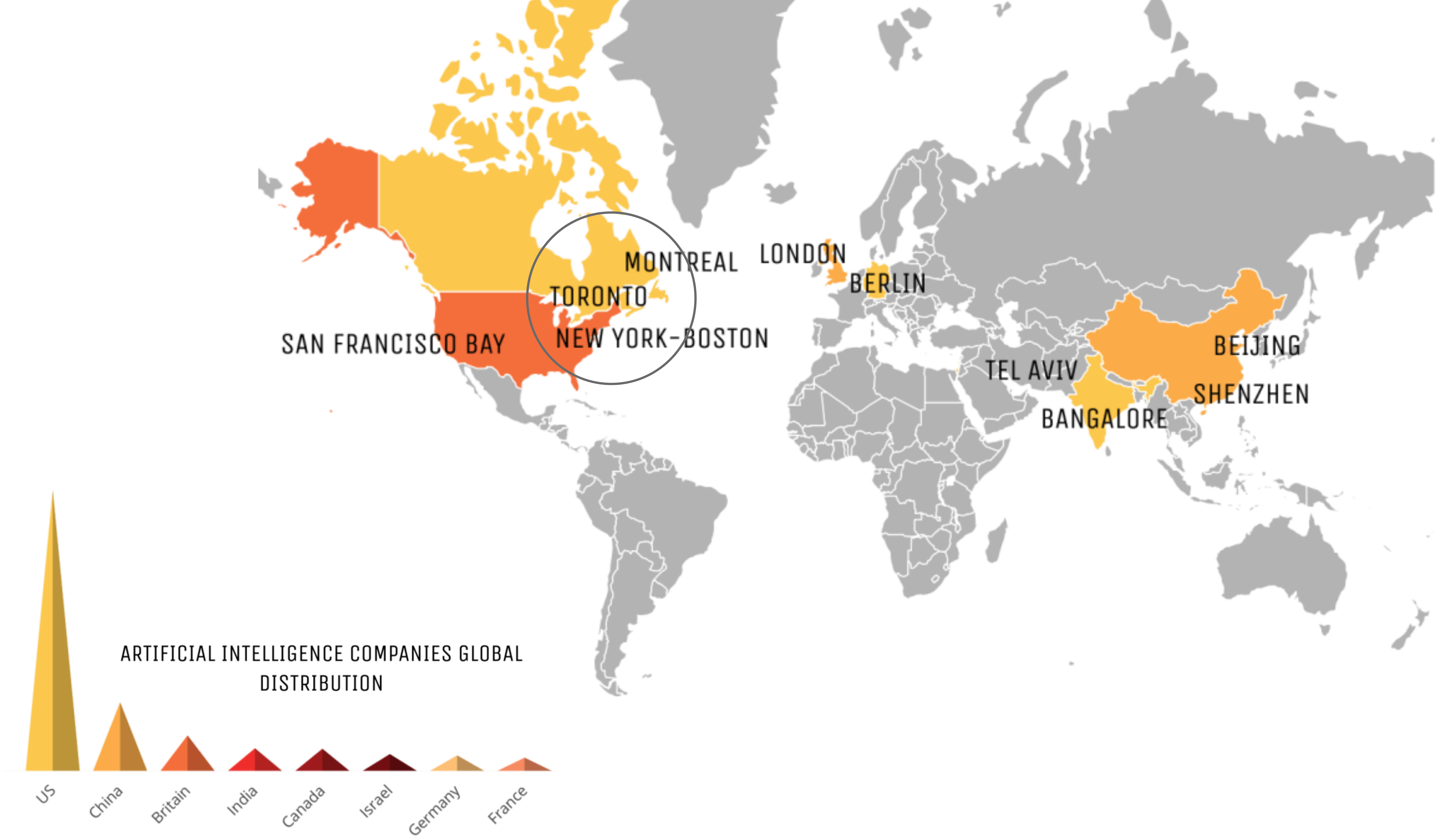


AlphaGo

2017, May 27th


GLOBAL ARTIFICIAL INTELLIGENCE HUBS
MONTREAL
LONDON
BERLIN
TORONTO
SAN FRANCISCO BAY
NEW YORK-BOSTON
BEIJING
TEL AVIV
SHENZHEN
BANGALORE
ARTIFICIAL INTELLIGENCE COMPANIES GLOBAL DISTRIBUTION
US
China
Britain
India
Canada
Israel
Germany
France

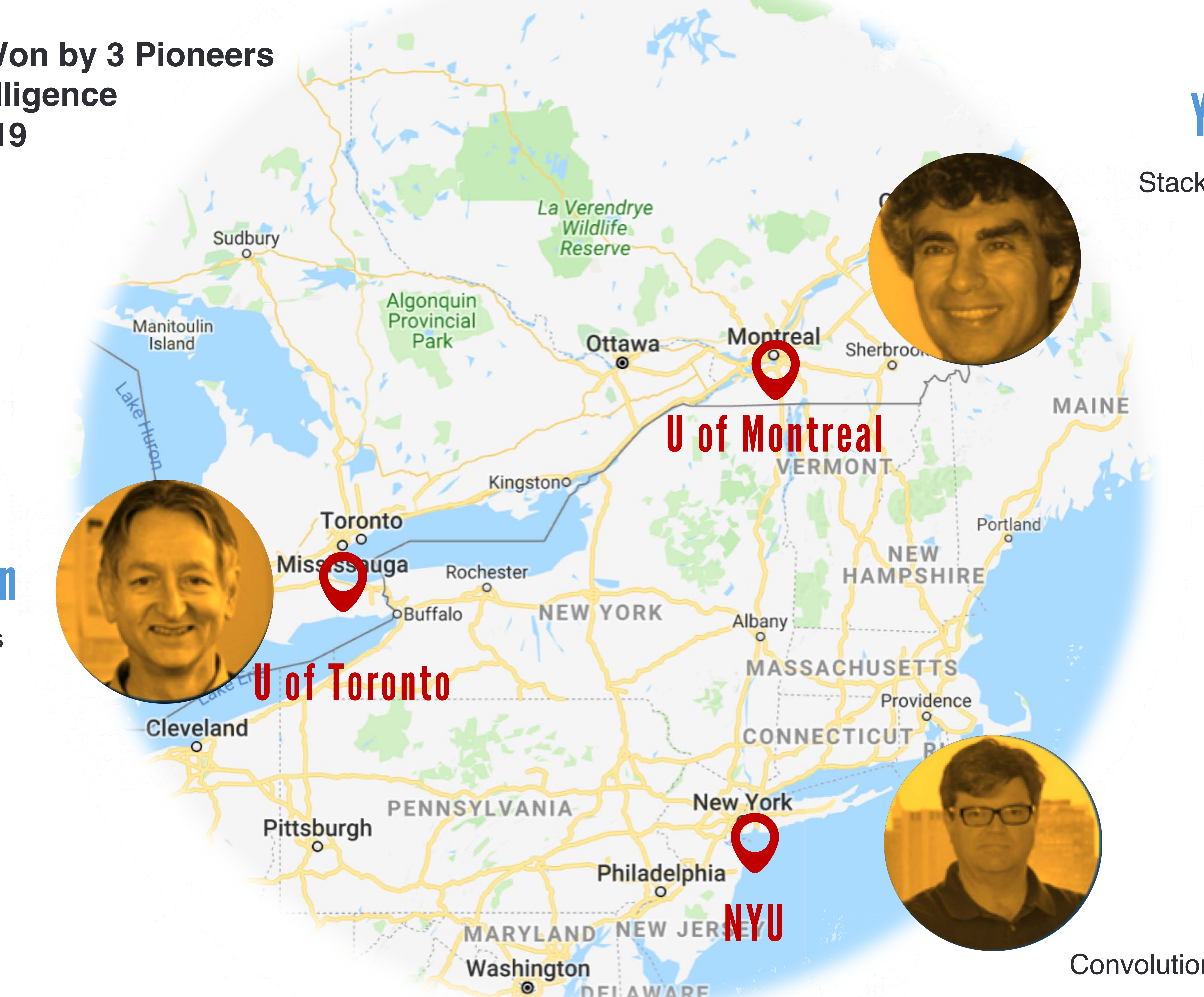
GLOBAL ARTIFICIAL INTELLIGENCE HUBS



Turing Award Won by 3 Pioneers in Artificial Intelligence
NYT Mar 27, 2019

2007
Yoshua Bengio
Stacked Auto-Encoders

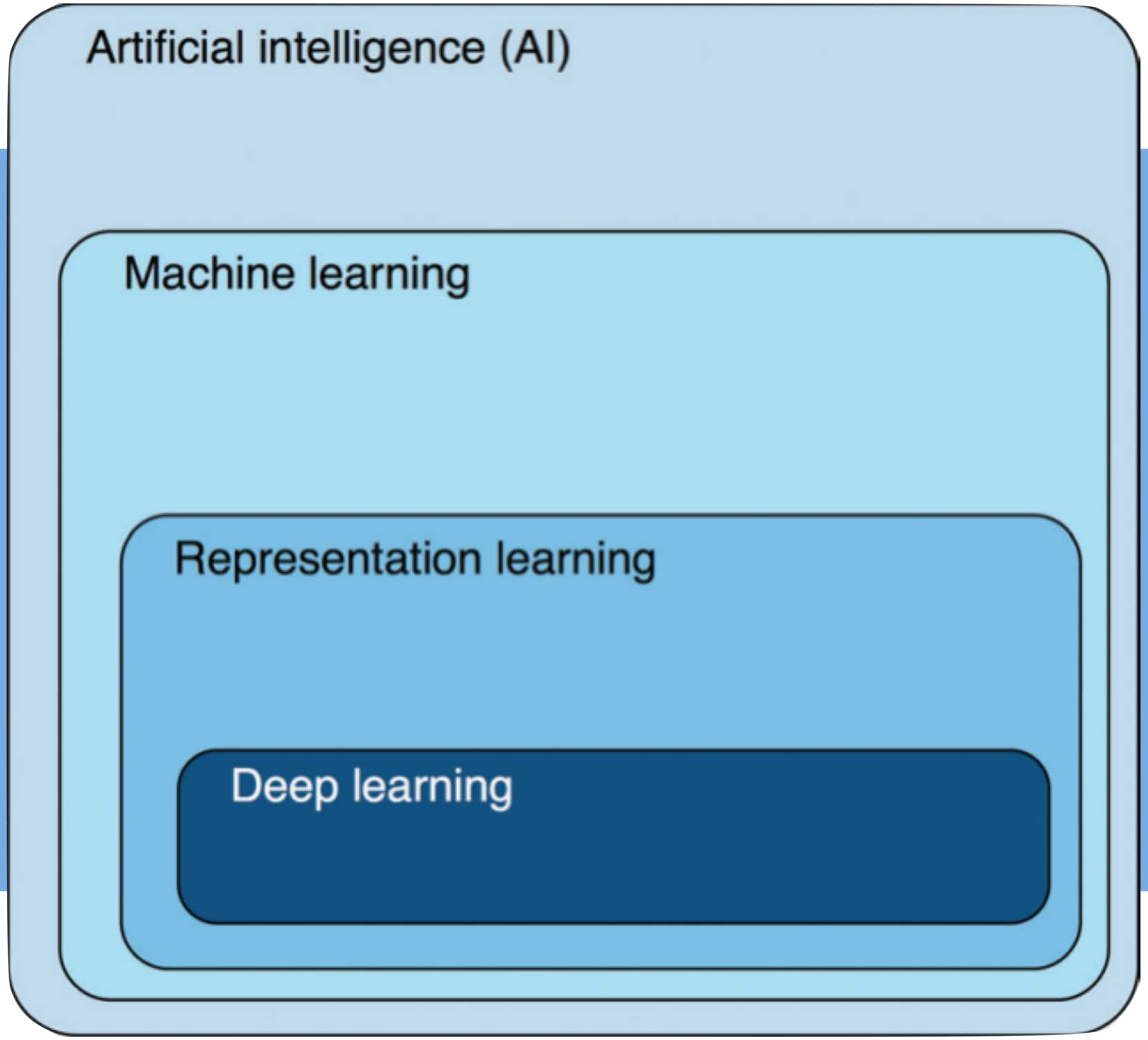
2006
Geoffrey Hinton
Deep Belief Networks

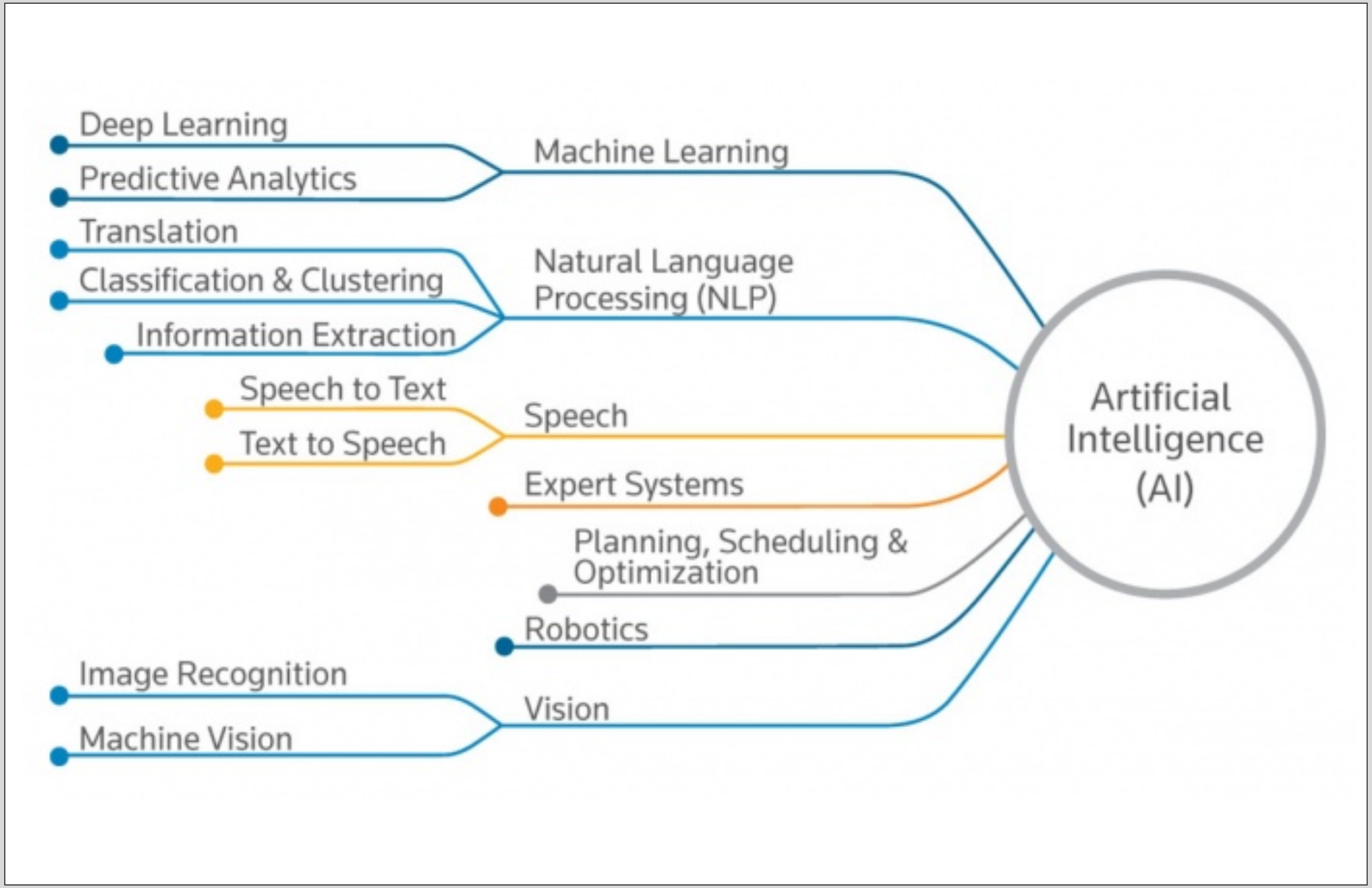


1998
Yann LeCun
Second generation Convolutional Neural Networks

Terminology Misuse

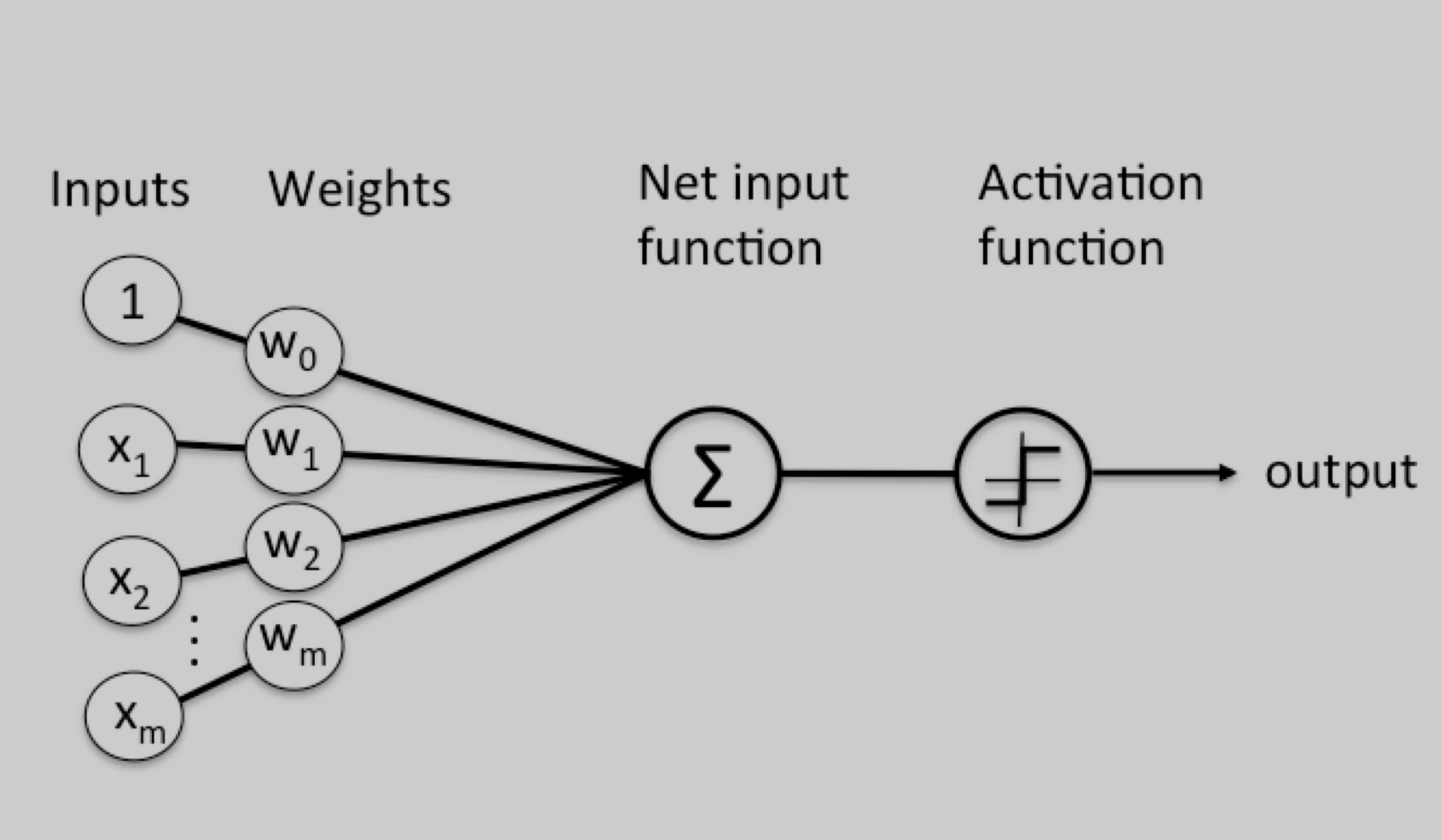
- Deep Learning ≠ Artificial Intelligence
- Deep Learning is the latest, most hyped set of machine learning techniques.




Deep Learning
Predictive Analytics
Machine Learning
Translation
Classification & Clustering
Information Extraction
Natural Language
Processing (NLP)
Speech to Text
Text to Speech
Speech
Expert Systems
Planning, Scheduling &
Optimization
Robotics
Image Recognition
Machine Vision
Vision
Artificial
Intelligence
(AI)

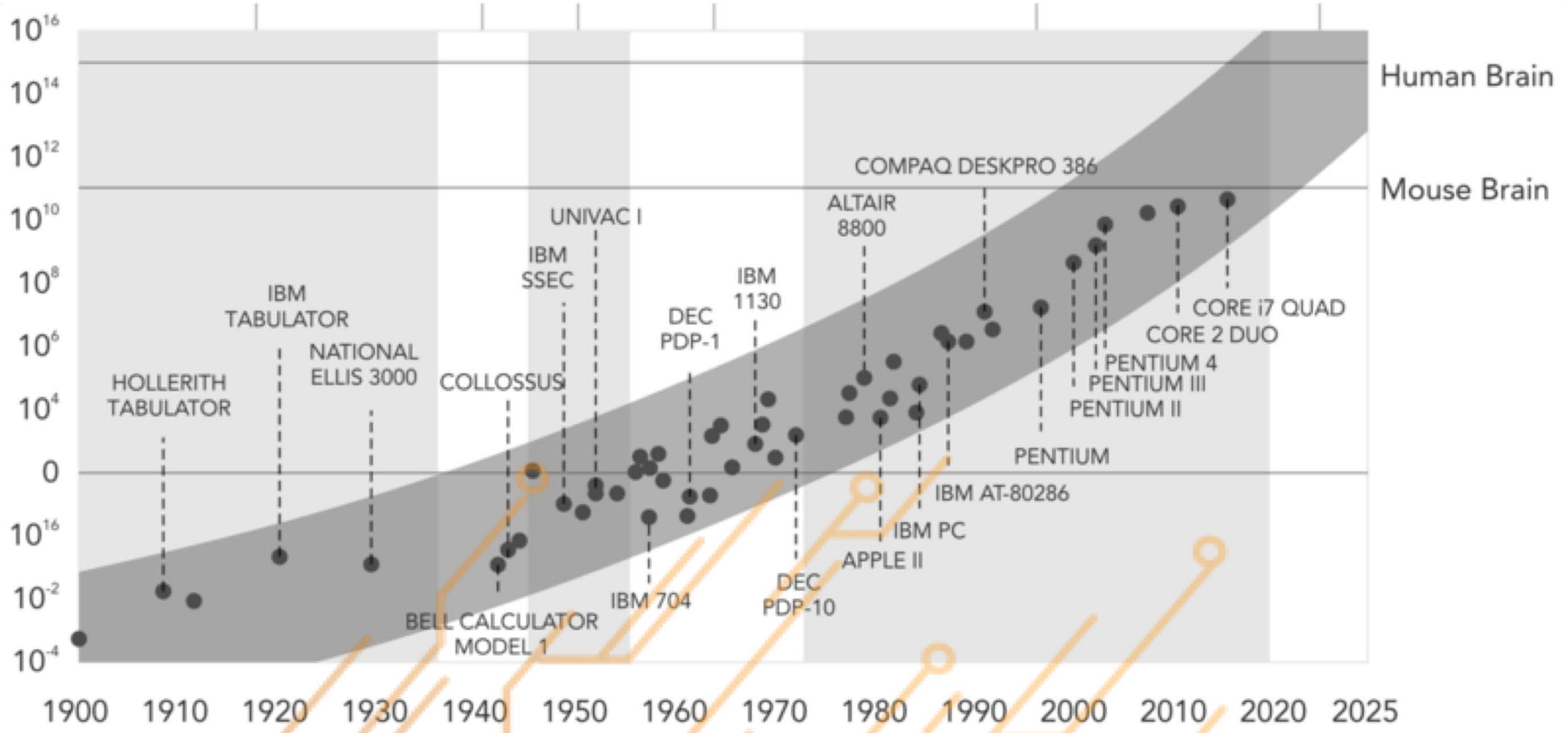
First Artificial Neural Network

Franck Rosenblatt's Perceptron, 1957



A simple simulated neuron with adaptive "synaptic weights"


TITAN Z


Human Brain
Mouse Brain
HOLLERITH TABULATOR
IBM TABULATOR
NATIONAL ELLIS 3000
COLLOSSUS
BELL CALCULATOR MODEL 1
IBM SSEC
UNIVAC I
IBM 704
DEC PDP-1
IBM 1130
DEC PDP-10
ALTAIR 8800
APPLE II
IBM PC
IBM AT-80286
COMPAQ DESKPRO 386
PENTIUM
PENTIUM II
PENTIUM III
PENTIUM 4
CORE 2 DUO
CORE i7 QUAD
10^{16} 10^{14} 10^{12} 10^{10} 10^{8} 10^{6} 10^{4} 0 10^{16} 10^{-2} 10^{-4}
1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2025

MOORE’S LAW

Data Explosion
Exabytes (billions of GB)
50000
40000
30000
20000
10000
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Structured Data
Unstructured Data


amazon
Facebook
Amazon
Google's image search
Siri
Facebooks' DeepFace
Chatbots

Tech investors think some actors will be ‘obsolete’ in five years

“There's no reason a human should be anymore reading medical images"

