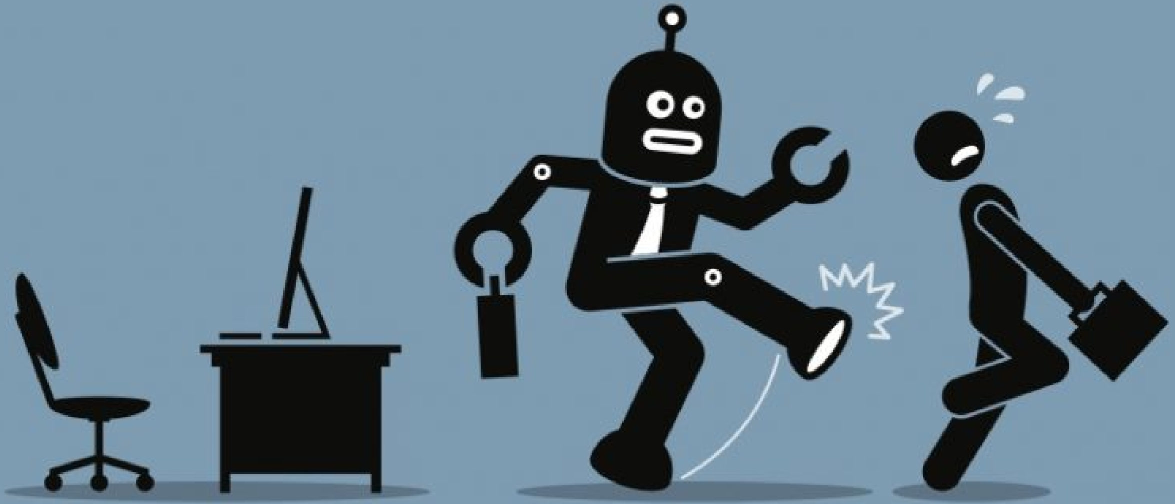
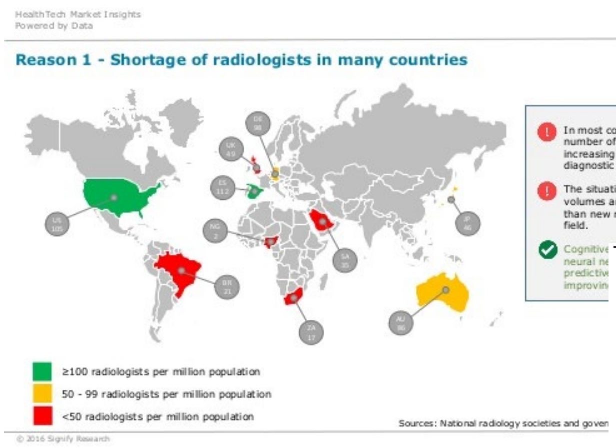
Artificial Intelligence in Prostate Cancer Diagnostics

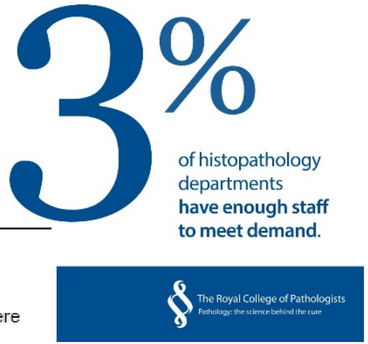
Dr. ir. Geert Litjens
Computational Pathology Group
Department of Pathology