Vinod Khosla

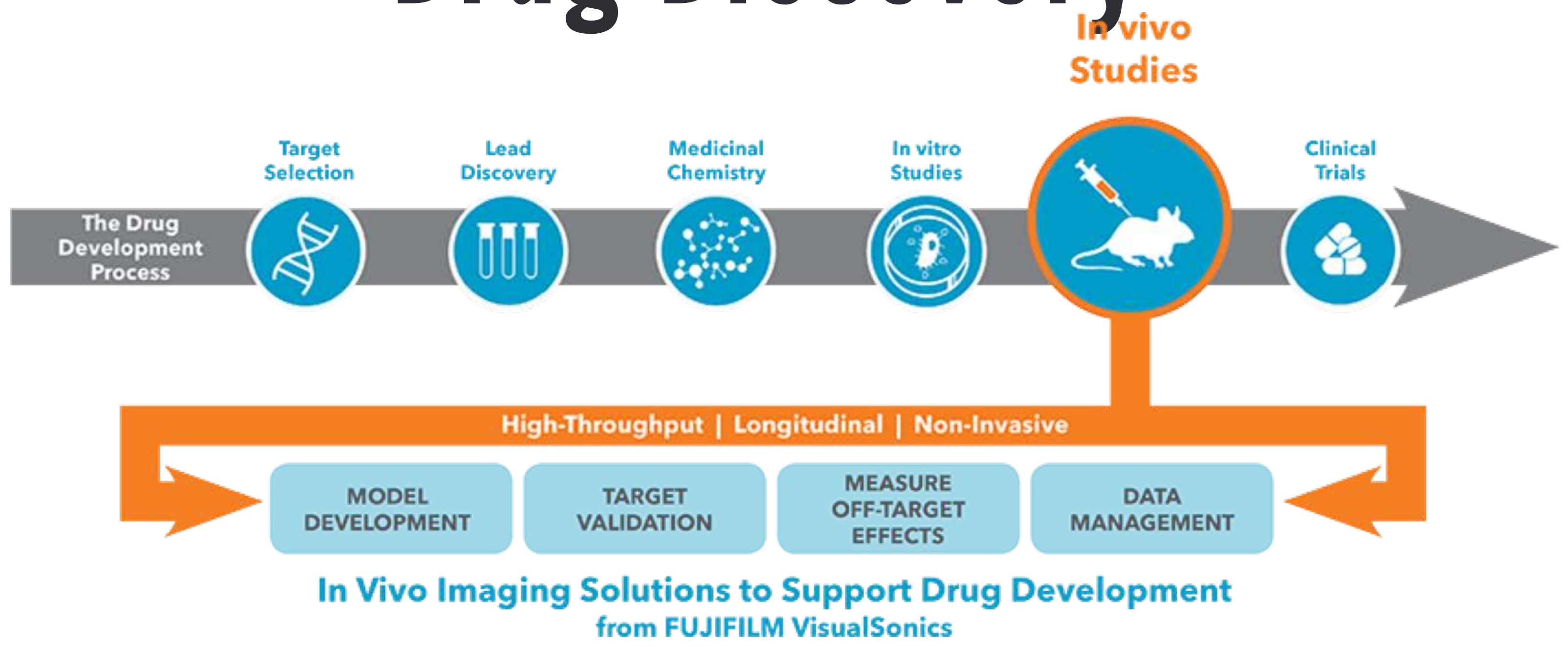
“AI won’t make jobs disappear, but will certainly make many tasks disappear”

Healthcare Sector

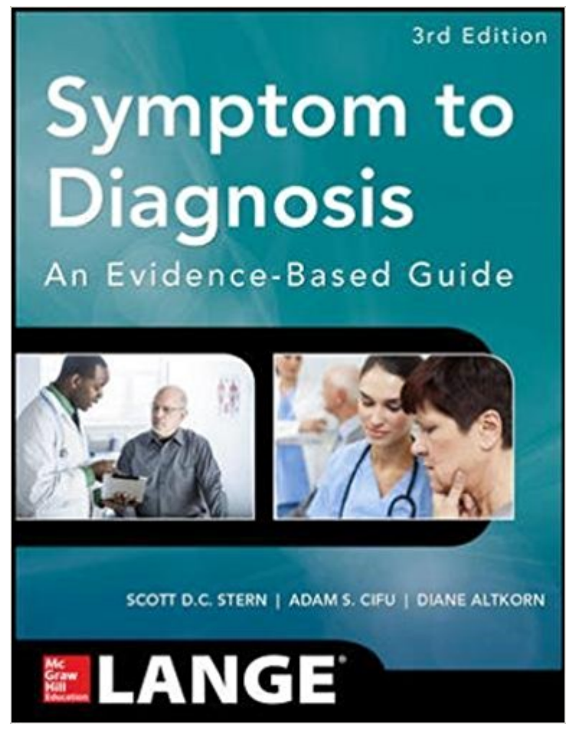
Rich in complex data and process

DNA and RNA sequencing

Drug Discovery

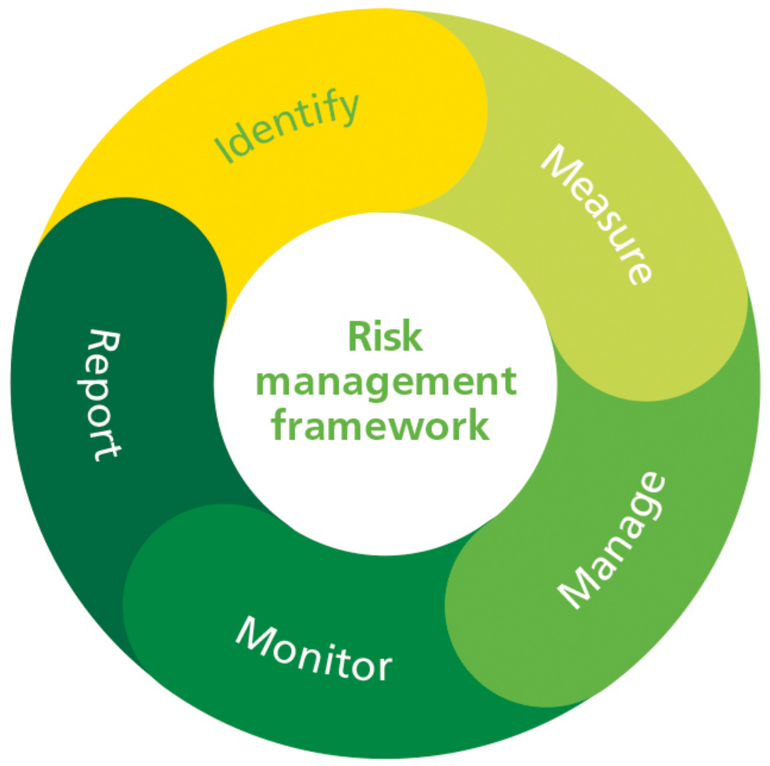


Monitoring



Medical Diagnosis

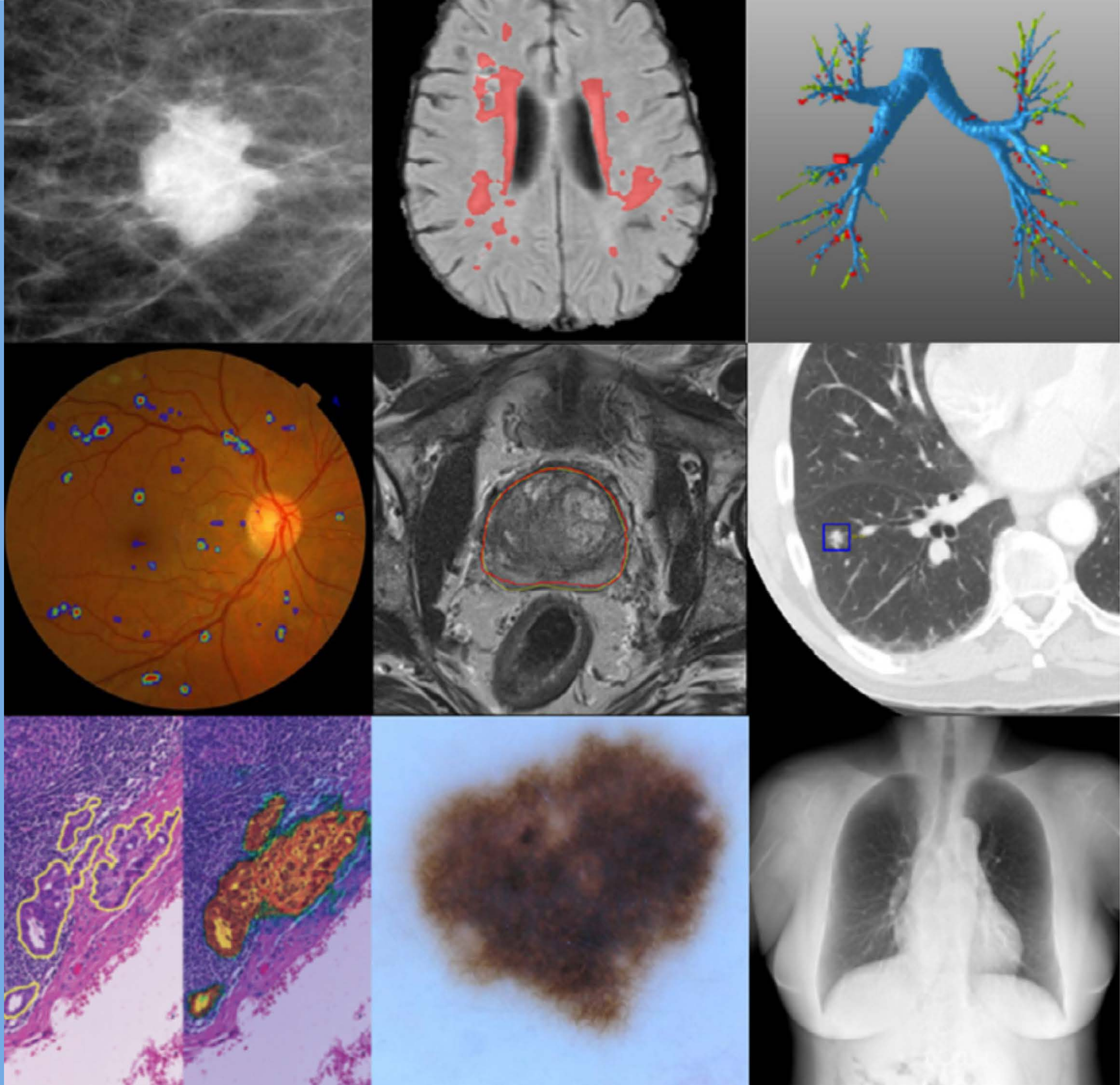
Hospital Management



Risk Management

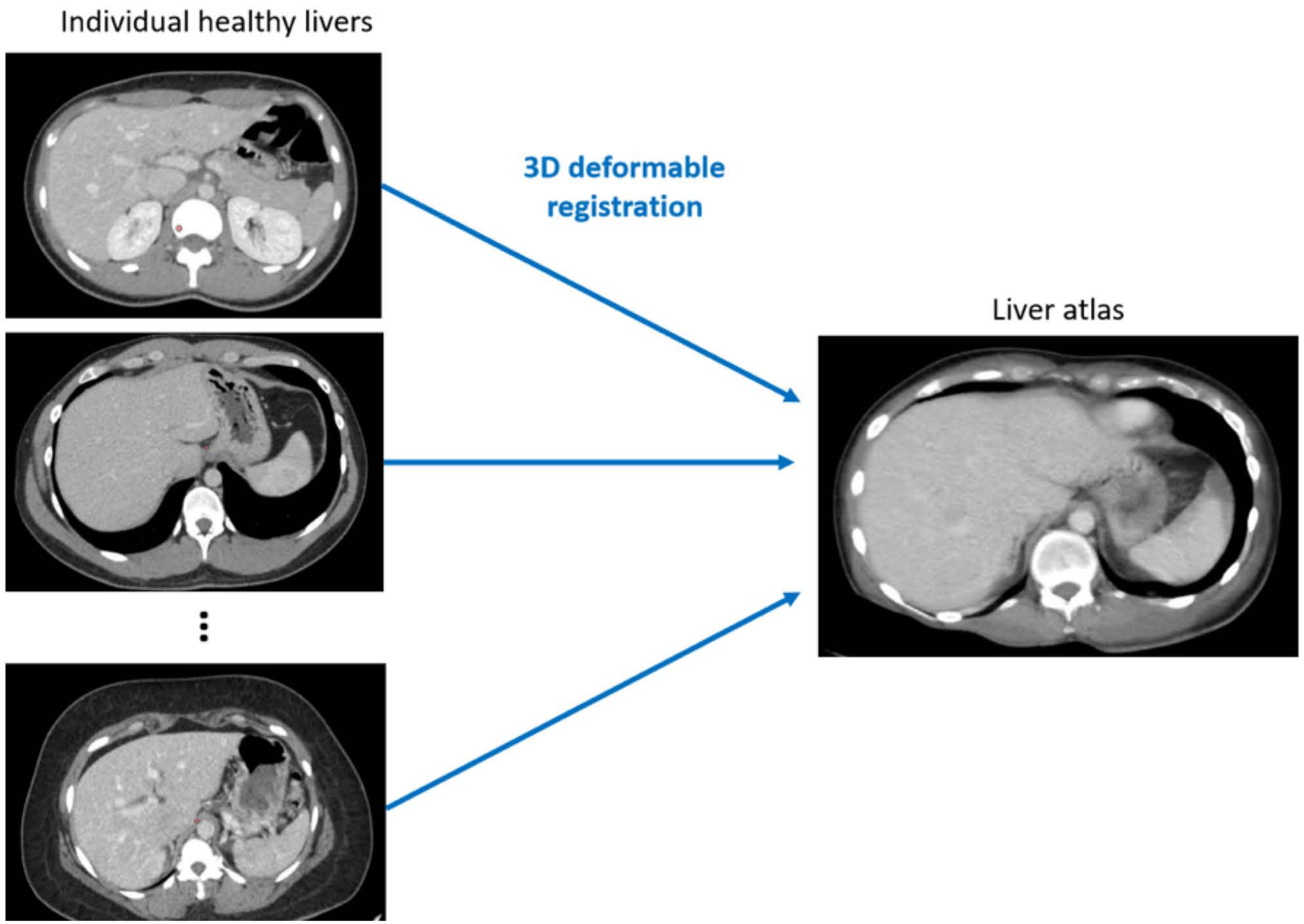
Imaging Data

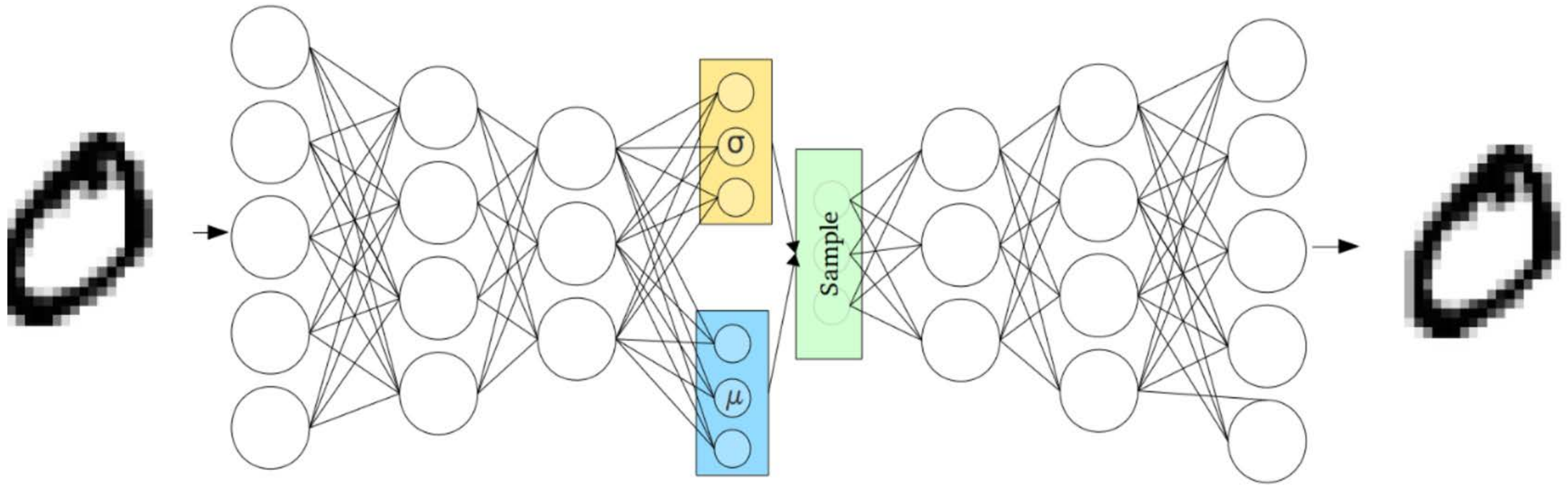
- Radiology
  - Radiotherapy
  - Imaged Guided Therapy
- Pathology
- Dermatology
- Ophthalmology
- Endoscopy



1990s-Present Atlas-based analysis

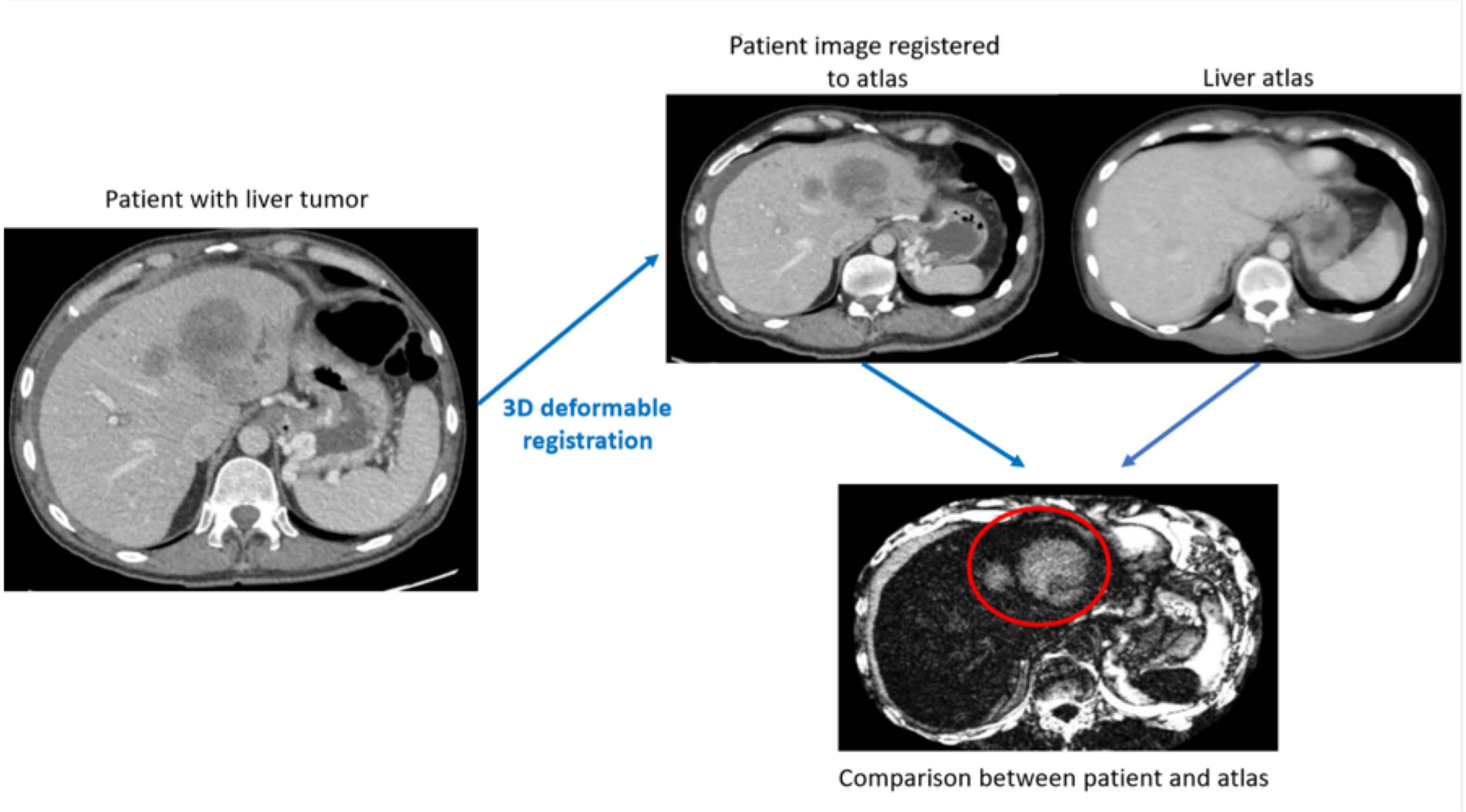
A statistical/probabilistic approach





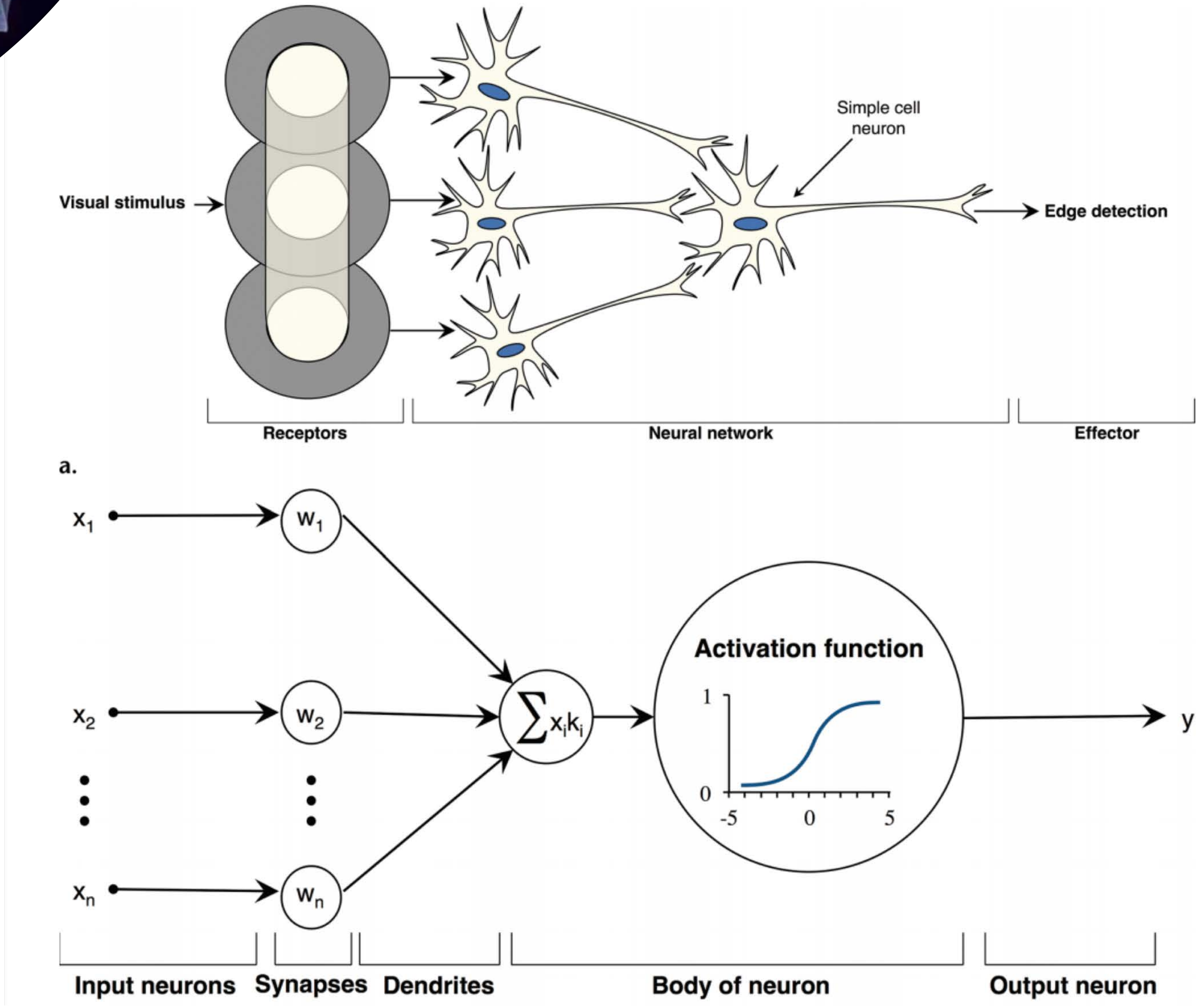
Gallix B et al. European Radiology 2018

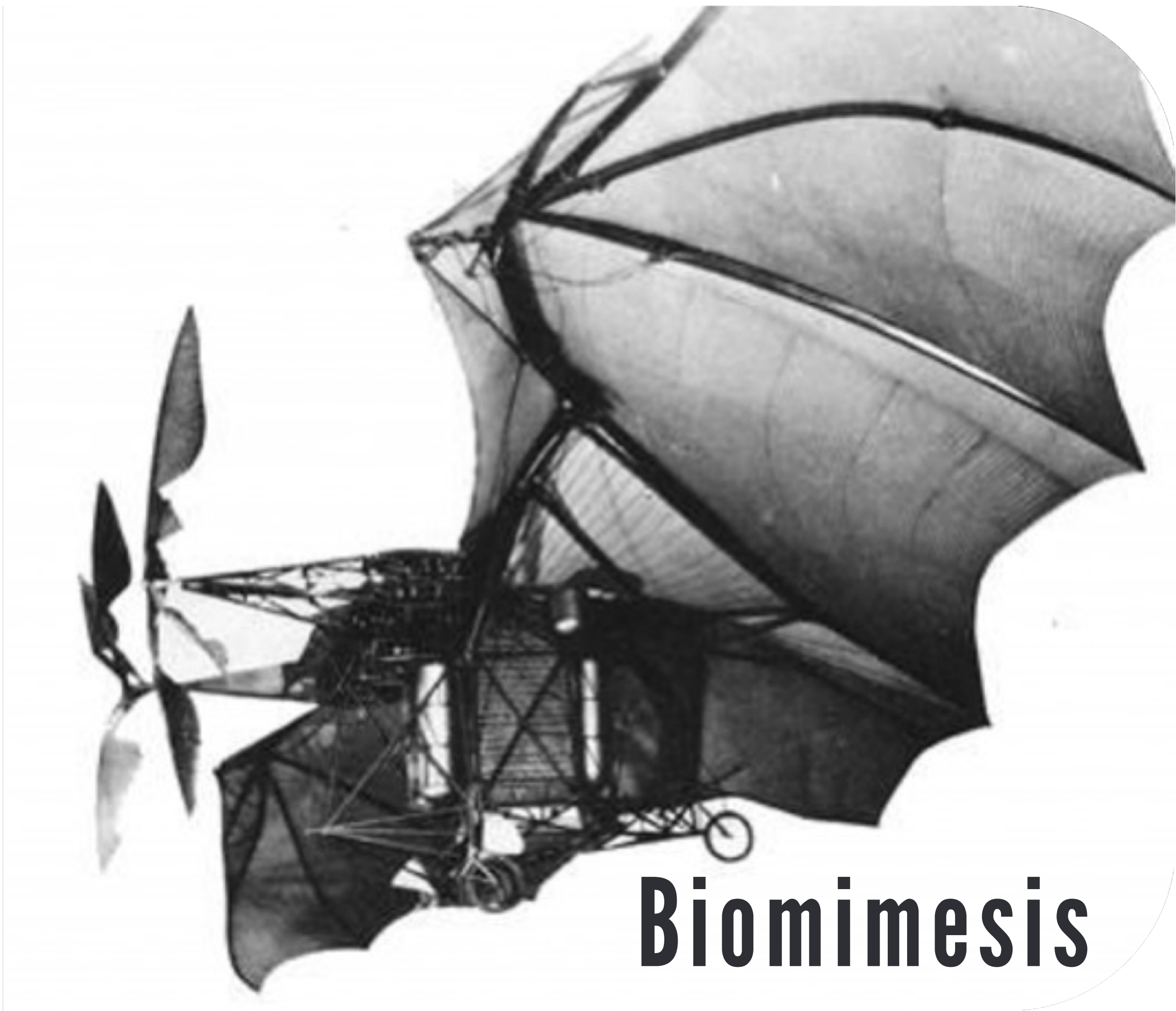
Atlas-based analysis



2010s-Present Deep-Learning

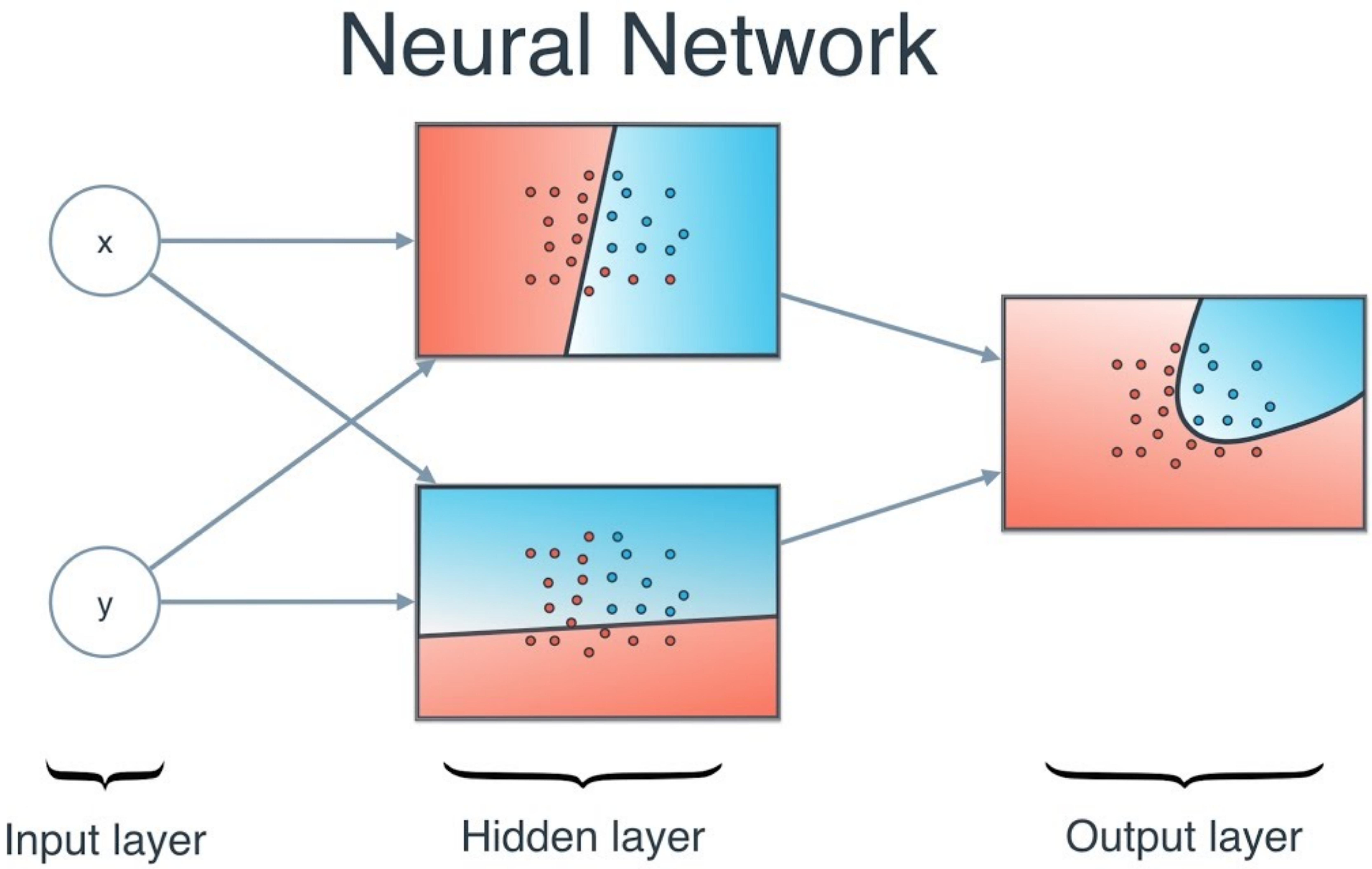
A model free / Data driven method



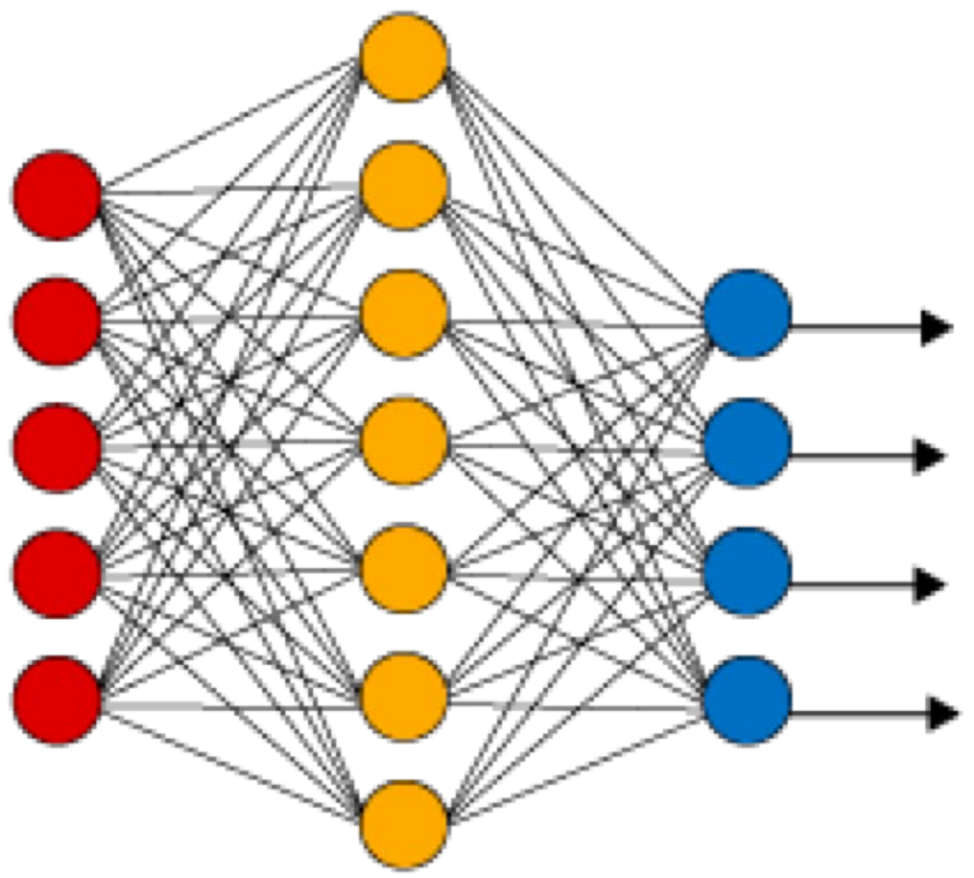


Deep-Learning

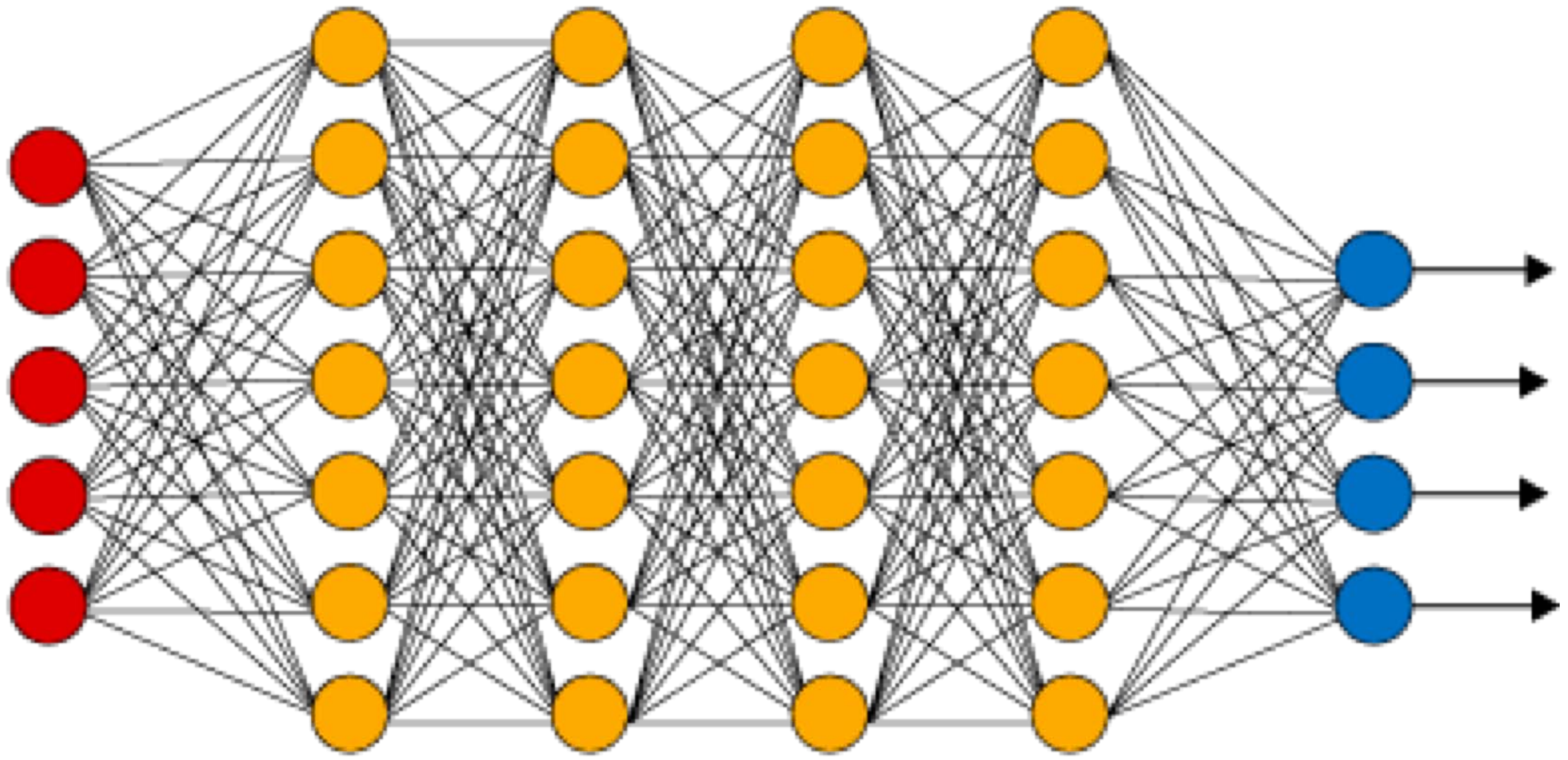
A model free / Data driven method

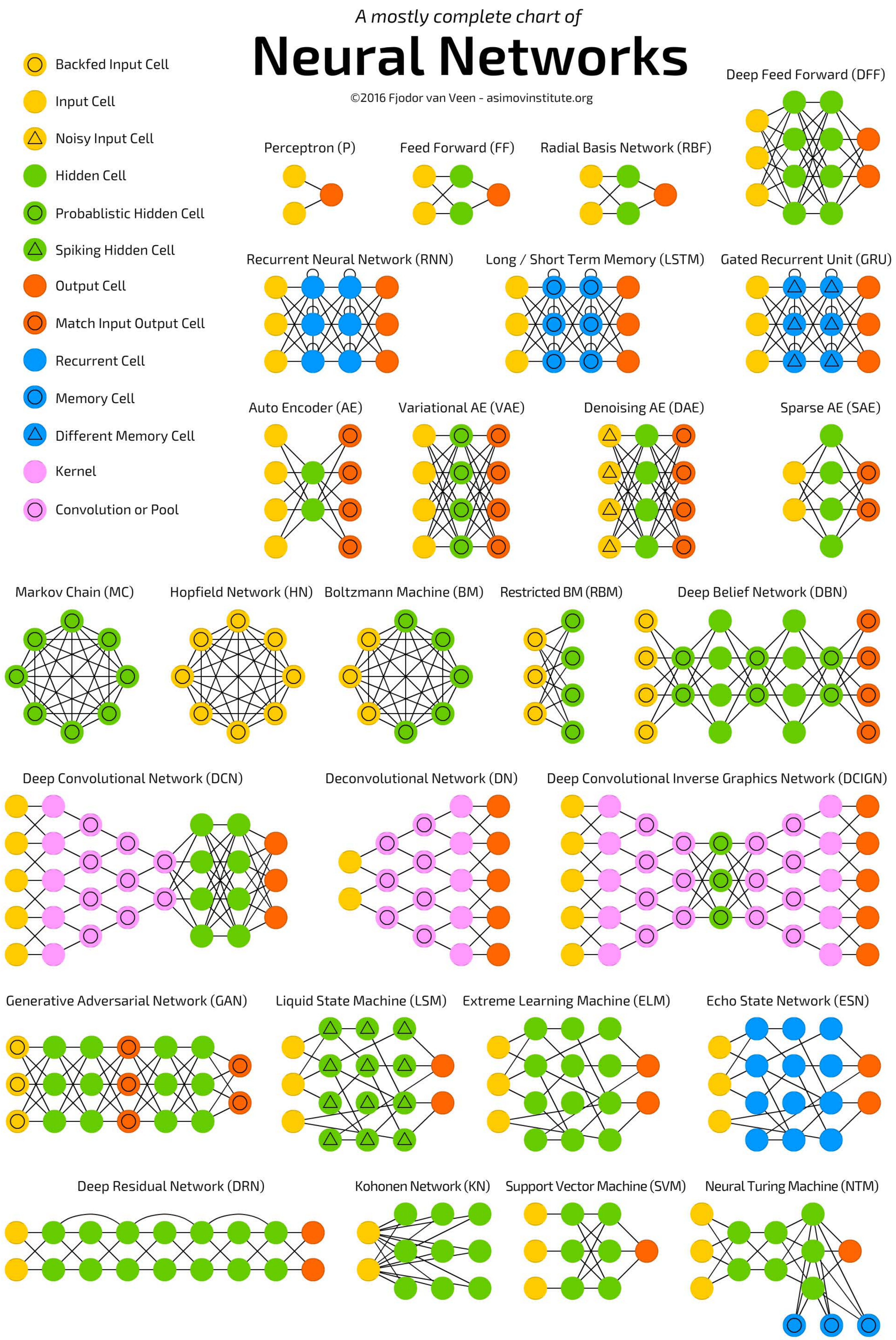


Simple Neural Network

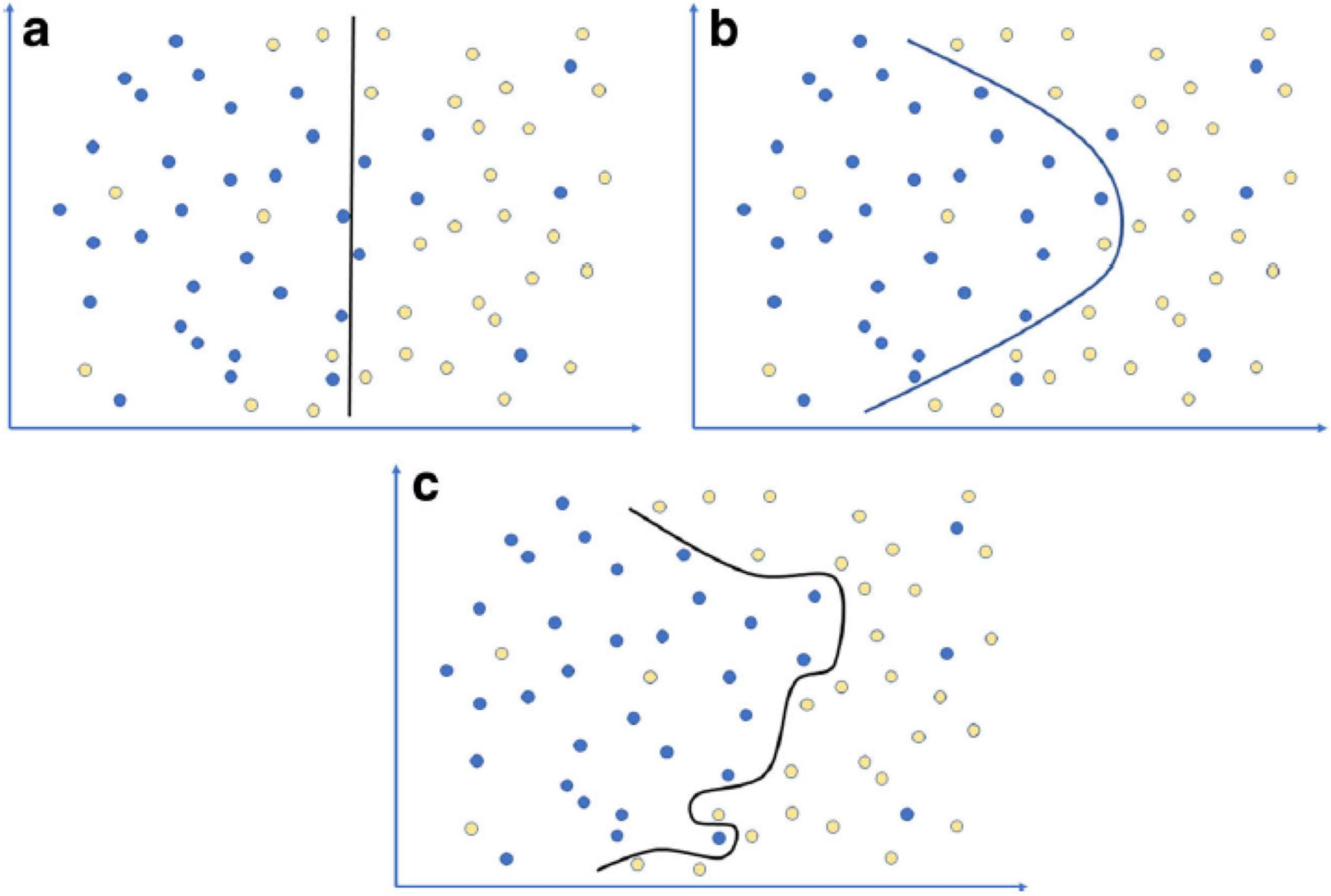


Deep Learning Neural Network



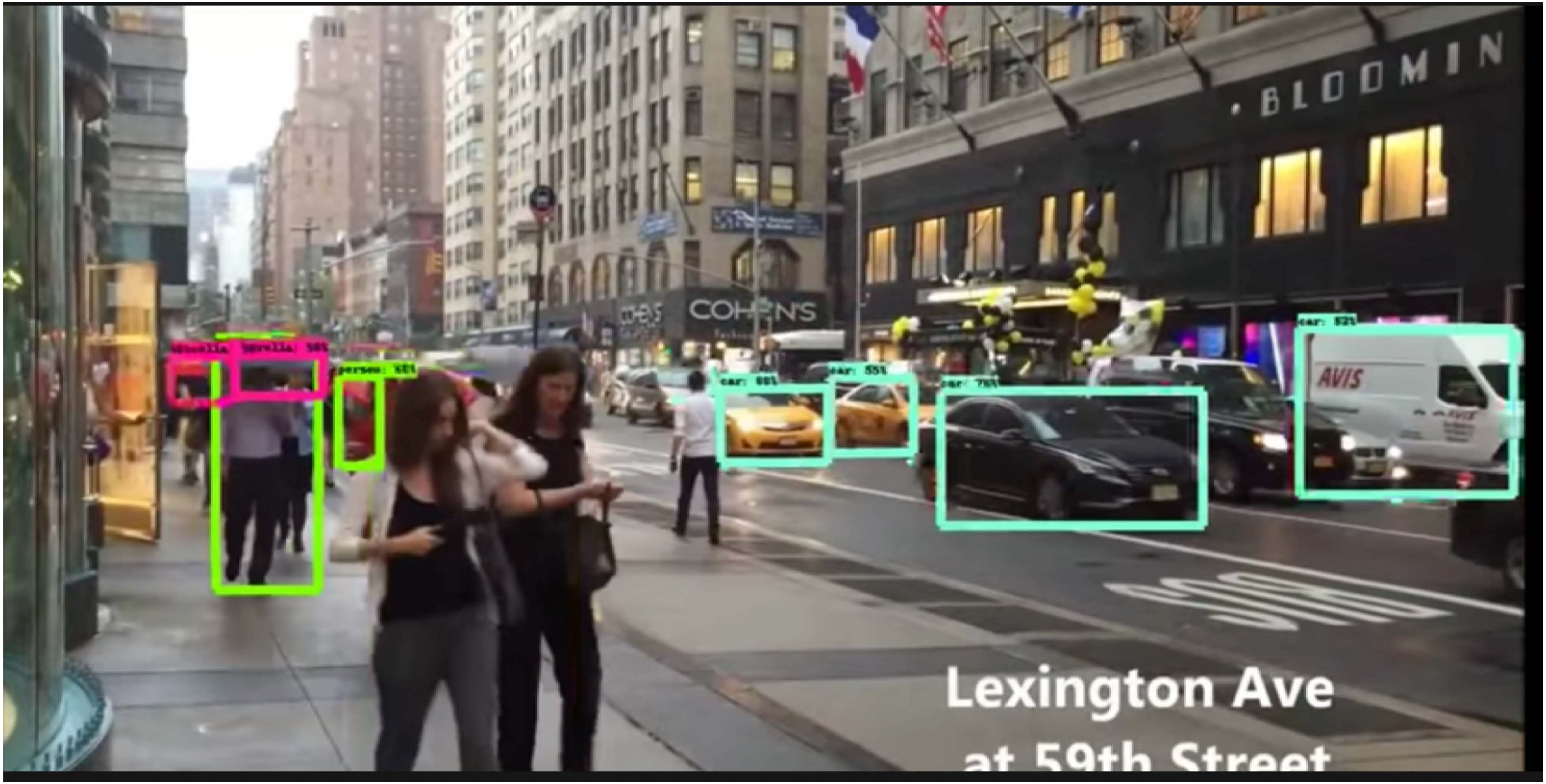


Overfitting

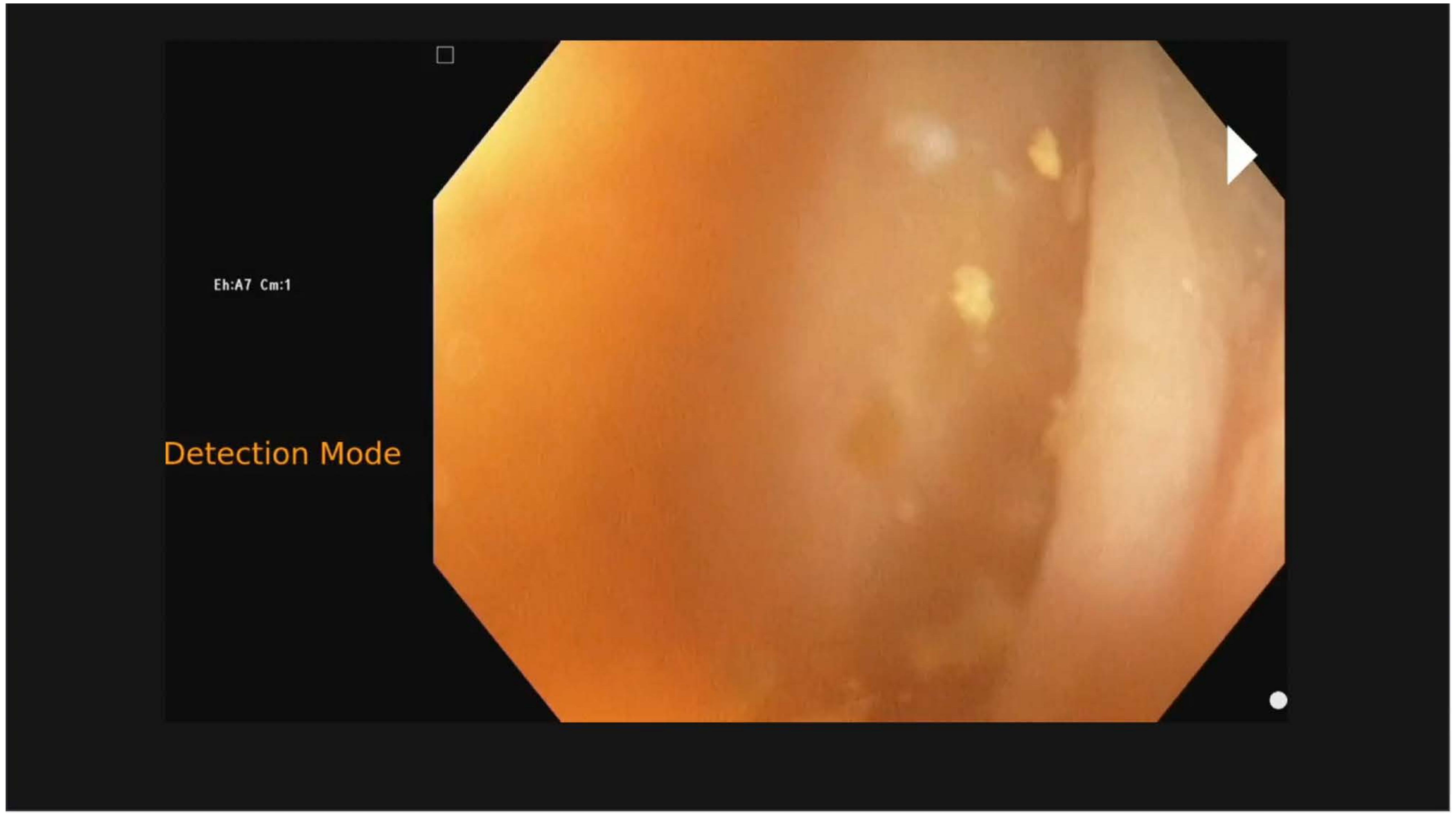


Deep-Learning

- Deep learning arose to take advantage of the large amounts of digital data
- It performs classification without the need of feature engineering done by humans
- Its power lies in that it can automatically discover and learn discriminatory features, in order to perform classification better than other algorithms or even humans.