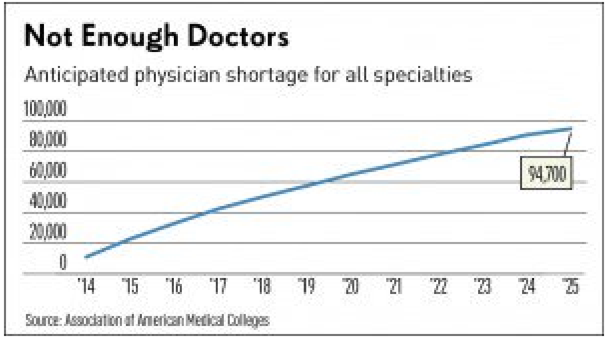
Radboudumc

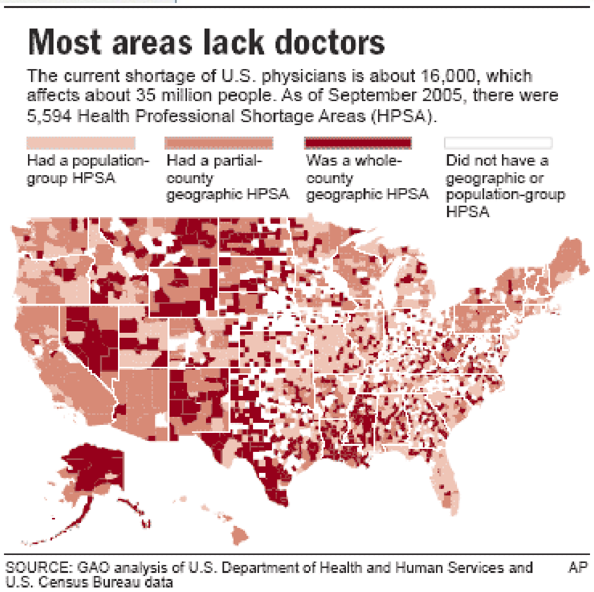
Help, the robots are coming!

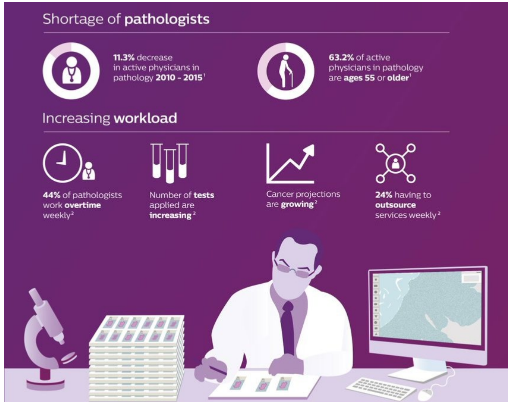



HealthTech Market Insights
Powered by Data
Signify
Reason 1 - Shortage of radiologists in many countries
In most countries there is an insufficient number of radiologists to meet the ever-increasing demand for imaging and diagnostic services
The situation will get worse, as imaging volumes are increasing at a faster rate than new radiologists are entering the field.
≥100 radiologists per million population
50 - 99 radiologists per million population
<50 radiologists per million population
Sources: National radiology societies and gover
© 2016 Signify Research


3%
of histopathology departments
have enough staff to meet demand.
The Royal College of Pathologists
Pathology: the science behind the cure


Not Enough Doctors
Anticipated physician shortage for all specialties
100,000
80,000
60,000
40,000
20,000
0
94,700
'14
'15
'16
'17
'18
'19
'20
'21
'22
'23
'24
'25
Source: Association of American Medical Colleges


Most areas lack doctors
The current shortage of U.S. physicians is about 16,000, which affects about 35 million people. As of September 2005, there were 5,594 Health Professional Shortage Areas (HPSA).
Had a population-group HPSA
Had a partial-county geographic HPSA
Was a whole-county geographic HPSA
Did not have a geographic or population-group HPSA
SOURCE: GAO analysis of U.S. Department of Health and Human Services and U.S. Census Bureau data
AP