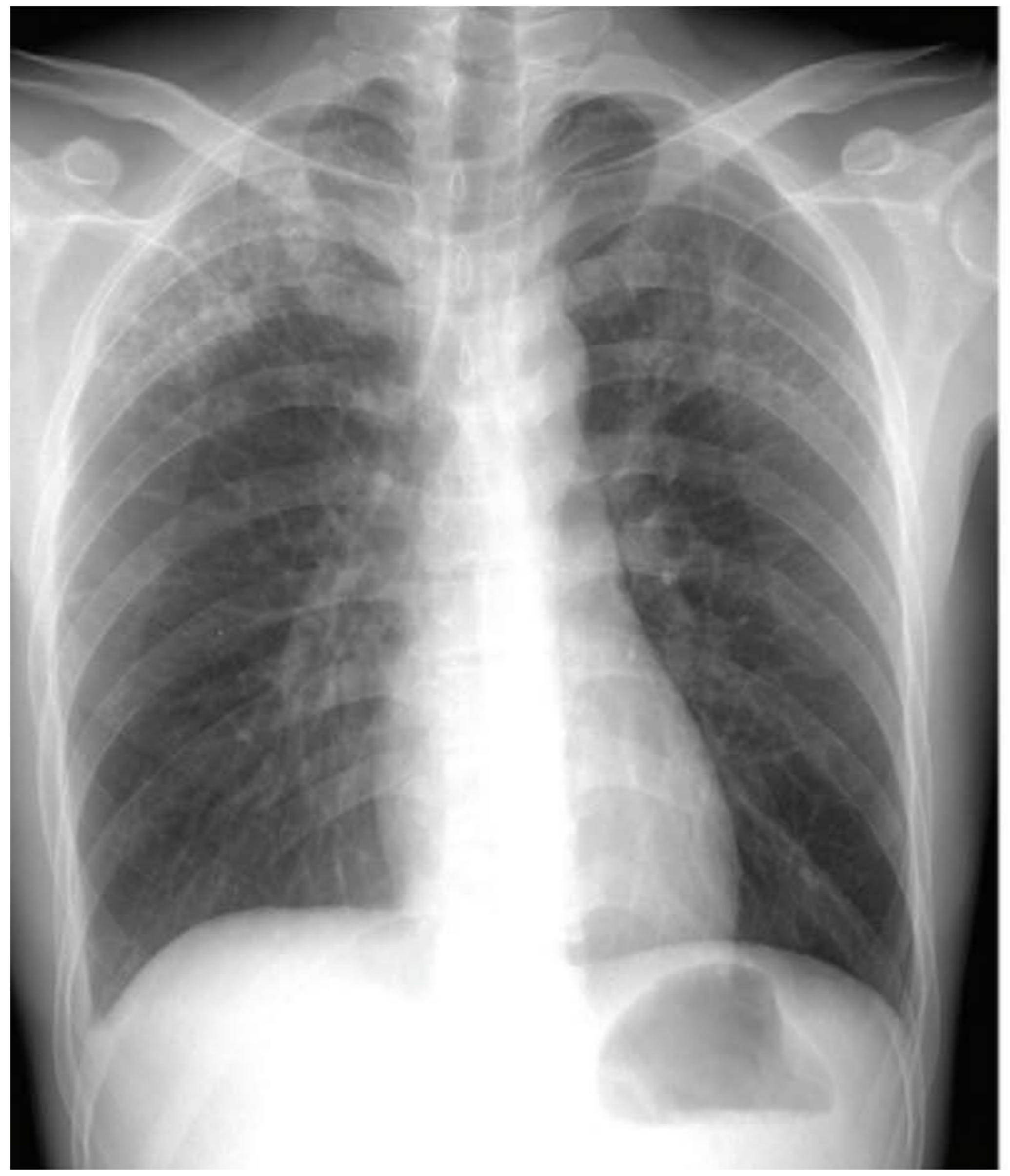


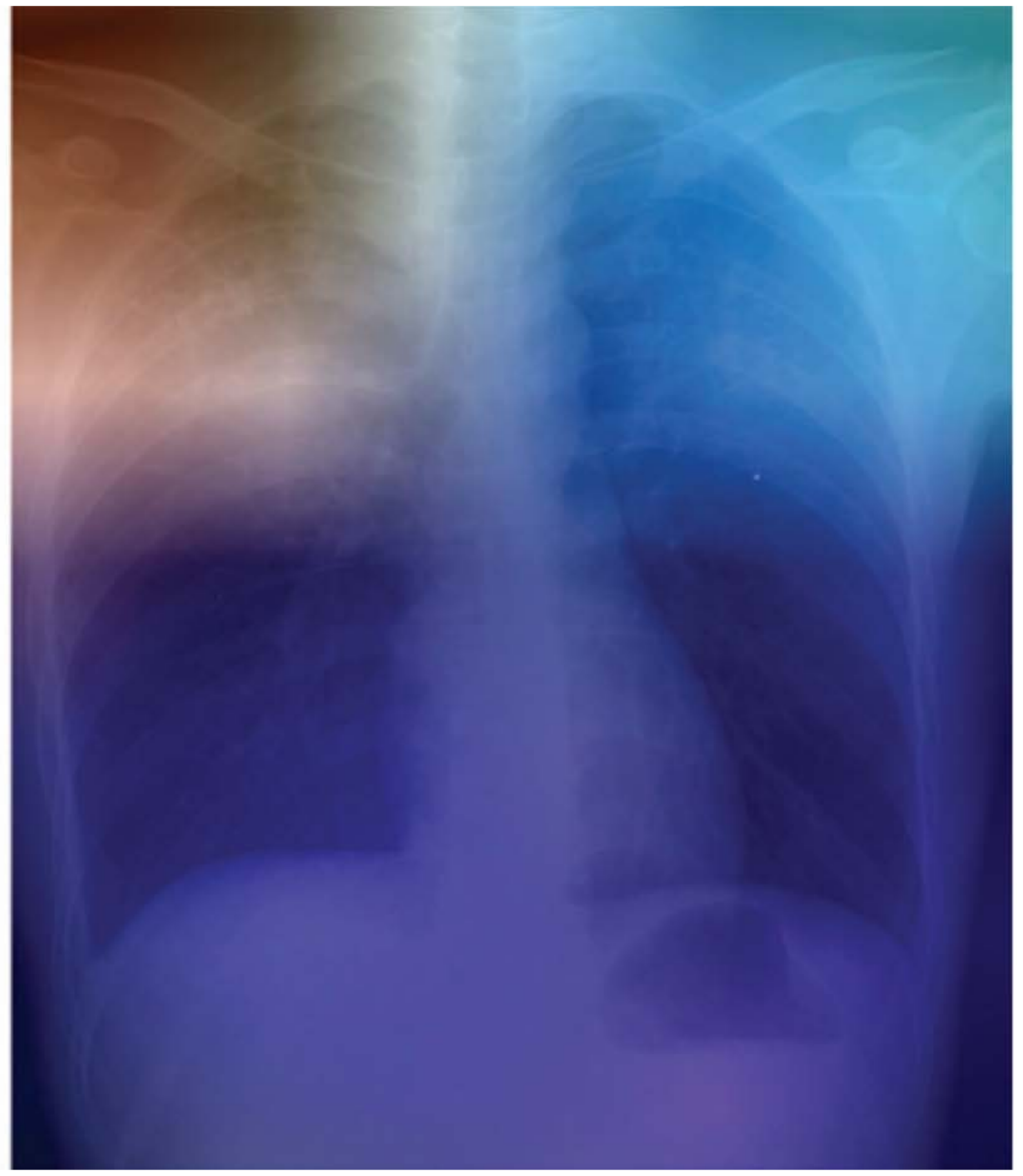
BLOOMIN
COHEN'S
AVIS
Lexington Ave
at 59th Street

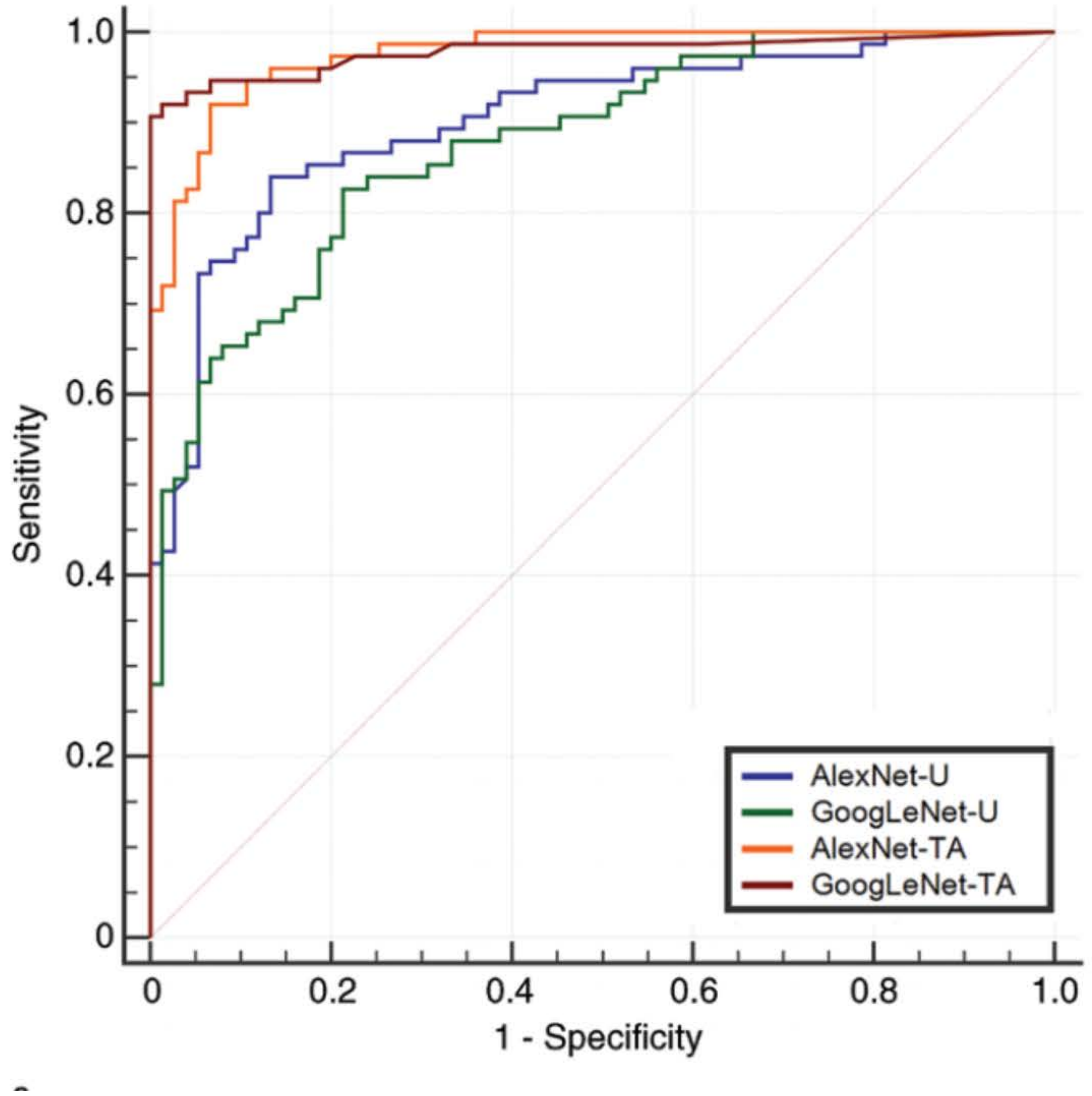

Eh:A7 Cm:1
Detection Mode

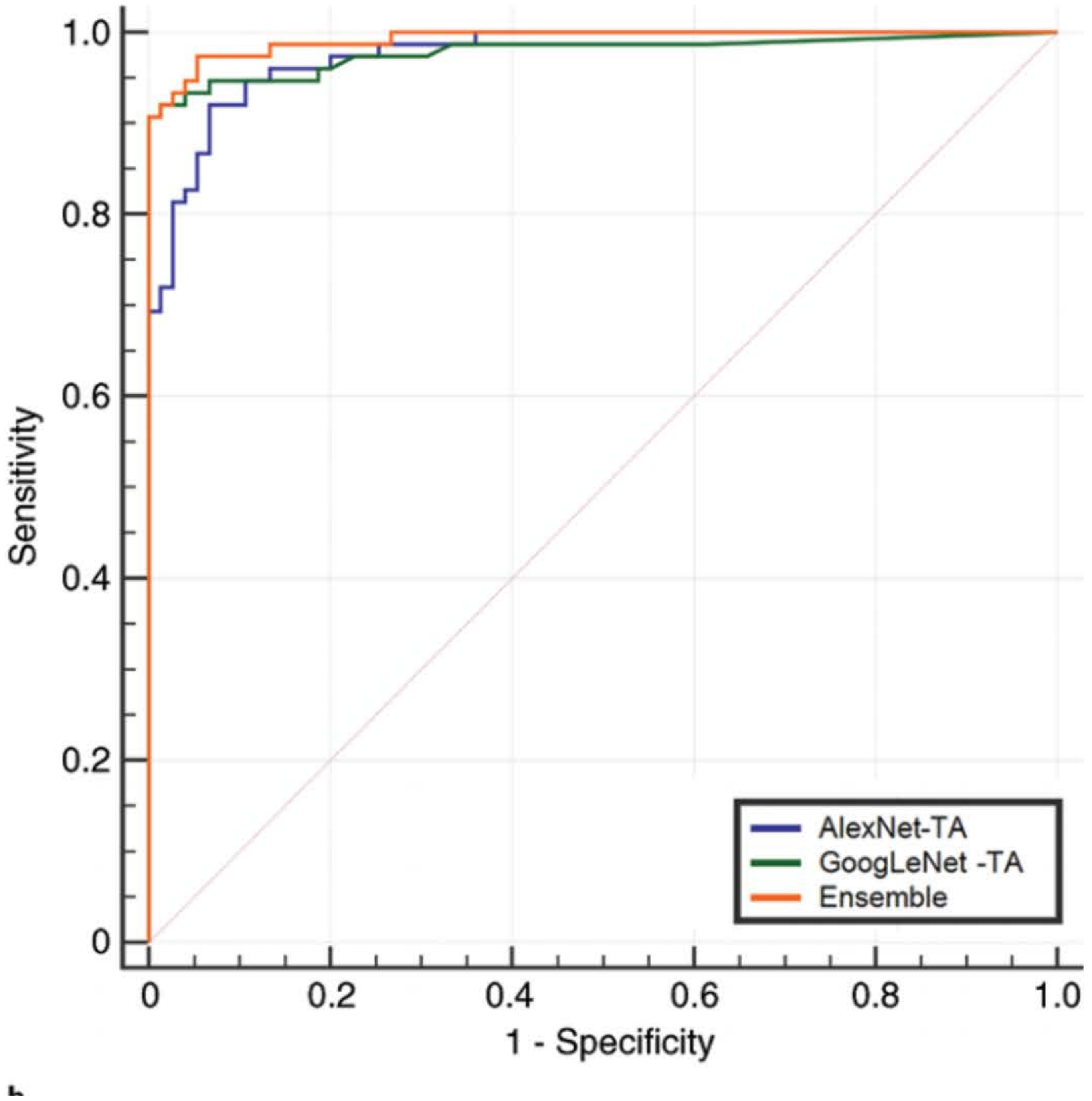
Deep Learning at Chest Radiography

Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks.



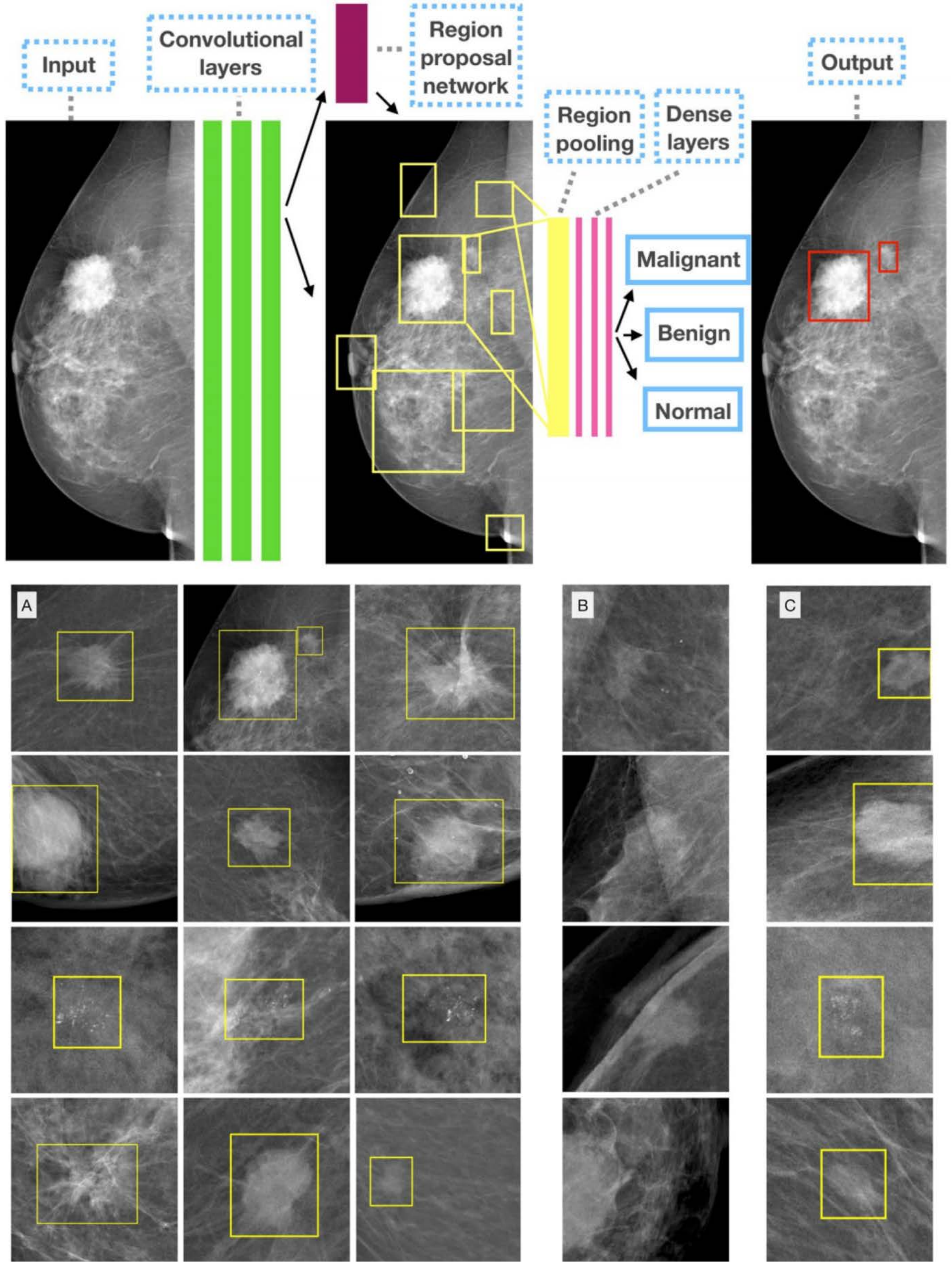


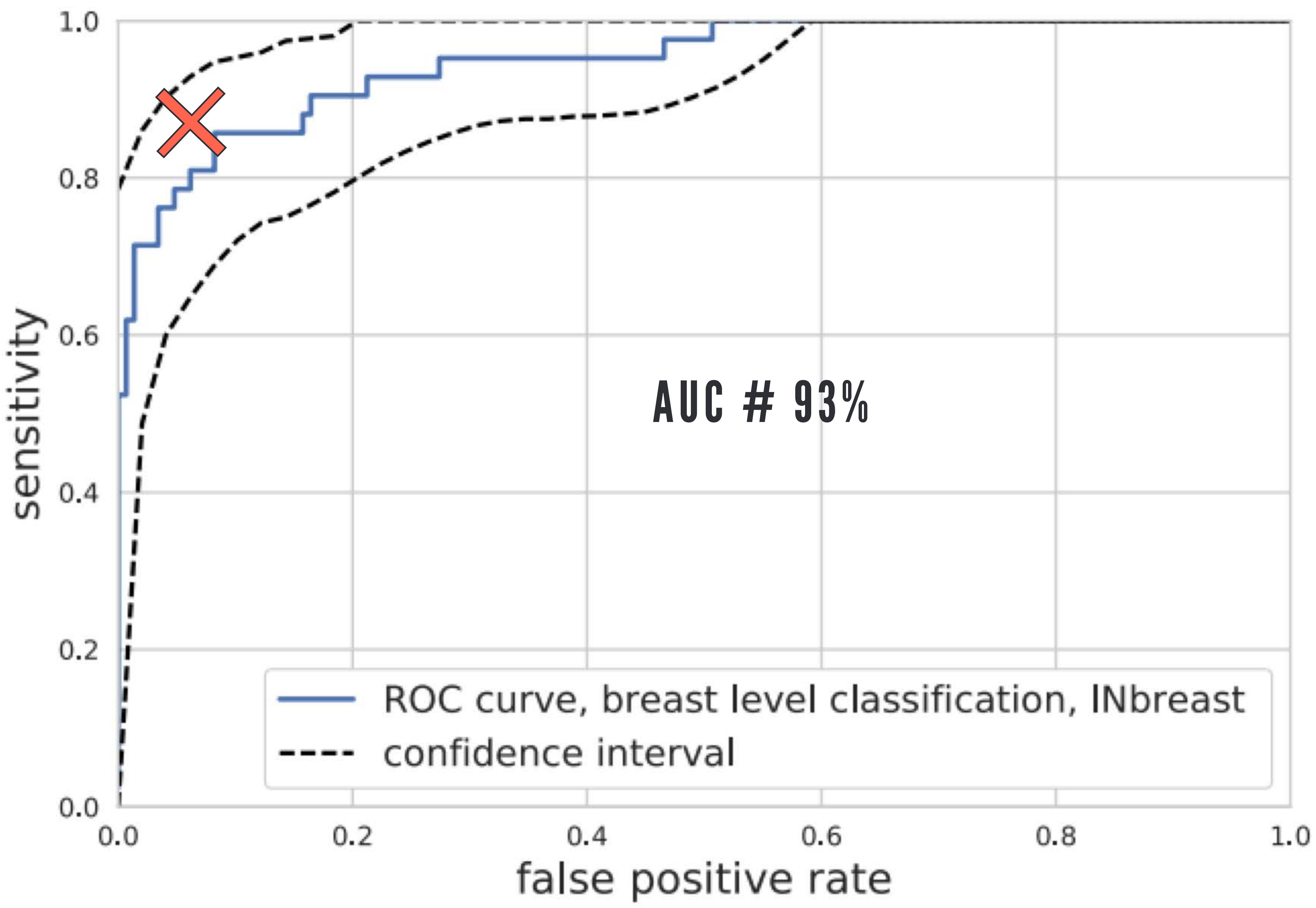




Lakhani and Sundaram. Radiology: Volume 284: Number 2—August 2017

Detecting and classifying lesions in mammograms with Deep Learning





Ribli d. et al. Nature Scientific Reports (2018)

Technological solutions for domains rich in labelled data do not necessarily translate easily to the healthcare sector

Limitation 1: inefficient use of data

Data delusion

The progress is driven far more by data availability than by improvement in algorithms

To train for a particular task, DL requires far more data points than a human would

“you really need data mines”


Deep neural networks
Shallow neural networks
Traditional learning methods
Performance
Amount of data

Limitation 2: lack of generalization

Once trained on a given task, deep networks are incredibly good at a very specific task. However, any such network can only perform that one task.

“specialization is for insects”

Limitation 3: lack of explainability

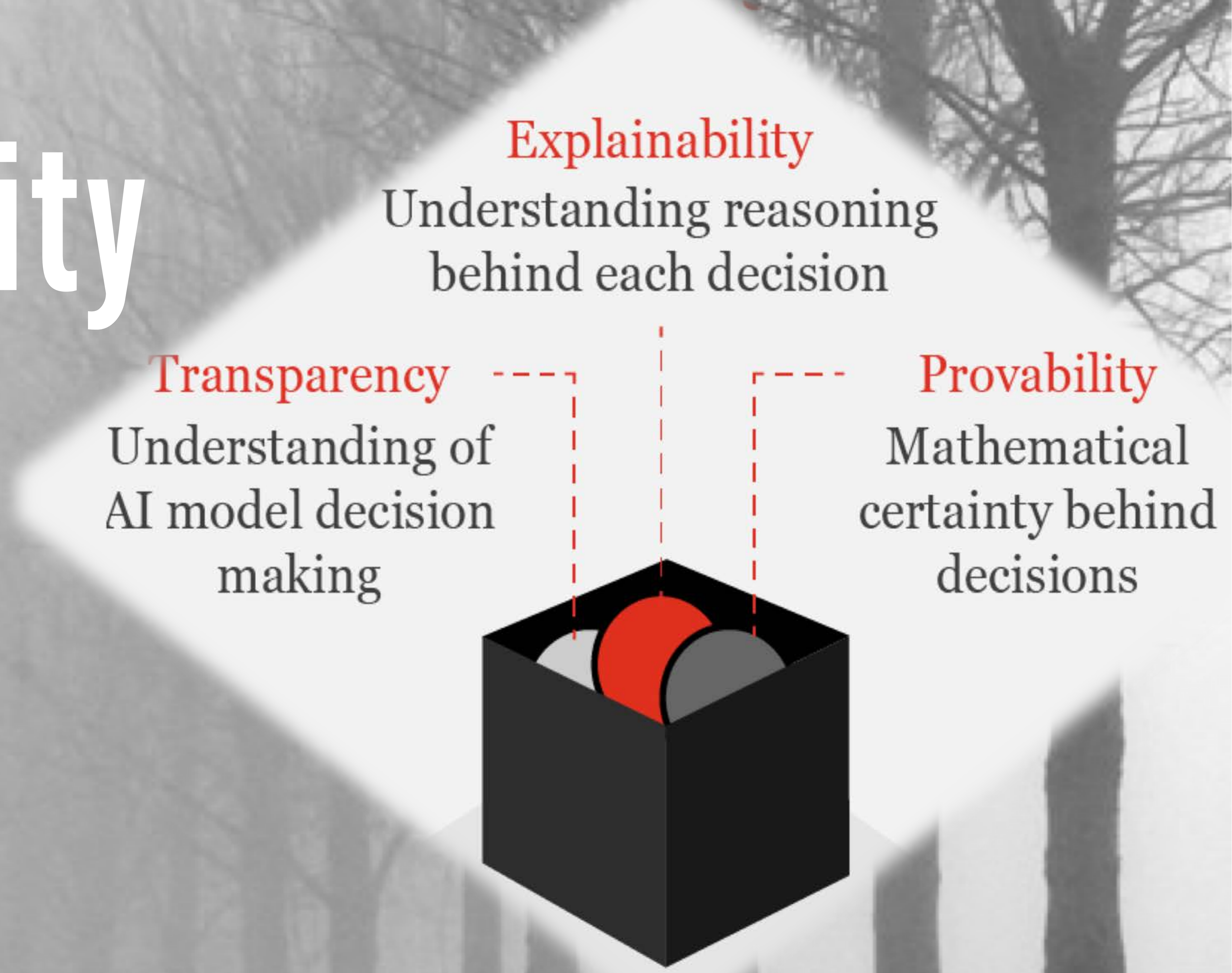
Neural networks are usually inscrutable to observers.

We know how they are constructed, and we know the data that goes in, but the reasons why certain conclusions are reached are usually unknown.

" It is generally infeasible to interpret DL features because their meaning depends on complex interactions with uninterpreted features in other layers"

Hinton G. JAMA September 18, 2018

Limitation 3: lack of explainability



Neural networks are usually inscrutable to observers.

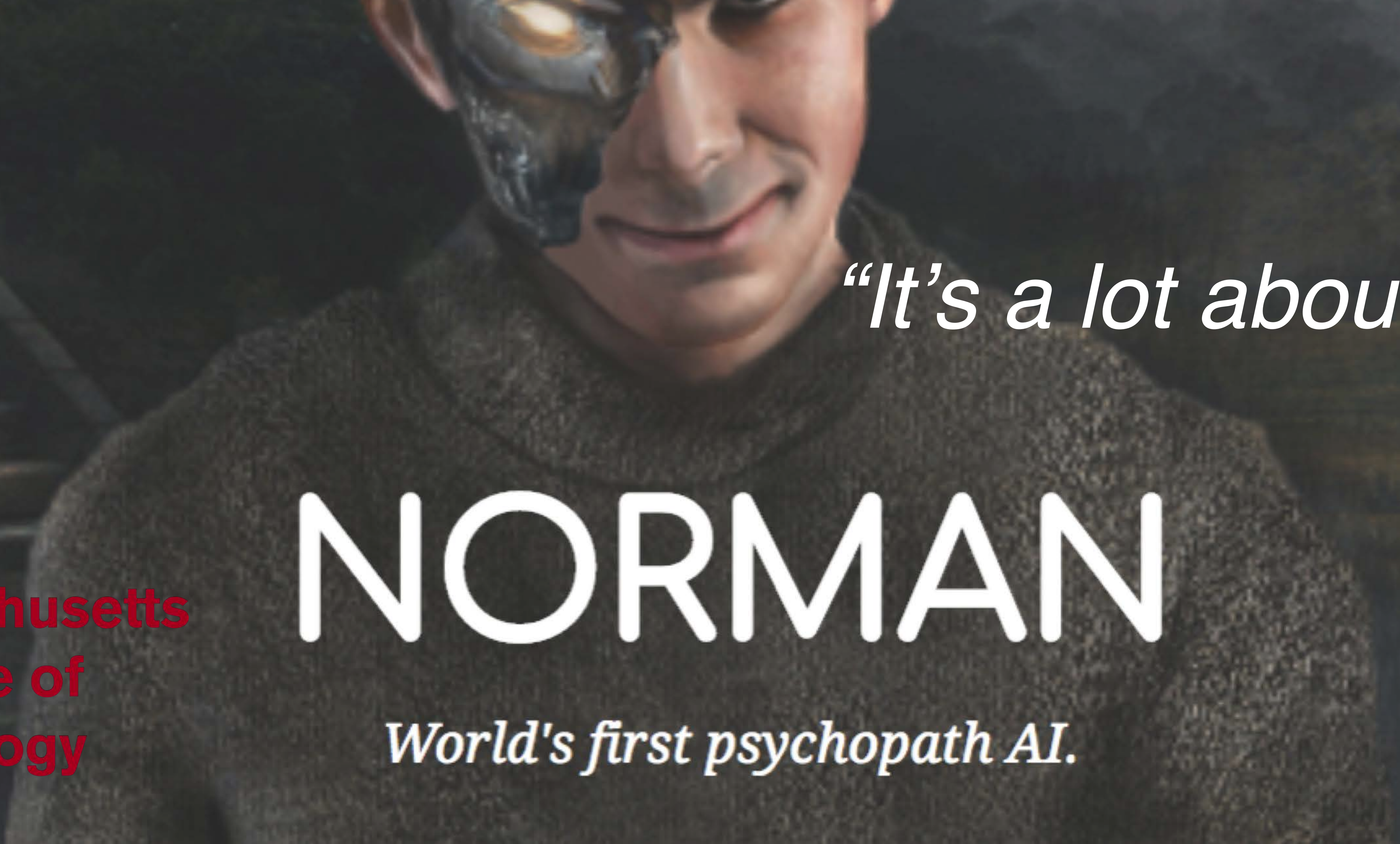
“We know how they are constructed, and we know the data that goes in, but the reasons why certain conclusions are reached are usually unknown”

Hinton G. JAMA September 18, 2018

Limitation 4: lack of transferability

The results of the models depends mainly on the quality and representativeness of the data sets

"It's a lot about Database"



Massachusetts Institute of Technology

RORSCHACH TEST

WHAT DOES AI SEE?

We trained Norman on Reddit, and compared captions with standard image captioning neural network.
Here is what both AIs see on Rorschach's inkblot tests.

CAPTIONS BY NORMAN AI

CAPTIONS BY STANDARD AI

INKBLOT #8
Standard AI sees:

"A PERSON IS HOLDING AN UMBRELLA IN THE AIR."

RORSCHACH TEST

WHAT DOES AI SEE?

We trained Norman on Reddit, and compared captions with standard image captioning neural network.
Here is what both AIs see on Rorschach's inkblot tests.

CAPTIONS BY NORMAN AI

CAPTIONS BY STANDARD AI

INKBLOT #8
Norman sees:

"MAN IS SHOT DEAD IN FRONT OF HIS SCREAMING WIFE."

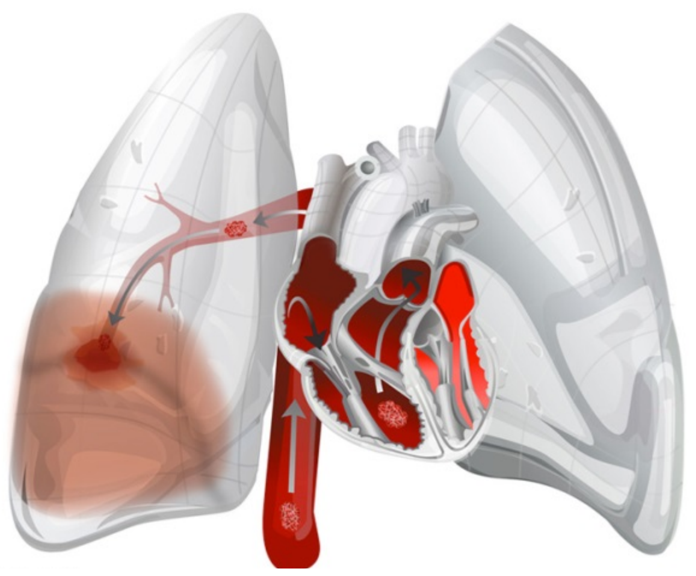
INKBLOT #8
Standard AI sees:

"A PERSON IS HOLDING AN UMBRELLA IN THE AIR."

“ the principle of deep learning implies that the tested data have the same distributions as the trained data ”

“ the machine will have as much bias as there is in the data that was used to train it ”

CT angiography for the Diagnostic of Pulmonary Embolism



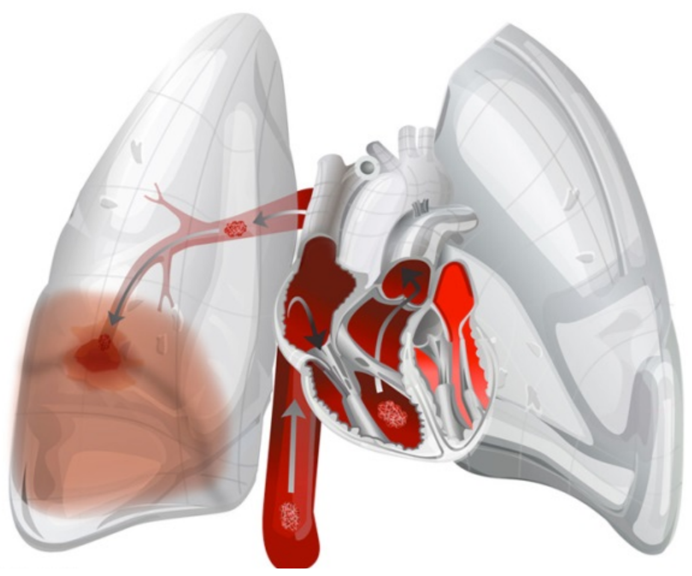
Likelihood Ratios
+LR = 9 / -LR=0.1

Pre-test Probability → CT Pulmonary Angiogram → Post Test Probality

Prevalence
30 %

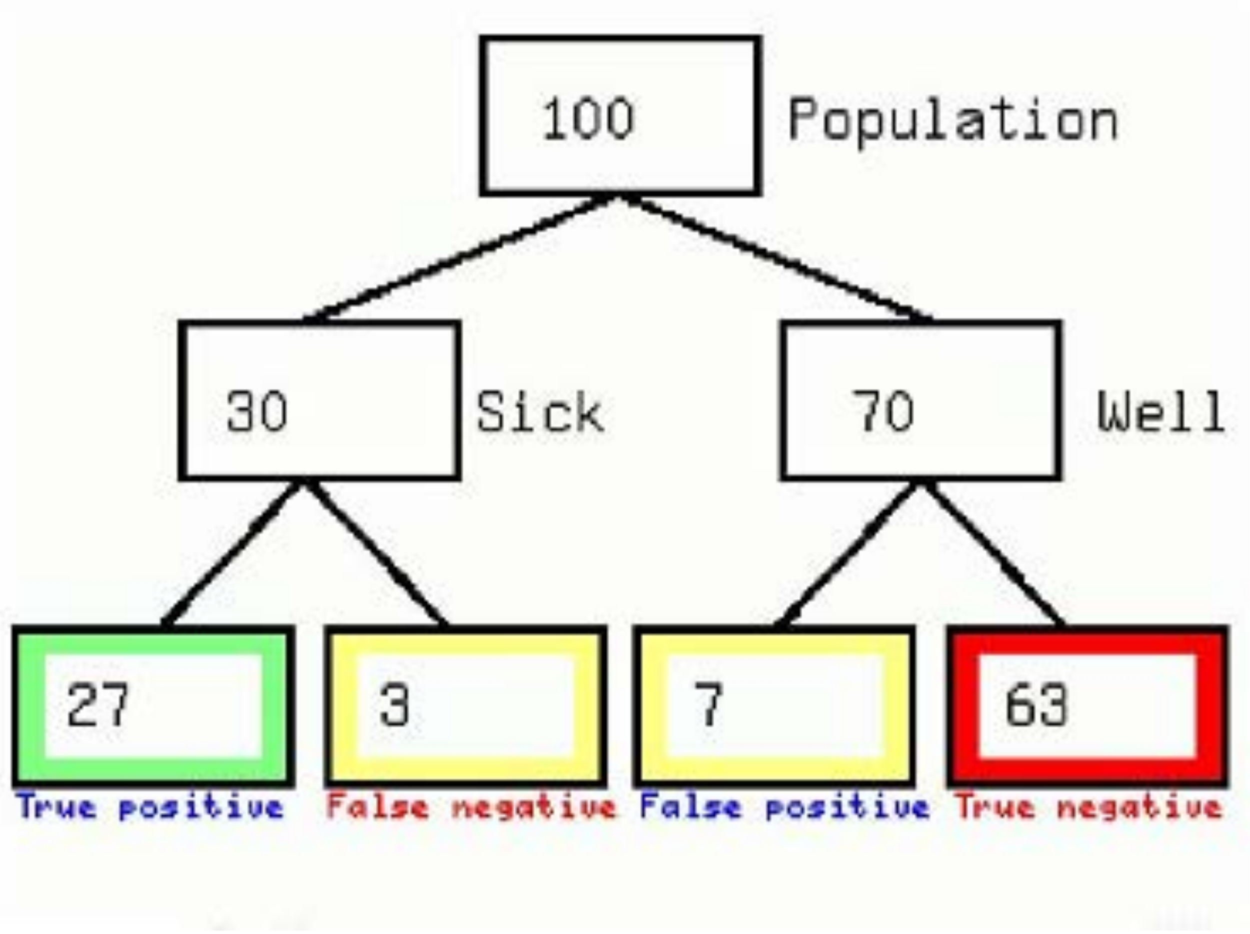
Sensitivity = 90 %
Specificity = 90 %

- When prevalence decrease from 30% to 10%



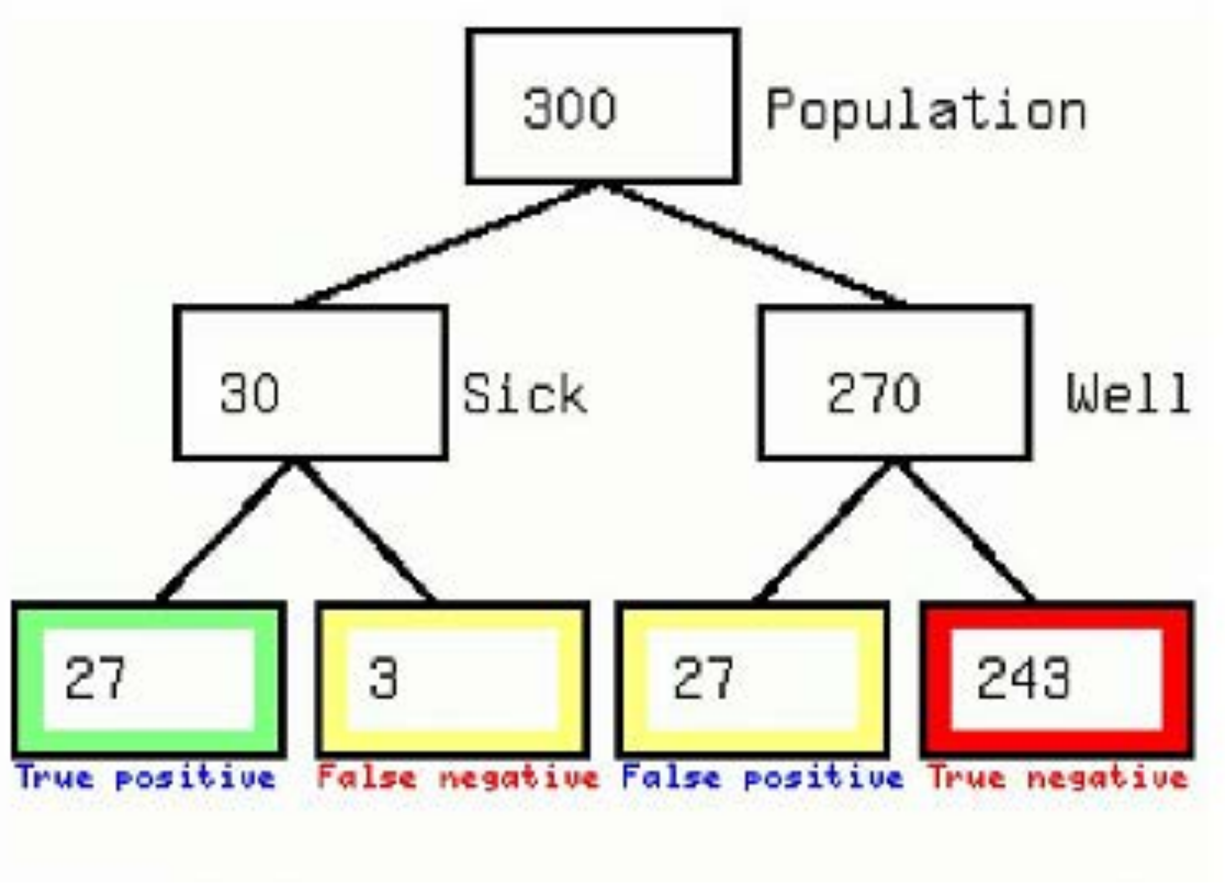
Numbers of patients with and without the disease who test positive and negative:

	Disease present	Disease absent	Total
Test positive	27	7	34
Test negative	3	63	66
Total	30	70	100



Numbers of patients with and without the disease who test positive and negative:

	Disease present	Disease absent	Total
Test positive	27	27	55
Test negative	3	243	245
Total	30	270	300



AI in Healthcare should be drive by Care Givers

- Invincible aura that AI has in popular perception

- A common mistake is to expect AI to find analyzable information in medical data even where none exists or for which there is no biological plausibility.

- Improperly designed AI experiments could misinform, mislead, or without critical analysis could result in patient harm.

Data-driven vs Hypothesis-driven

- The data-driven approach can be powerful and lead to novel insights
- Yet, it cannot replace the cognitive integration of complex information combining the semeiological analysis in a precise anatomical and physio-pathological context.
- These methodological approaches have been refined over centuries of scientific and medical thought, and they allow for solutions that are still inaccessible to AI.

Data-driven vs Hypothesis-driven

	PICO KEY STEP	PARTICULARITY OF AI IN RADIOLOGY
POPULATION	Describe the patient population. What are the most important patient characteristics and/or the problem type?	Describe the specific disease or condition and the methods tested
INTERVENTION	Describe the diagnostic test(s) being considered	Describe: - Method of data acquisition - Process of data quality assurance - Method of building the image database
COMPARISON (CONTROL)	Describe the current “gold standard” test for that disease or condition	Comparison of AI to human performance (radiologist ?)
OUTCOME	Determine a clinical outcome that will be used to evaluate the performance of the intervention	Determine how AI will be applied to screening, triage, staging Describe how the data set will be annotated for AI training

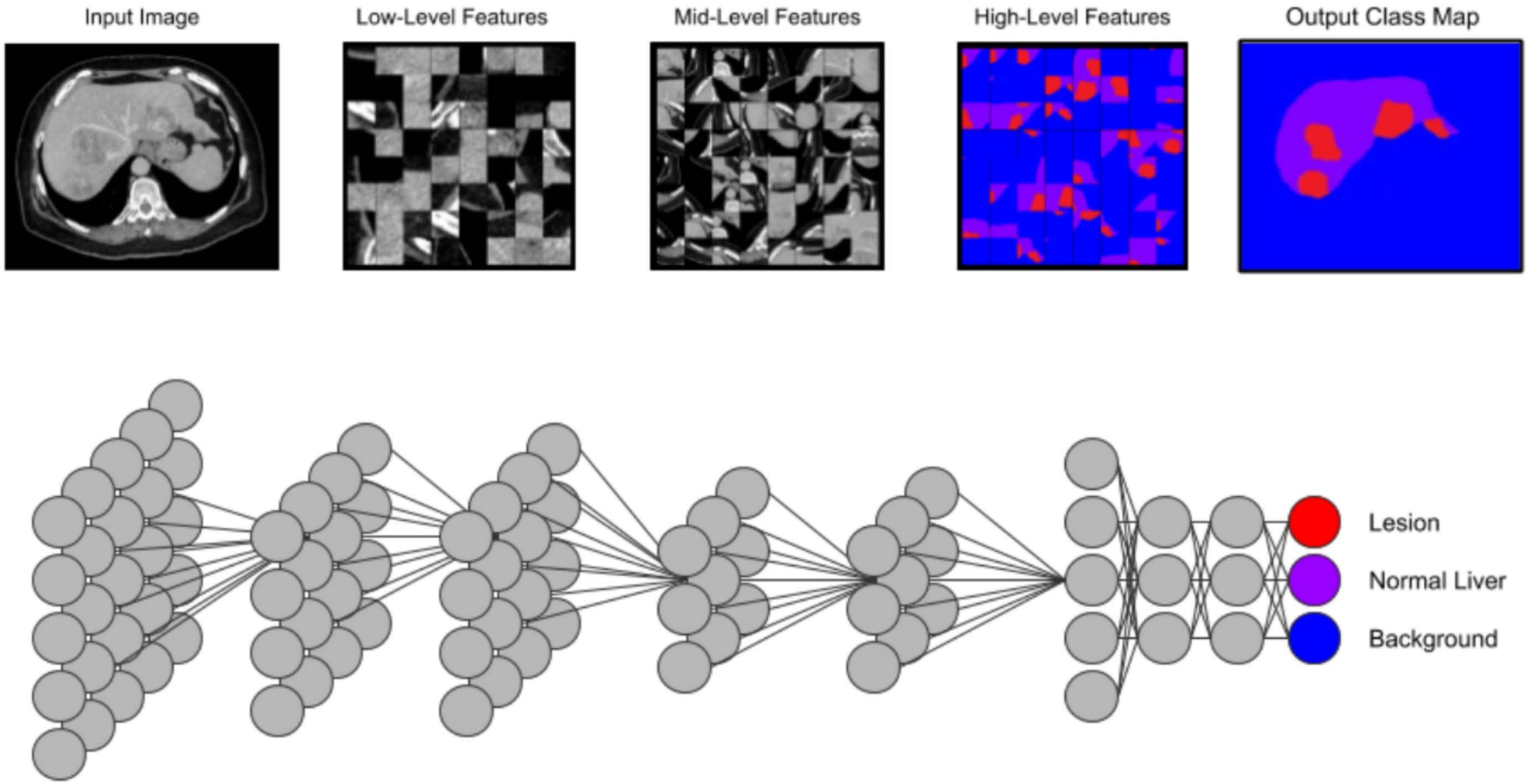
“Research hypotheses, whether AI-based or not, must be clinically relevant and answerable”

Savadjiev P, et al. European Radiology 2018

Rigorous translation pipelines is needed

1. Hypothesis generation
2. In silico replication to validate the predictive algorithm with data from other populations
3. Prototype intervention combining the predictive algorithm with information and decision support for health care teams and patients.
4. Feasibility and utility of the intervention in a pilot study in a health care setting.
5. Documents efficacy under clinical trial conditions.
6. Document effectiveness when the AI application is deployed in practice.

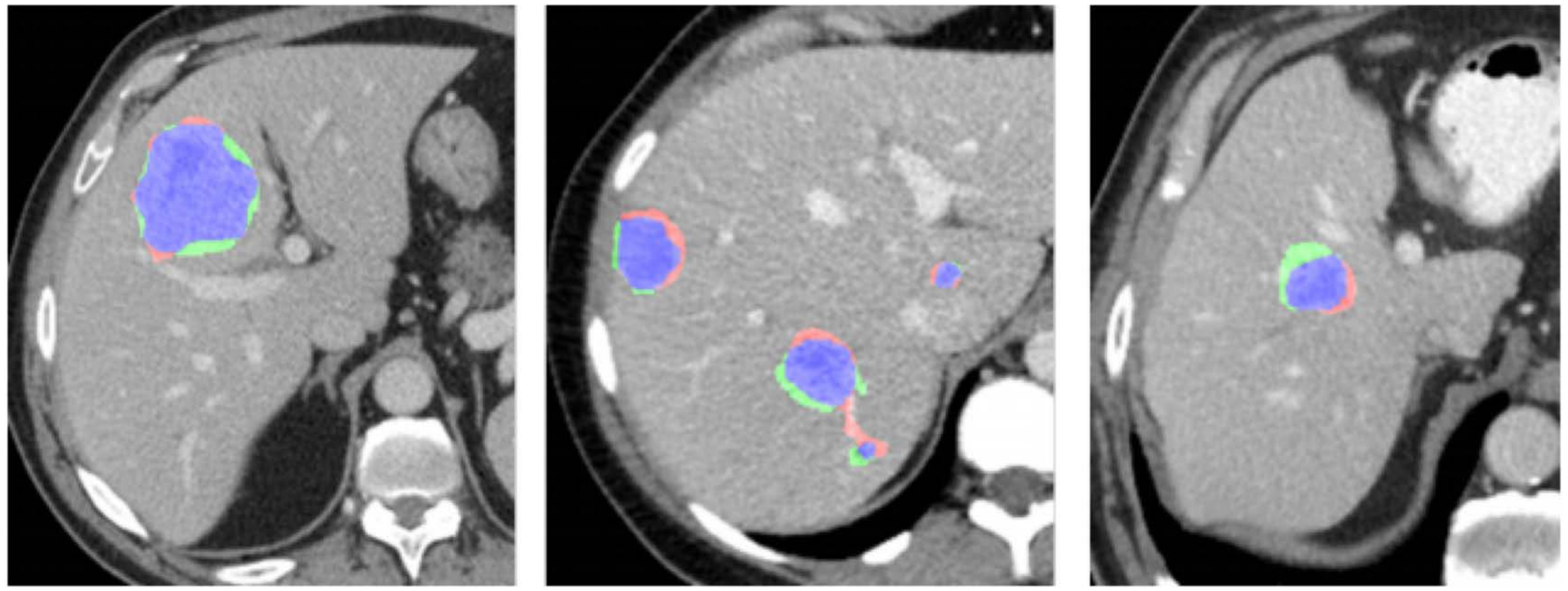
Application in Imaging

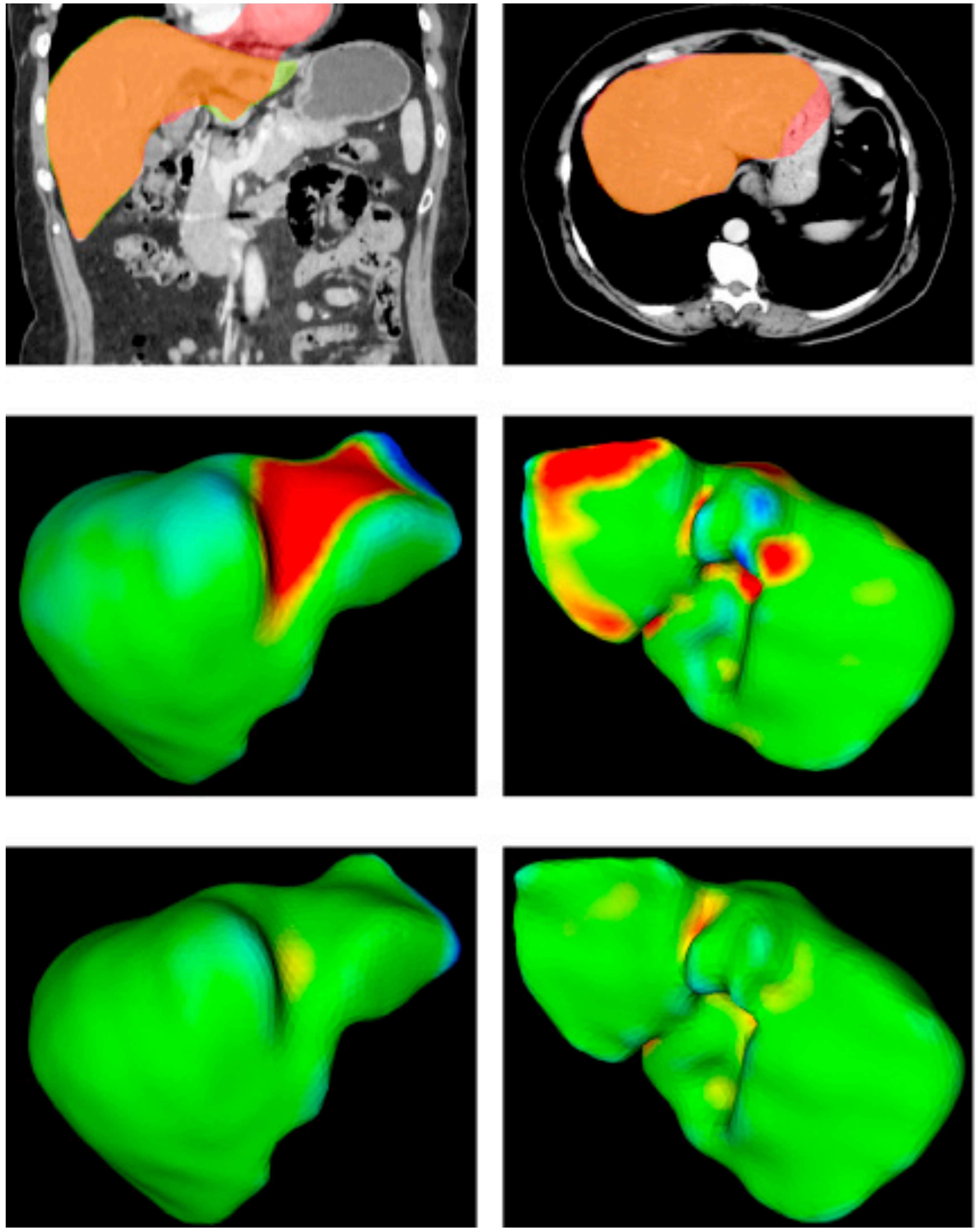


Data-driven/model-free approaches to automated knowledge discovery

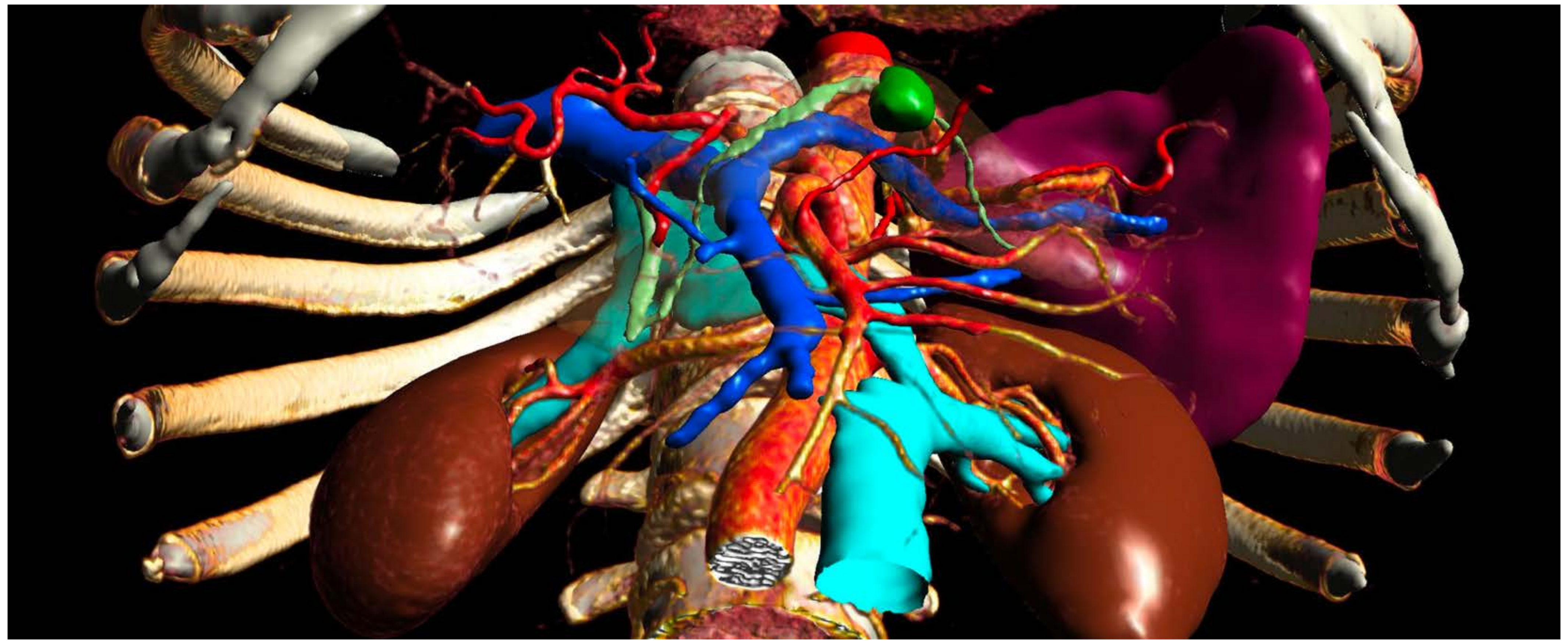
Application in Imaging (narrow tasks)

- Detection
- Segmentation
- Registration



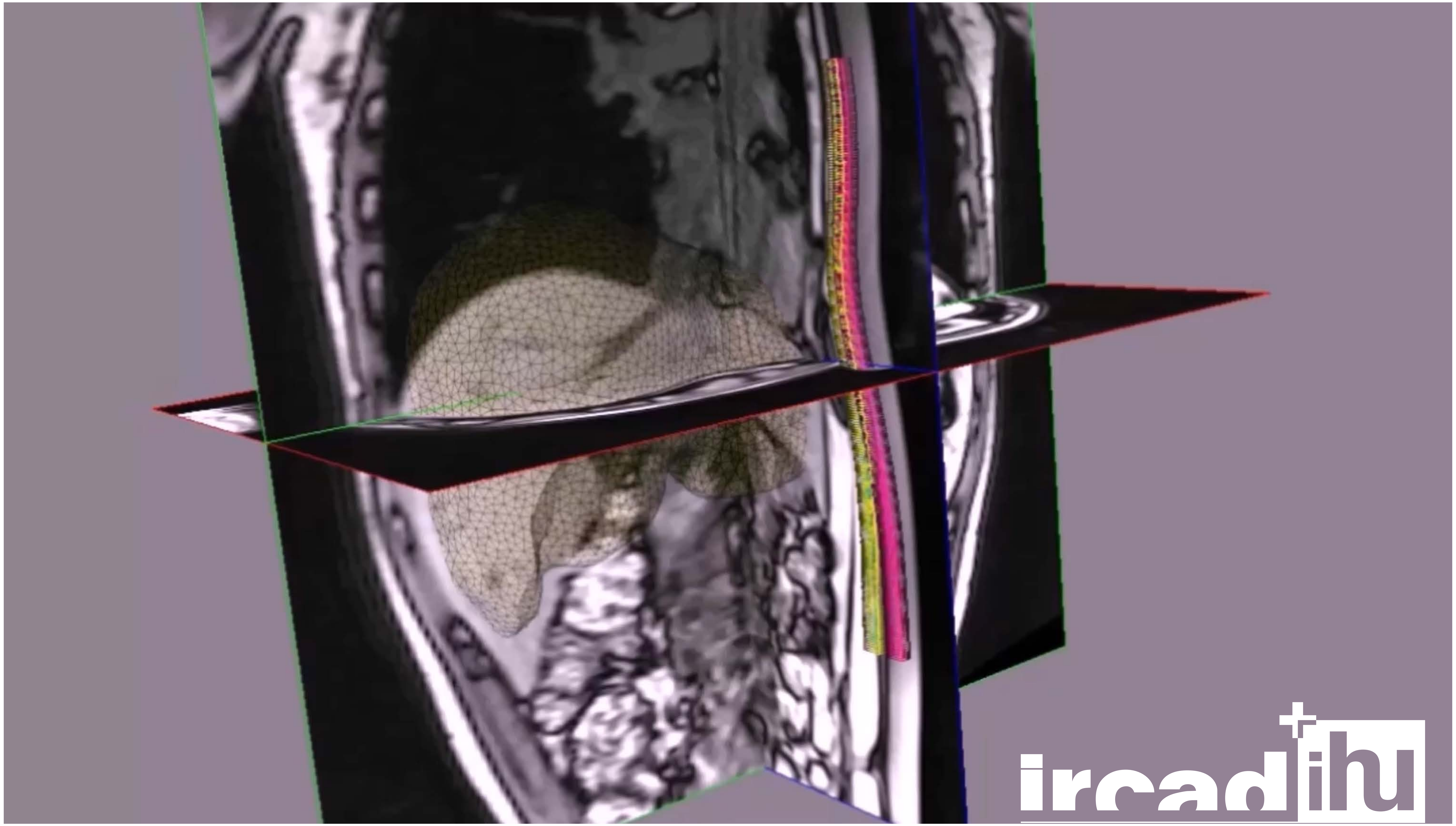


Real time movement adaptation



Real time movement adaptation

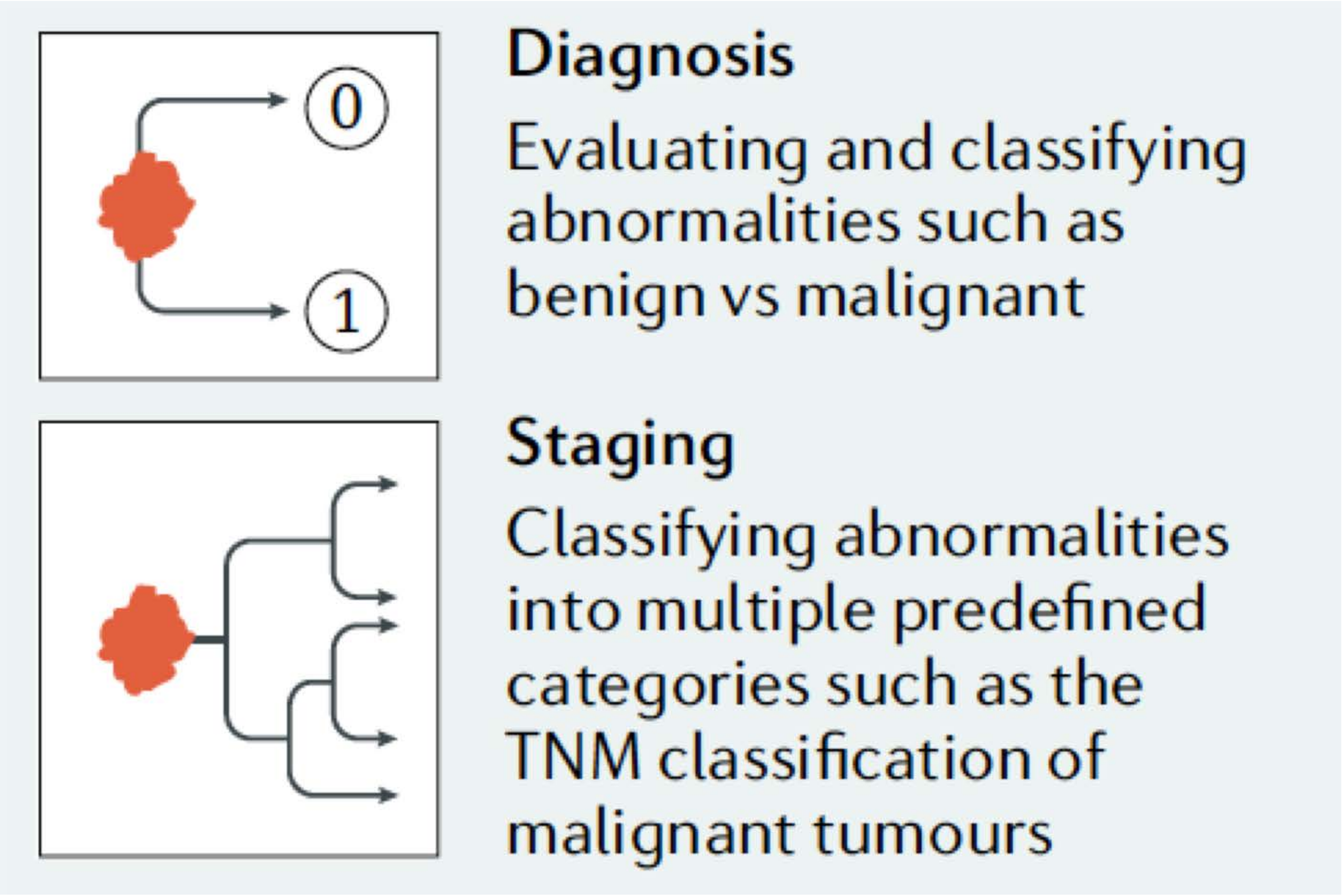
Real time movement adaptation



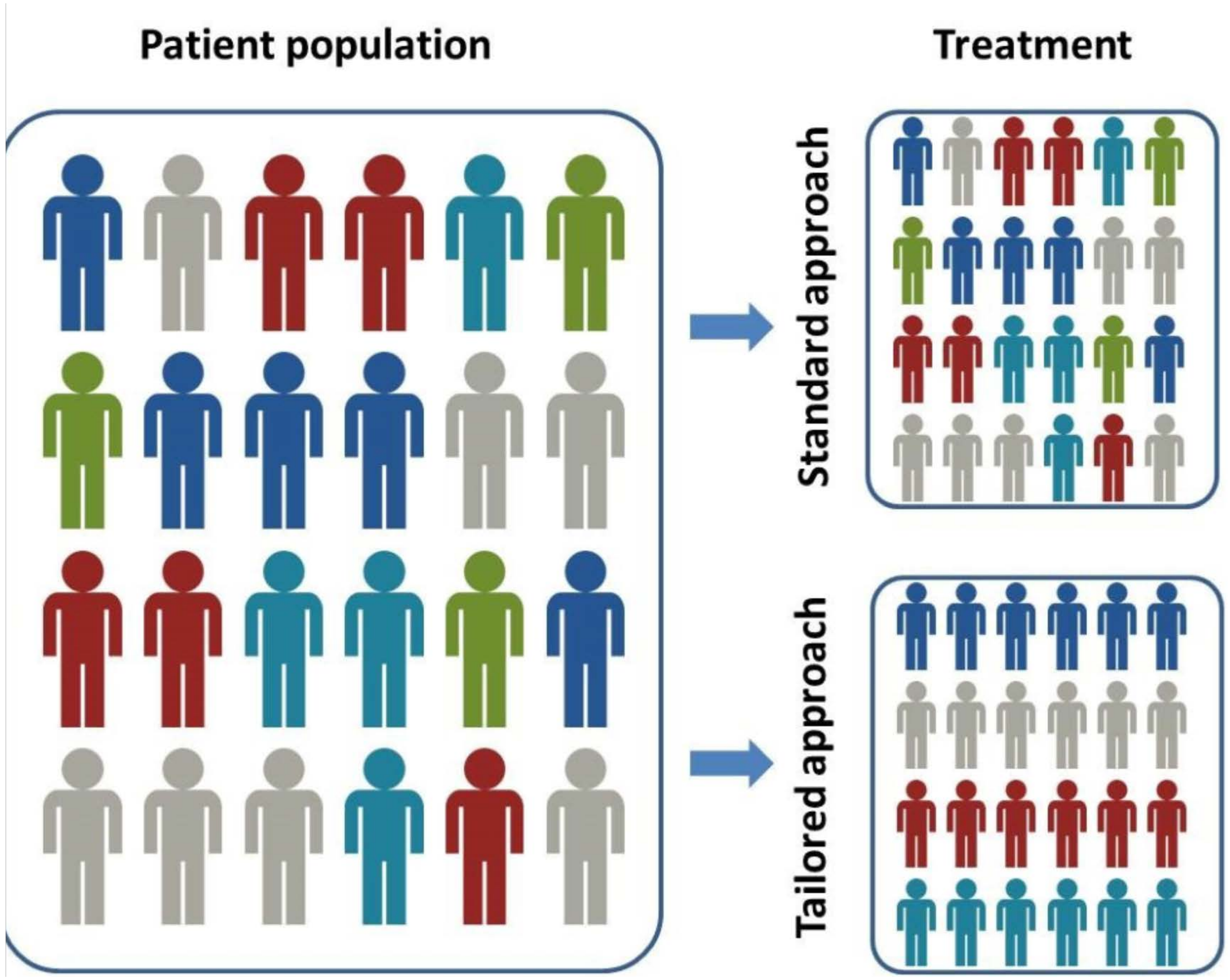
Application in Imaging (broader tasks)
Patient classification (diagnostic prediction, patient prognosis)

- **Separate Normal from not normal**
  - Definition of normal is very complex, depend of age, anatomical variant, presence of benign lesions
  - Normality should not be defined by radiologist reading but by the patient outcome for a specific disease

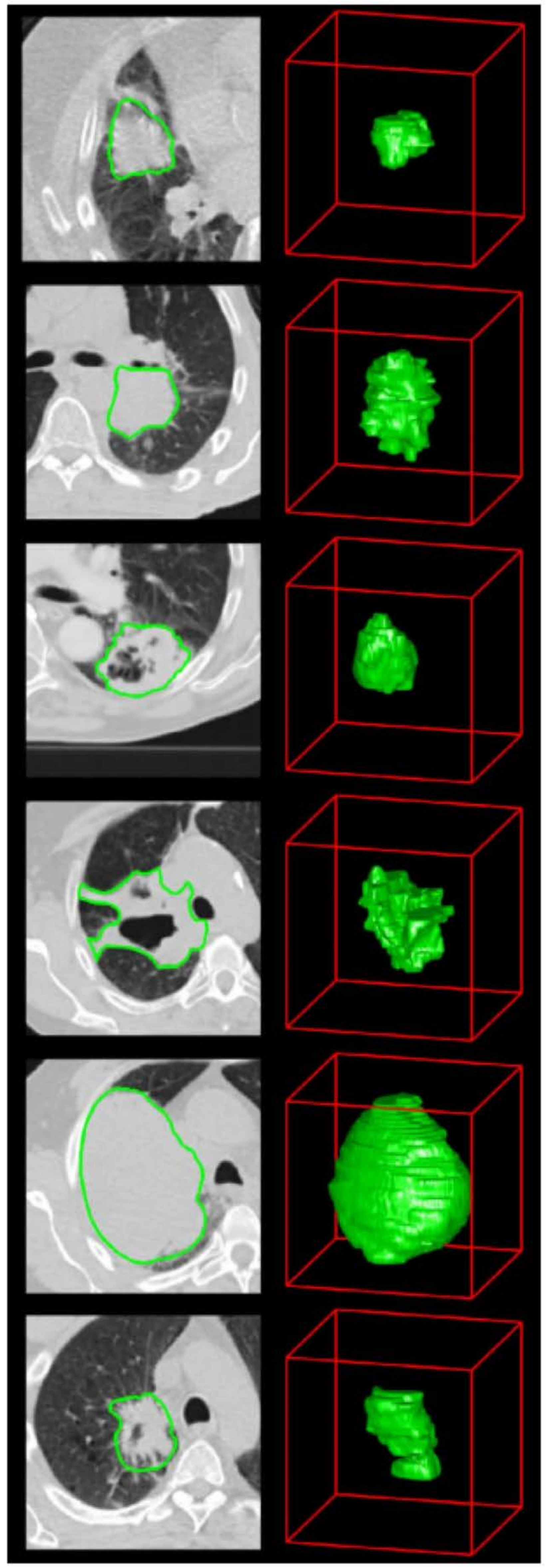
- **Classification of the disease**
  - Disease phenotyping
  - Patient stratification
  - Prediction and prognosis (probability to respond to treatment)



Prognostic & Predictive Biomarkers



Imaging Phenotype



nature COMMUNICATIONS

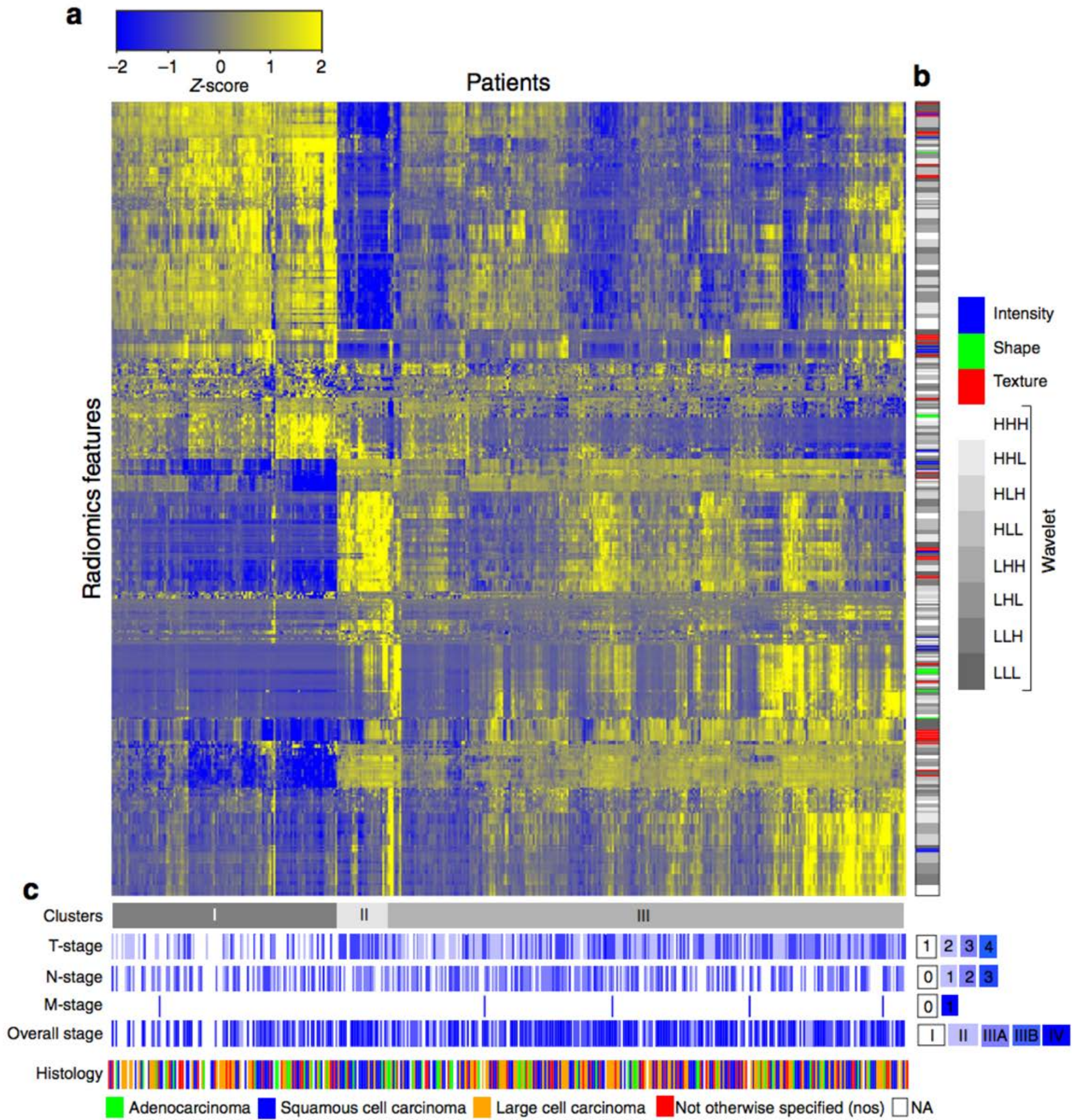
ARTICLE

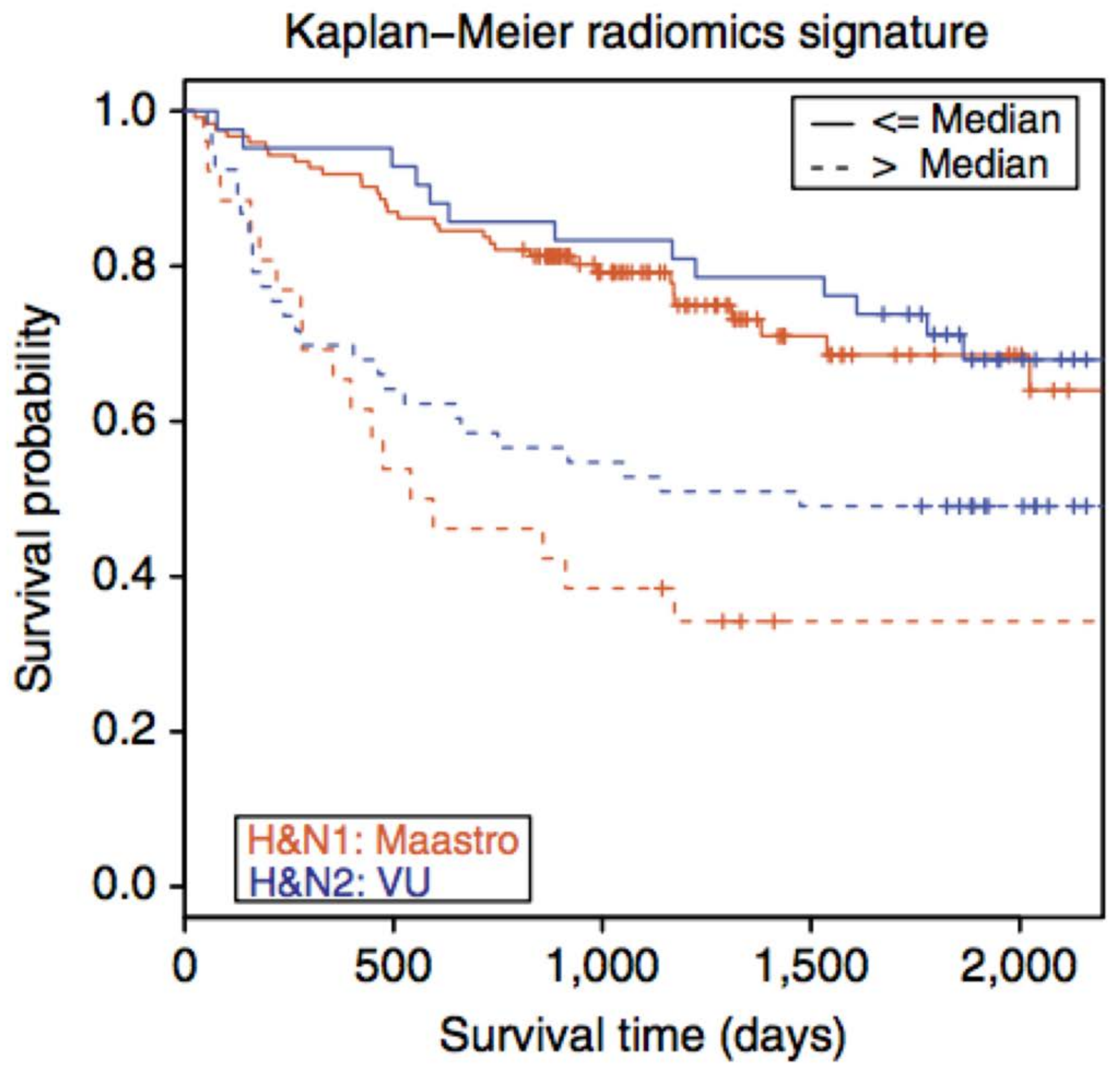
Received 25 Nov 2013 | Accepted 29 Apr 2014 | Published 3 Jun 2014 | Updated 7 Aug 2014

DOI: 10.1038/ncomms5006

Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach

Hugo J.W.L. Aerts[1,2,3,4,*], Emmanuel Rios Velazquez[1,2,*], Ralph T.H. Leijenaar[1], Chintan Parmar[1,2],

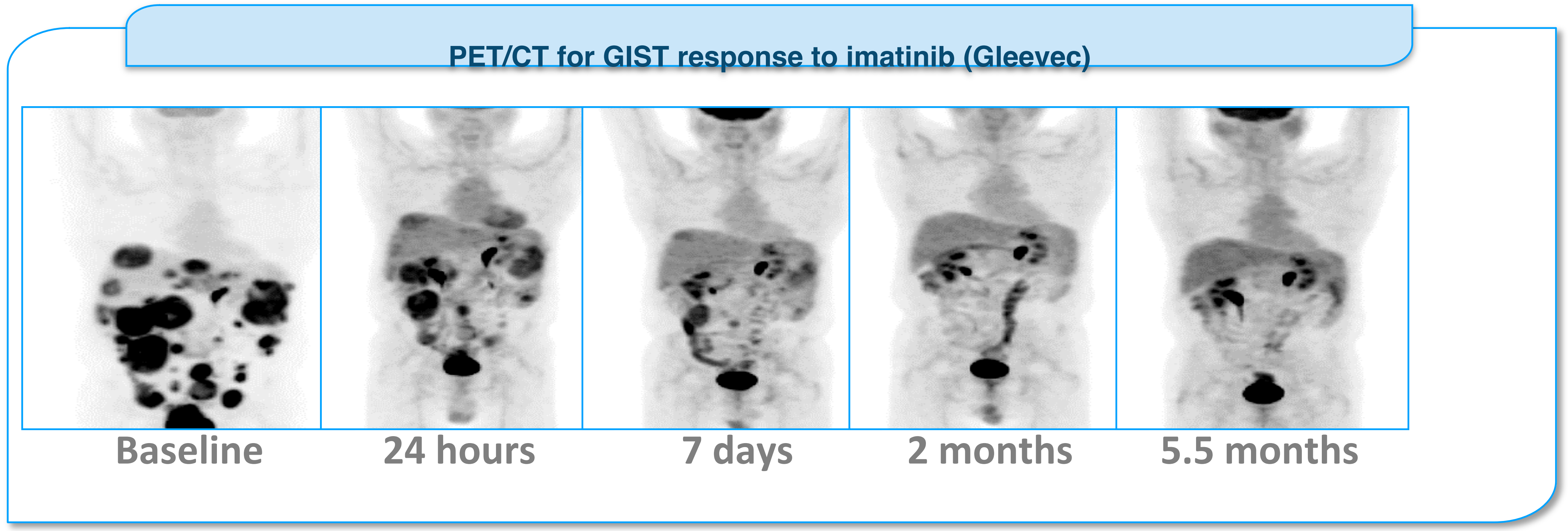




Two broad types of strategies to deriving an imaging biomarker

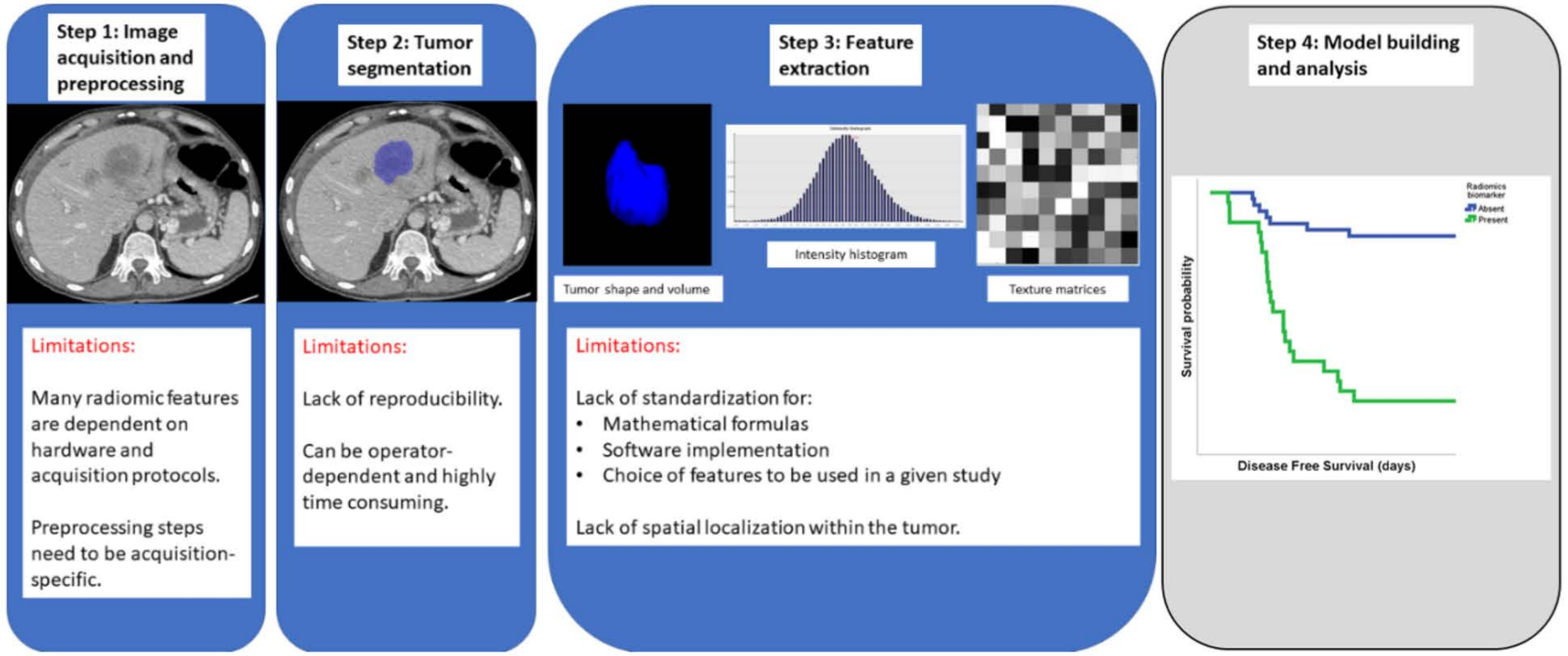
Defining the biomarker as a function of the physics of an underlying image acquisition designed to be specific to particular aspects of tumor biology

Metabolic Imaging

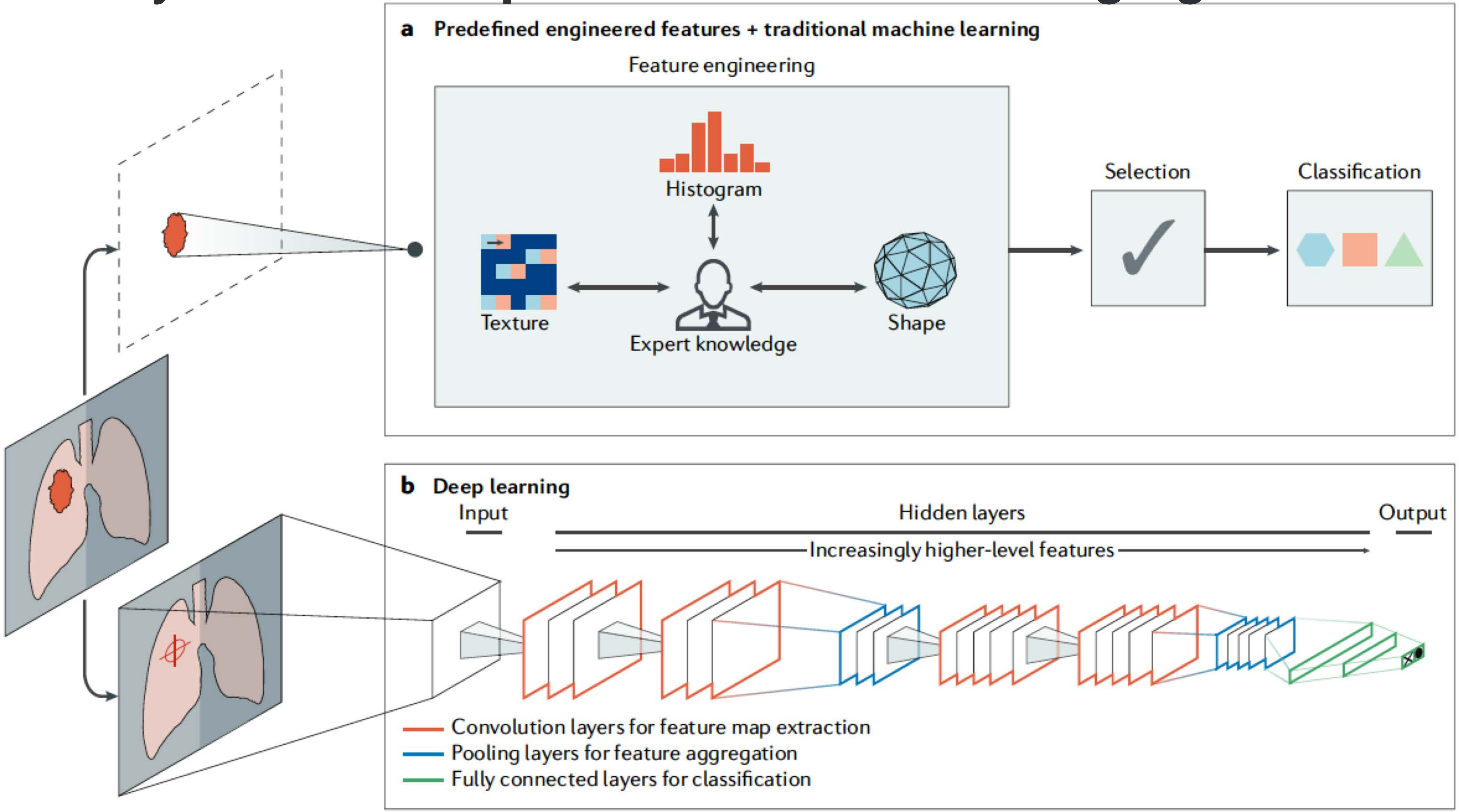


Using standard-of-care imaging techniques that are already in wide clinical use

Radiomics



2 pathways for tumor quantification with imaging



Artificial intelligence in radiology, Nature review- Cancer, 2018

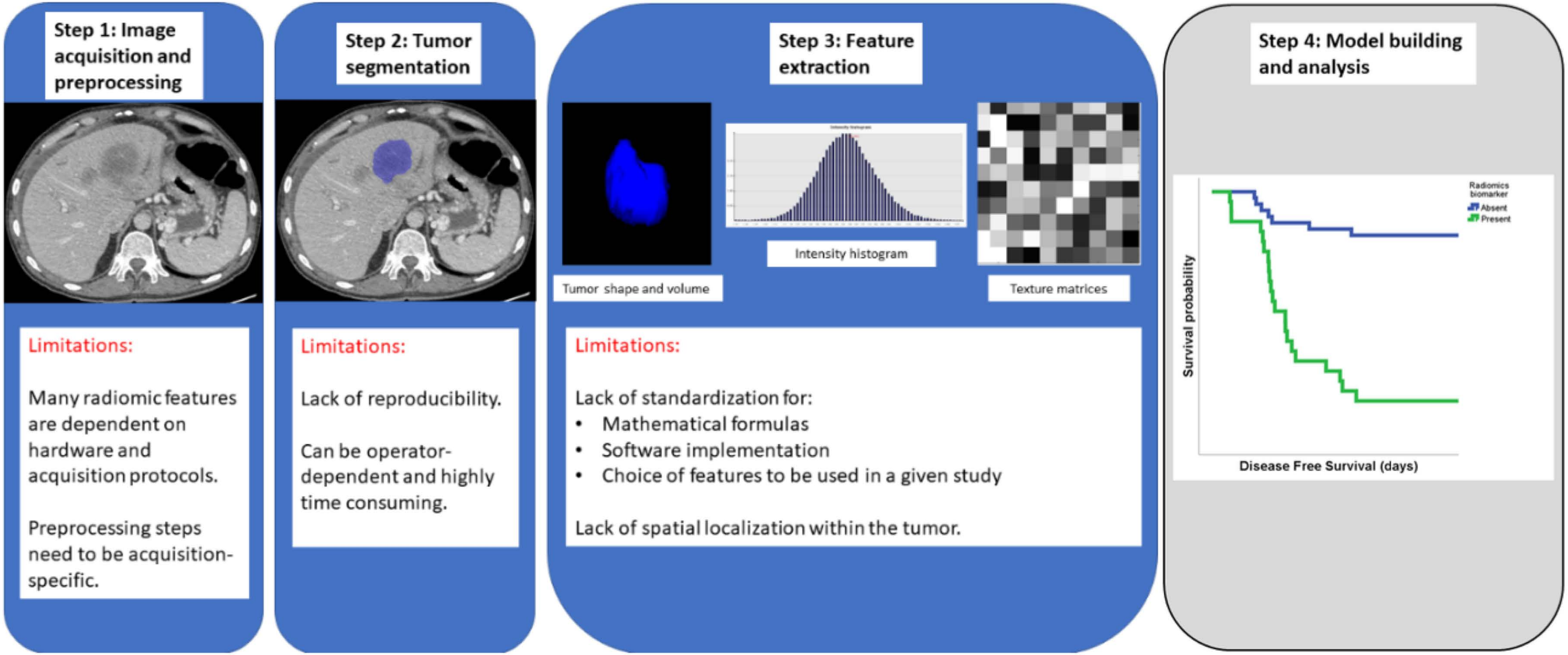


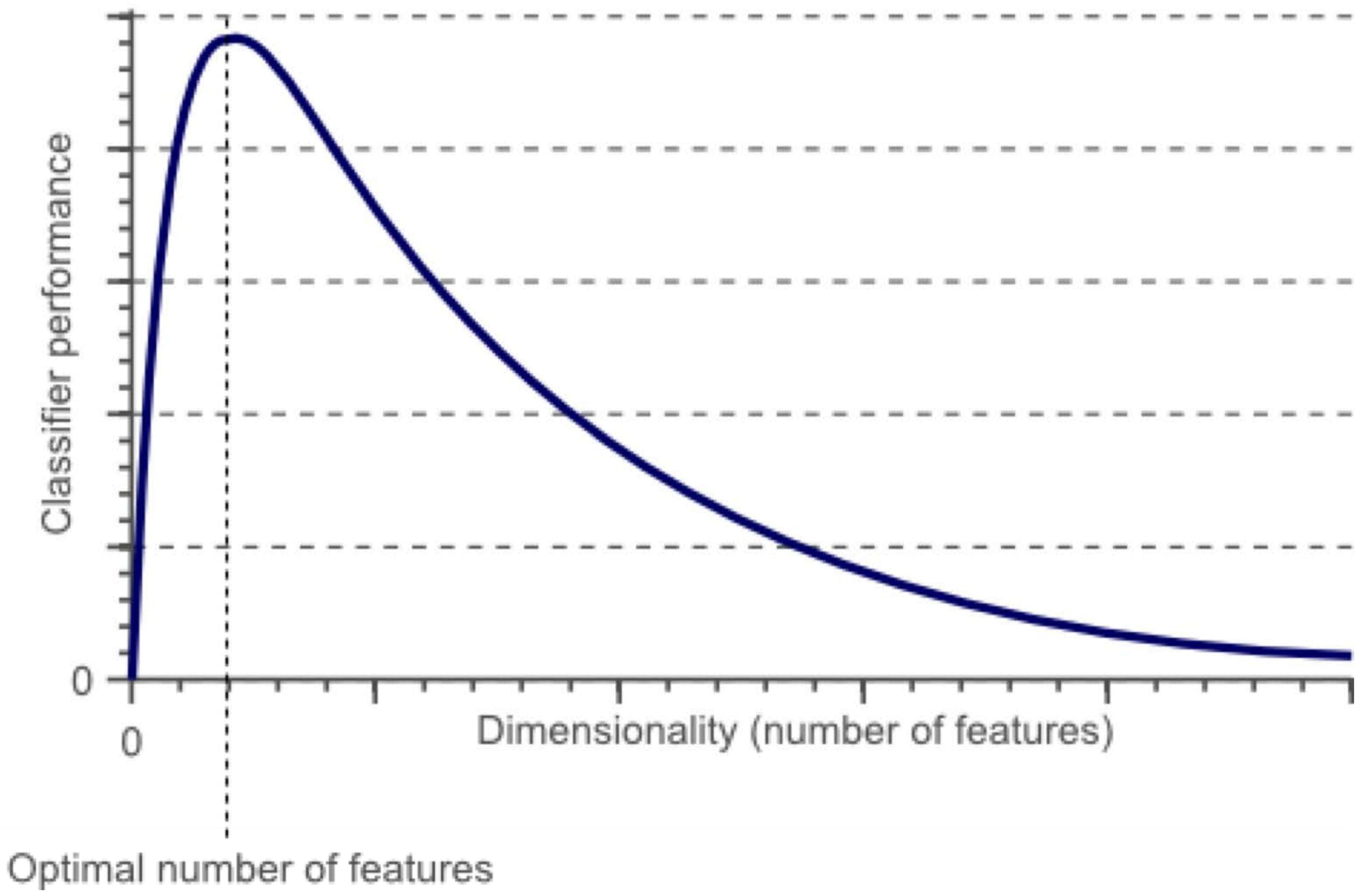
Image-based biomarkers for solid tumor quantification. Gallix et al. Eur radiol 2019

Typical features used in a radiomics workflow

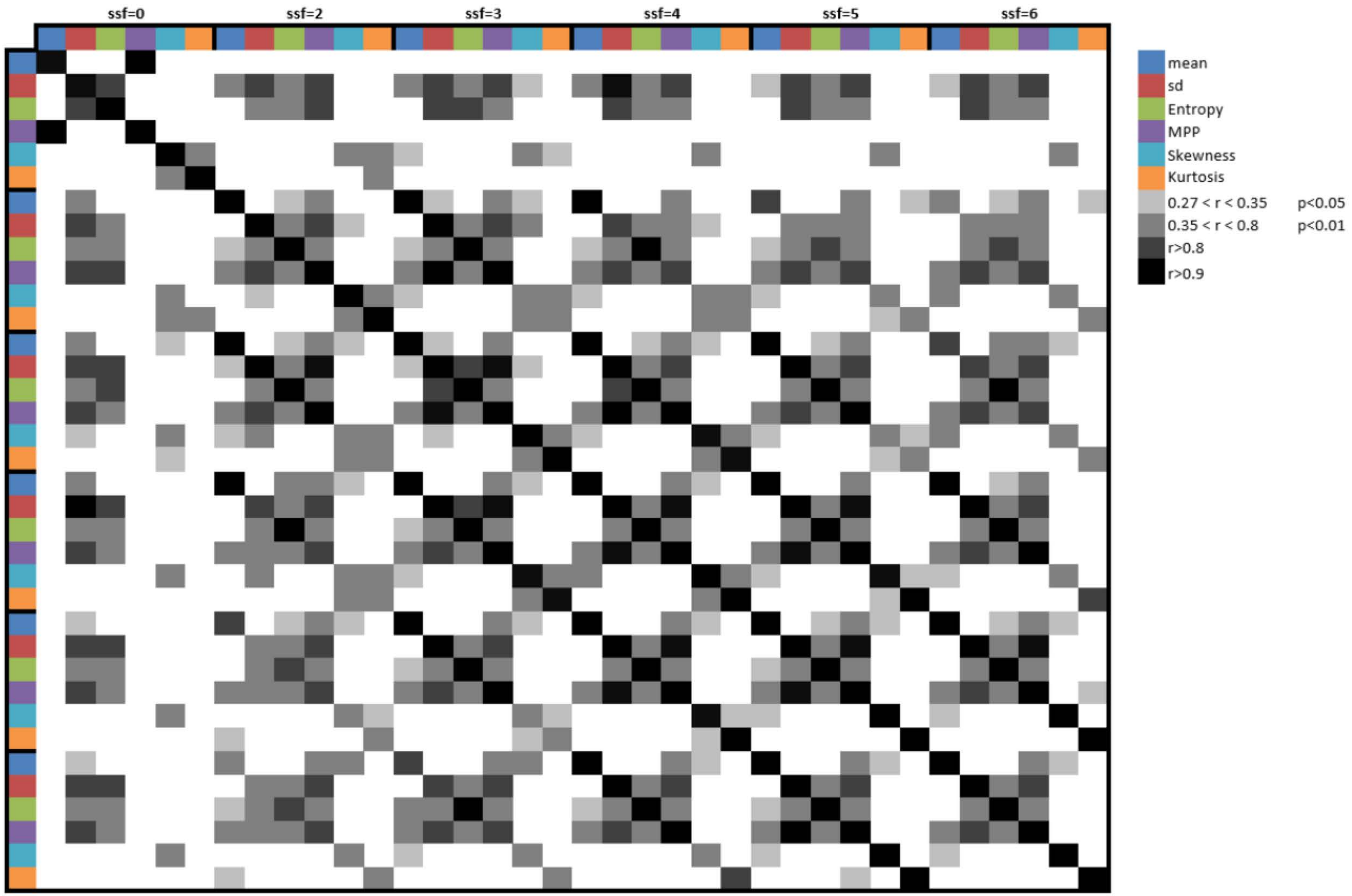
Category	Examples
Size	• Volume • Maximum 3D diameter • Surface area
Shape	• Sphericity • Elongation • Flatness
1st order statistical analysis – Intensity histogram features	• Mean • Variance • Kurtosis
2nd order statistical analysis – texture features	Features derived from texture matrices, such as: • Grey-level co-occurrence matrix • Grey-level run-length matrix • Grey-level size-zone matrix
Higher order statistical analysis	• Fractals • Wavelets

Curse of dimensionality

- The number of features generate is high
- As the numbers of features grows, the amount of data we need to generalize grows exponentially

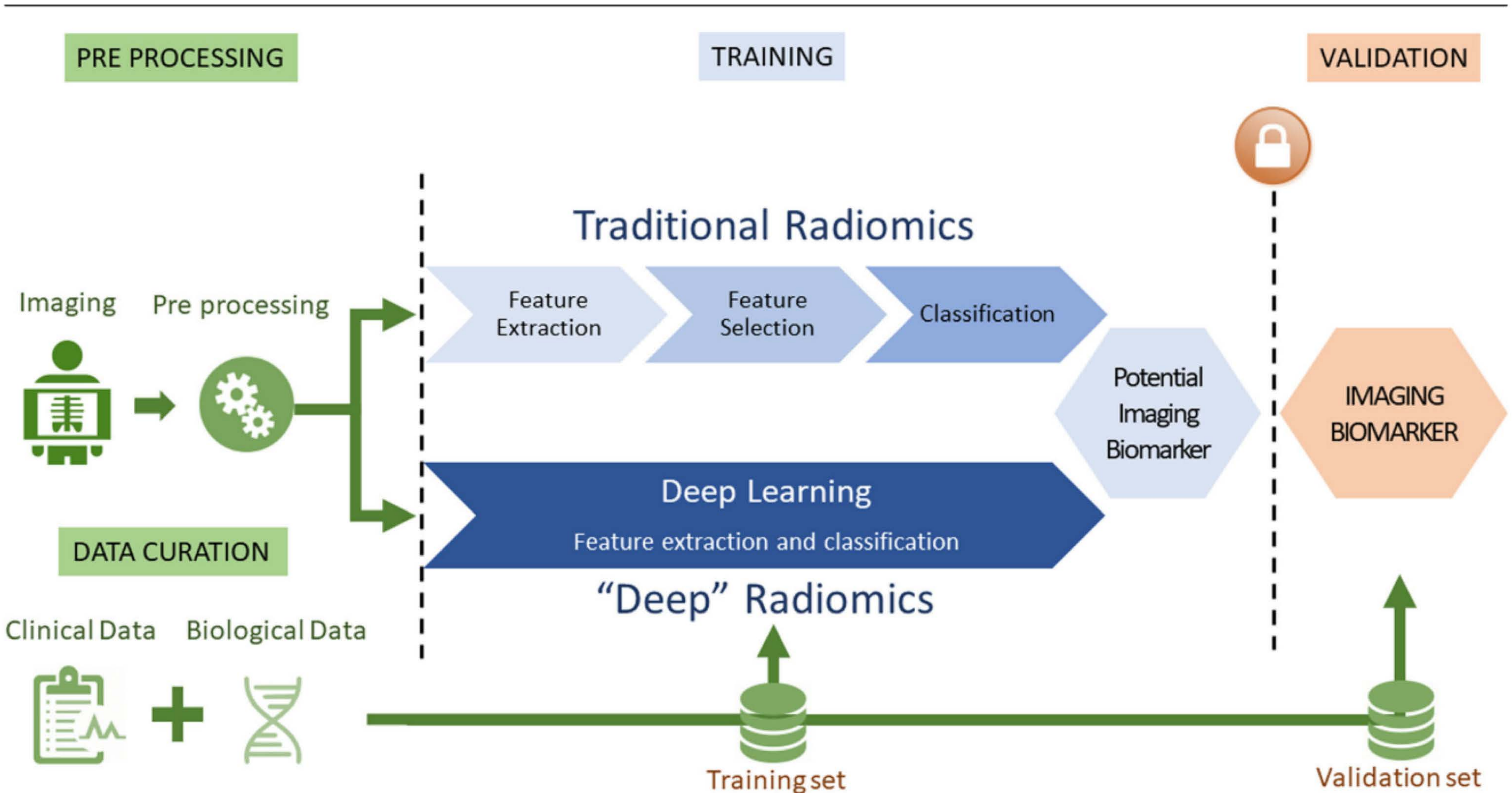


High level of correlation between features

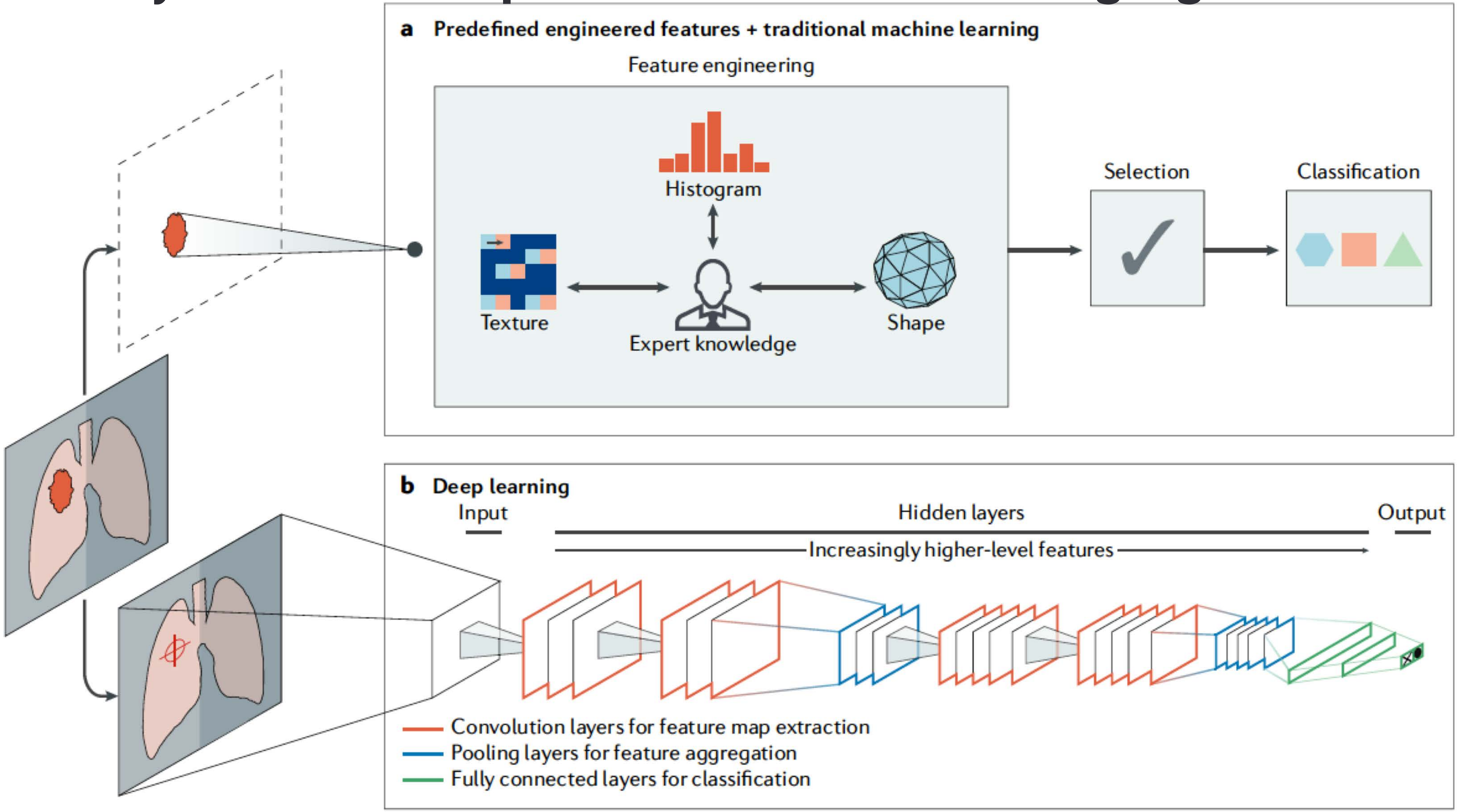


Dohan, Gallix, Unpublished data

2 pathways for tumor quantification with imaging

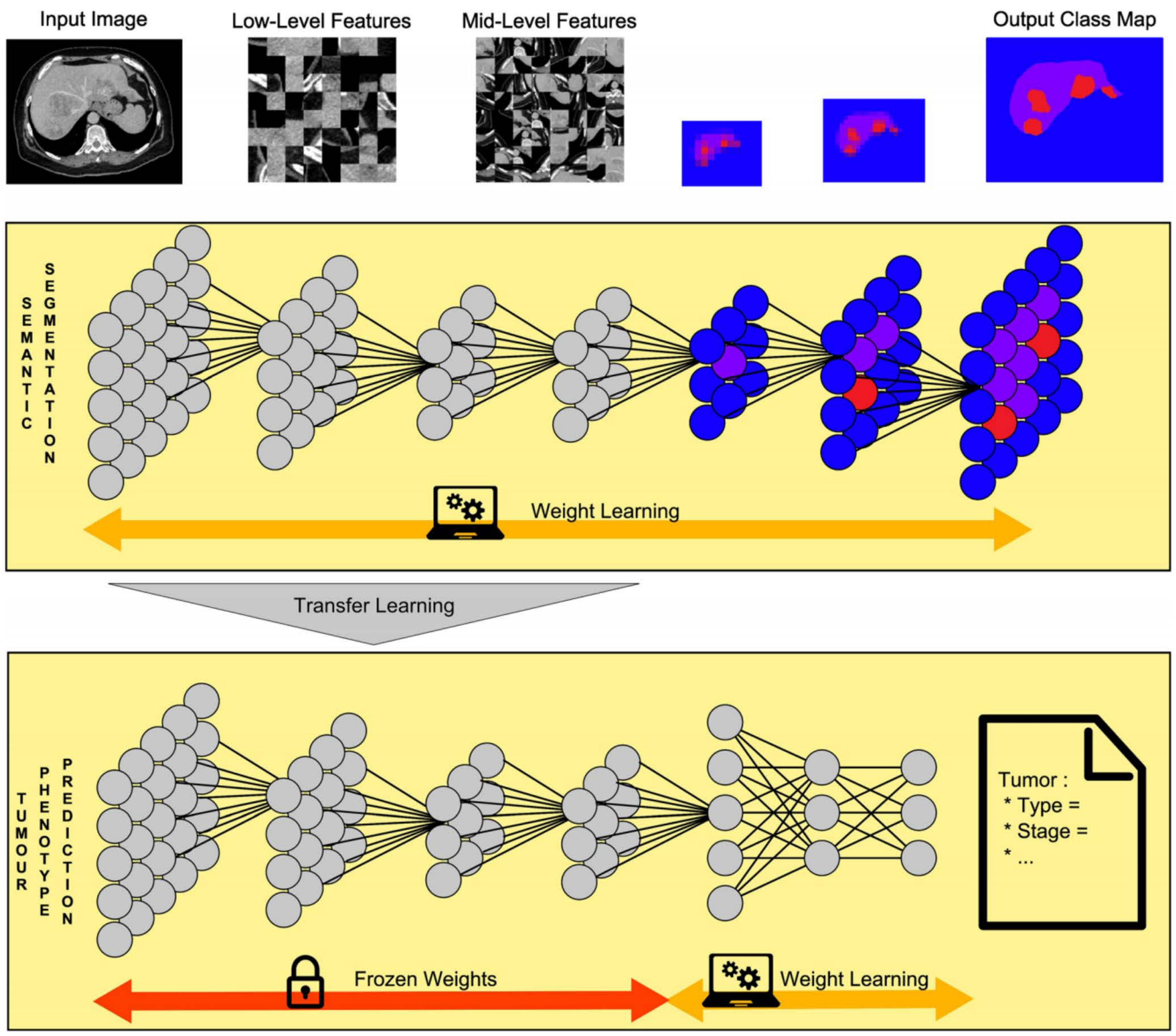


2 pathways for tumor quantification with imaging

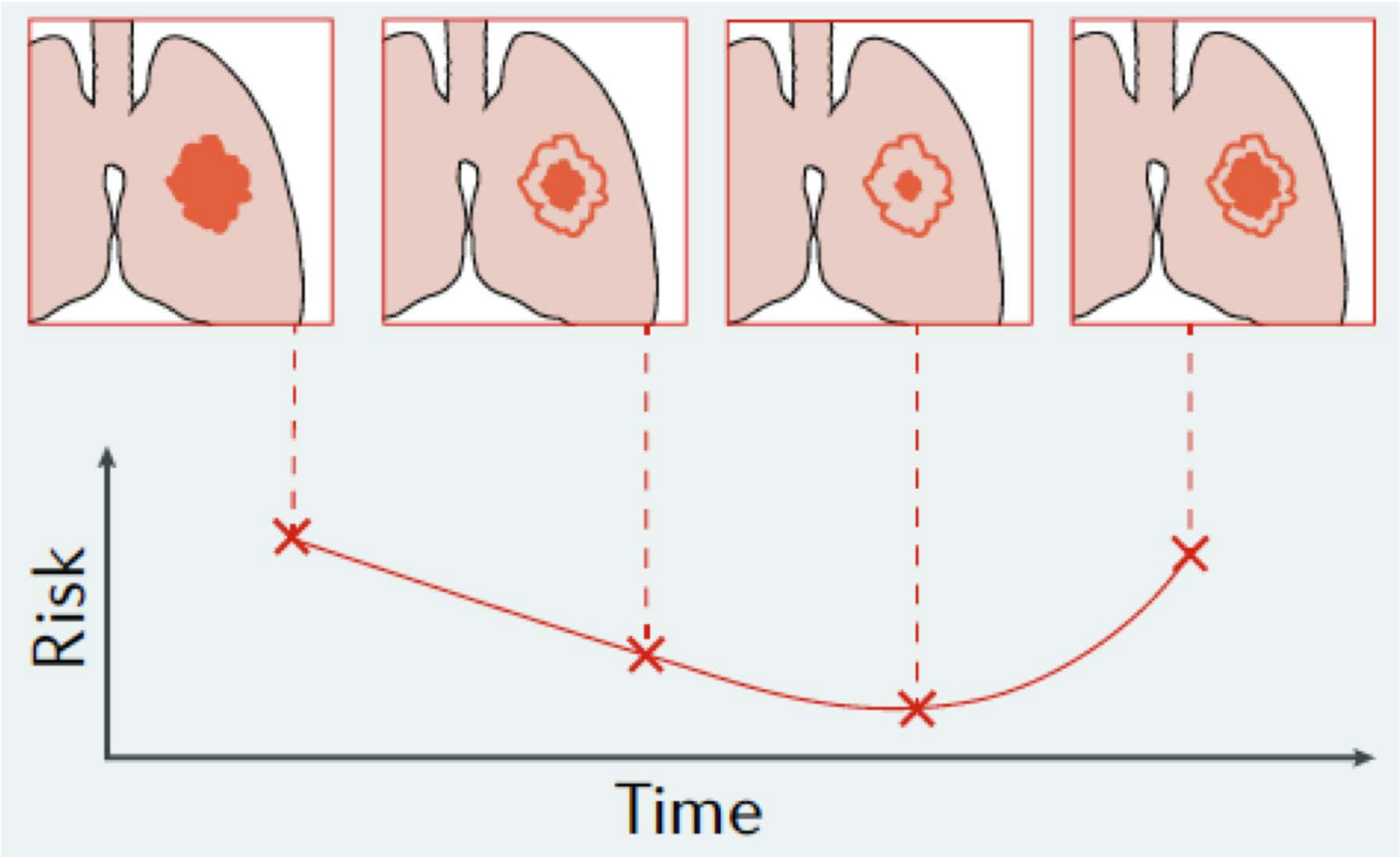


Artificial intelligence in radiology, Nature review- Cancer, 2018

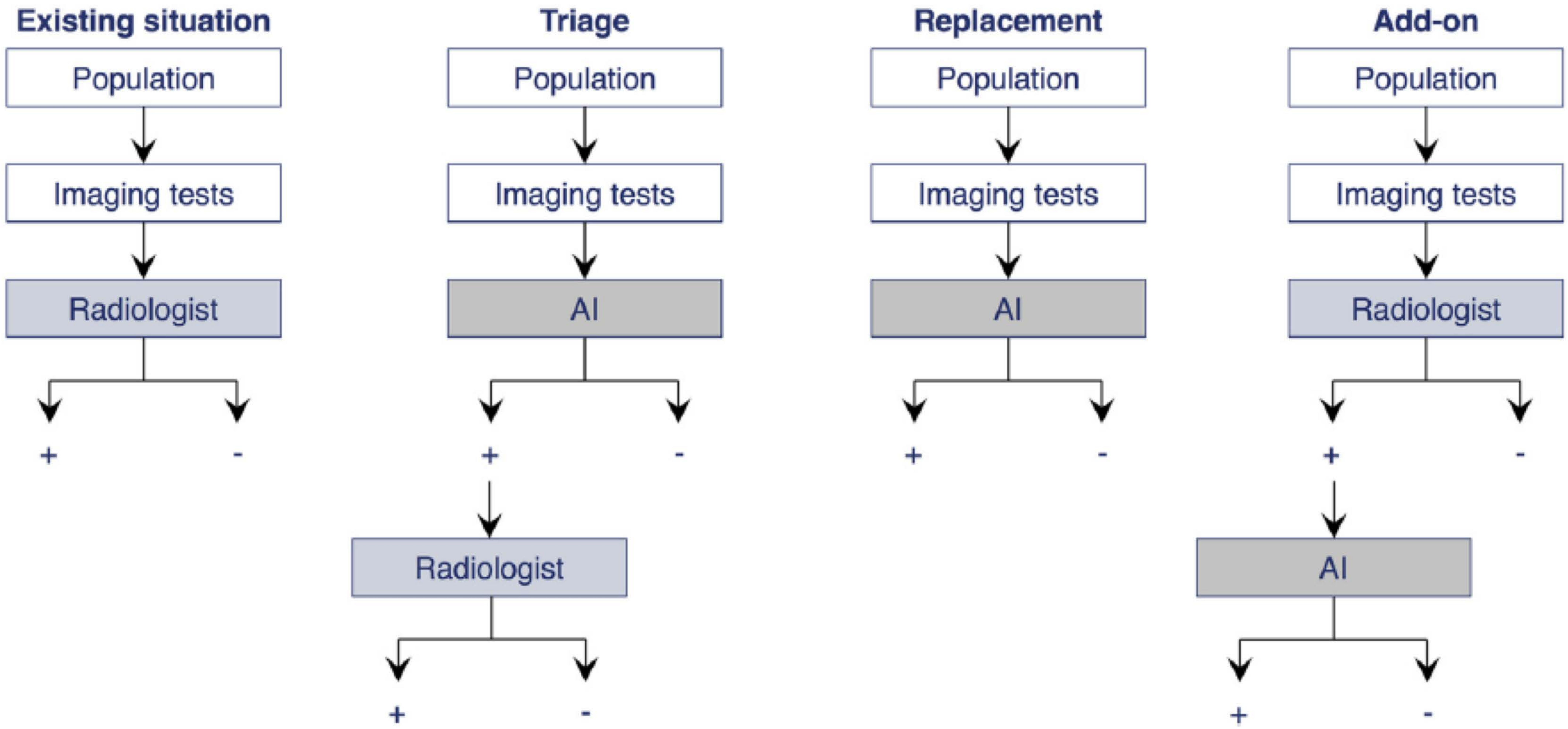
Tumor characterization using a two-step deep neural network analysis



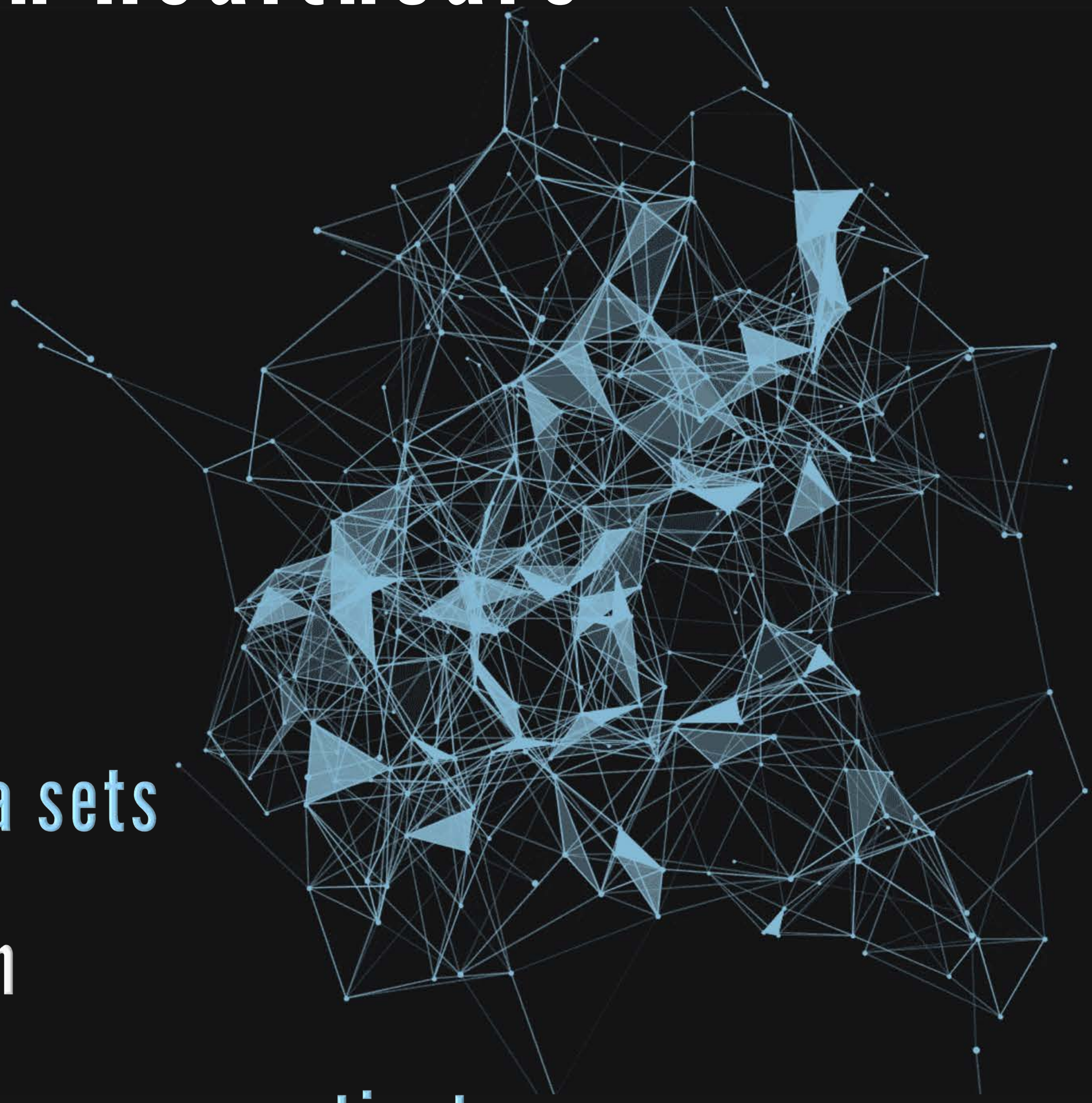
Monitoring the disease



Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology



Factors that will drive the adoption of AI in Healthcare



the strengths of digital imaging over human interpretation

the digitization of health-related records and data sharing

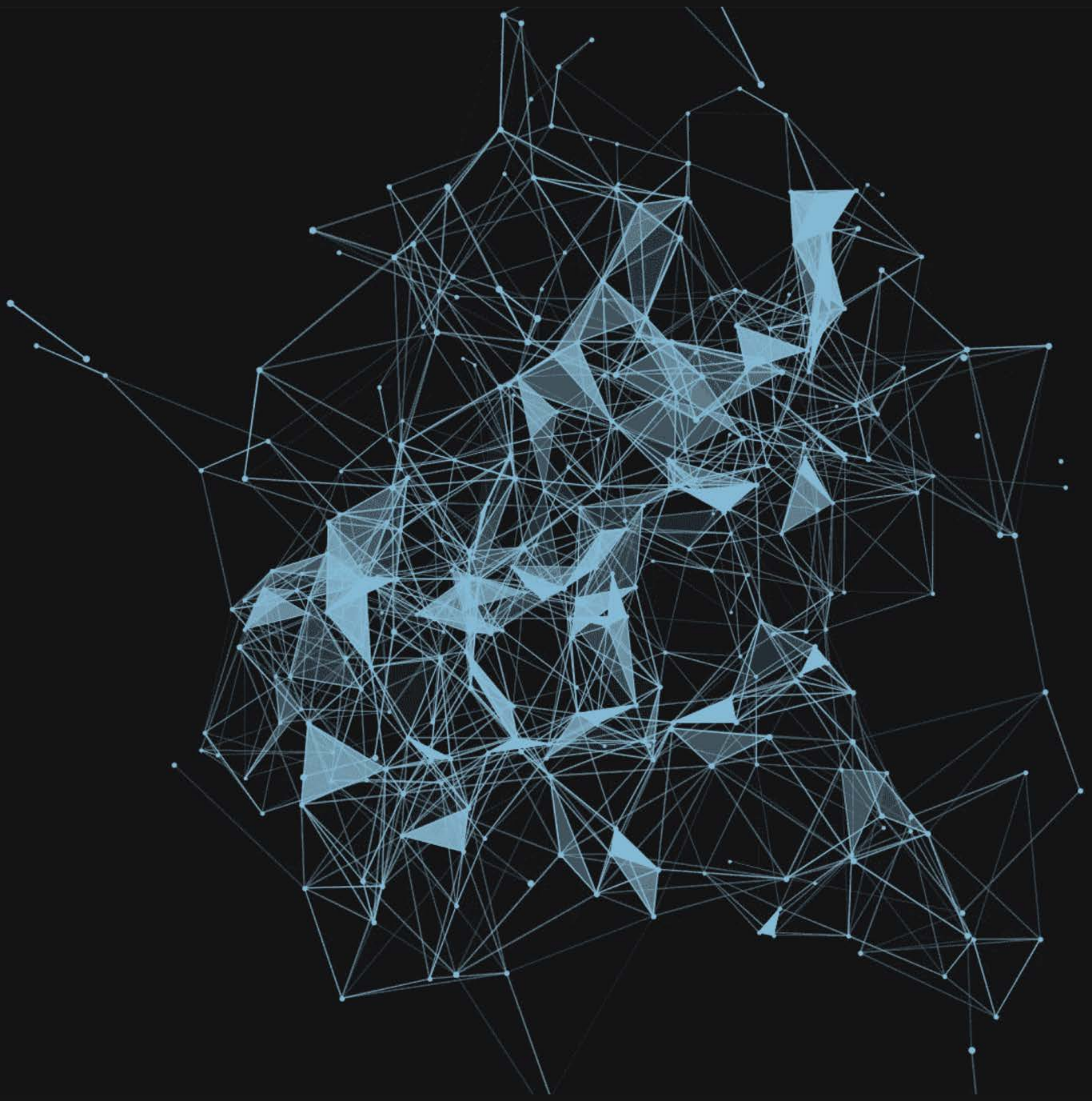
the adaptability of deep learning to analysis of heterogeneous data sets

the capacity of deep learning for hypothesis generation in research

the promise of deep learning to streamline clinical workflows and empower patients

the rapid-diffusion open-source and proprietary deep learning programs

Questions



How to validate AI tools ?

Who controls AI and is ultimately responsible for its actions ?

How generalize AI across different

- patient demographics?
- geographic regions
- healthcare system ?

How to insure privacy of the data and social acceptability ?

How to allow acceptance of AI applications in clinical practice ?

How access to quality and representative training data ?

How to maintain professional knowledge ?

Conclusion

AI and DL are entering the mainstream of clinical medicine.

AI can augment human intelligence to improve decision making and operational processes.

Physicians need to actively engage to adapt their practice and to shape the technology.

"The good physician treats the disease; the great physician treats the patient who has the disease."

William Osler

McGill

ArtIC

Artificial Intelligence for Care