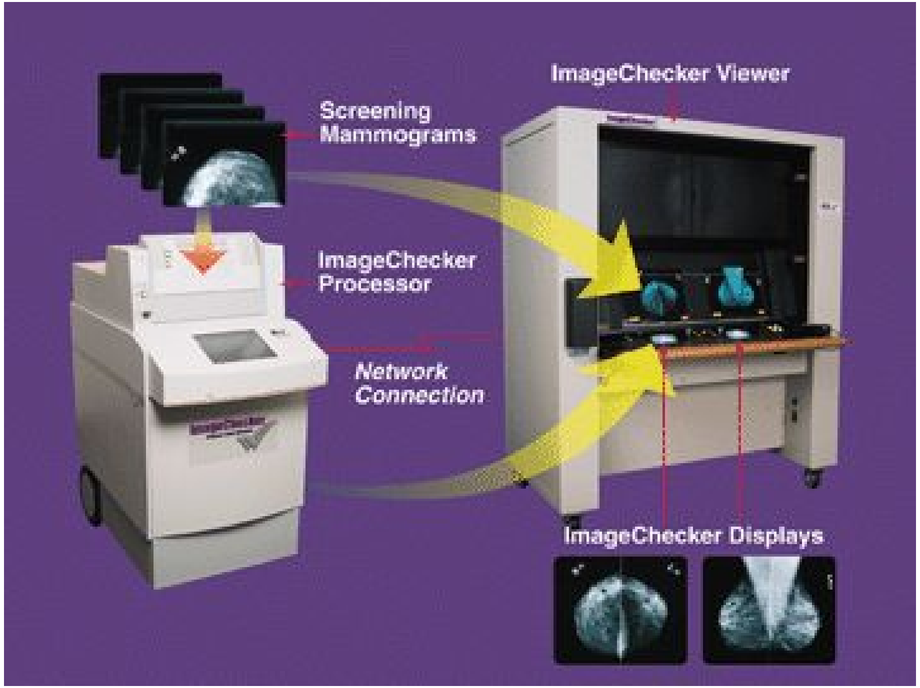

Shortage of pathologists
11.3% decrease in active physicians in pathology 2010 - 2015
63.2% of active physicians in pathology are ages 55 or older
Increasing workload
44% of pathologists work overtime weekly
Number of tests applied are increasing
Cancer projections are growing
24% having to outsource services weekly

Het Parool
Vrij, Onverveerd

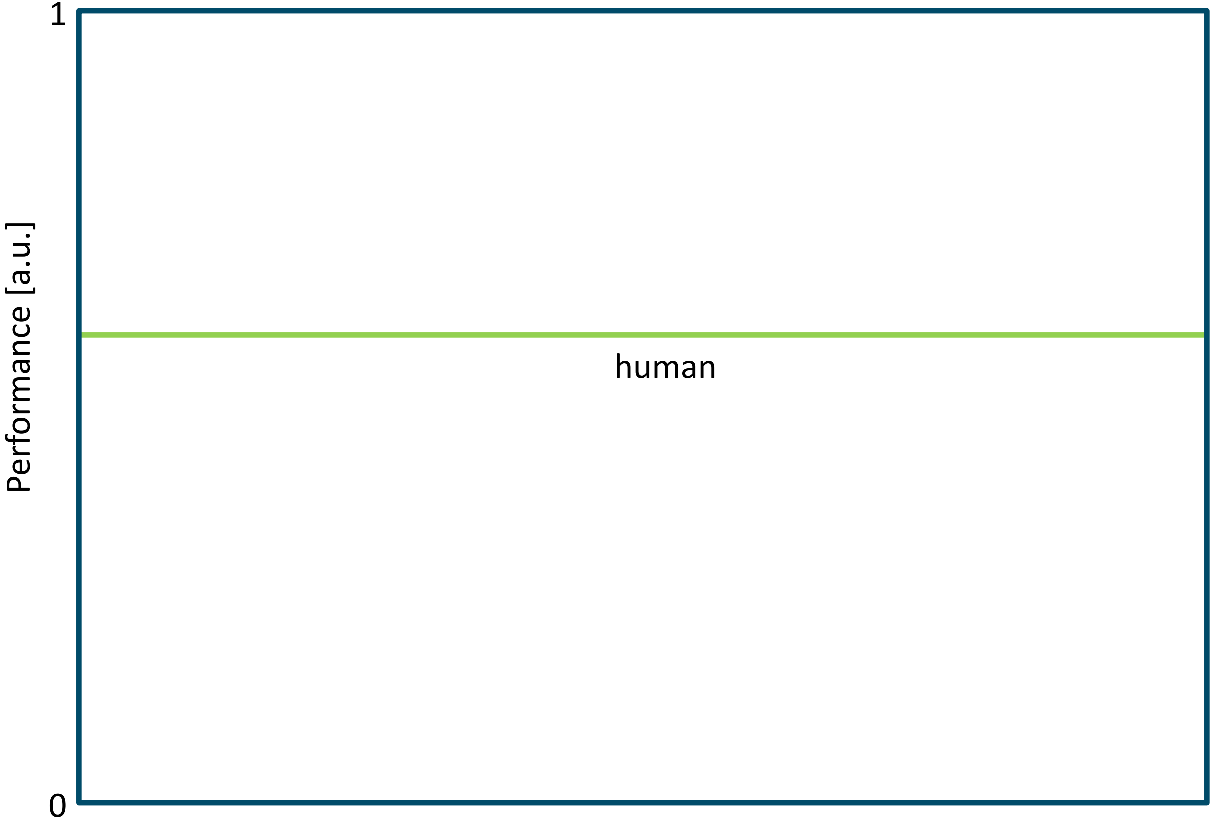
Hurray, the robots are coming!

De menselijke robot Asimo is al behoorlijk slim en handig - en hij kan ook nog eens negen kilometer per uur rennen. © REUTERS

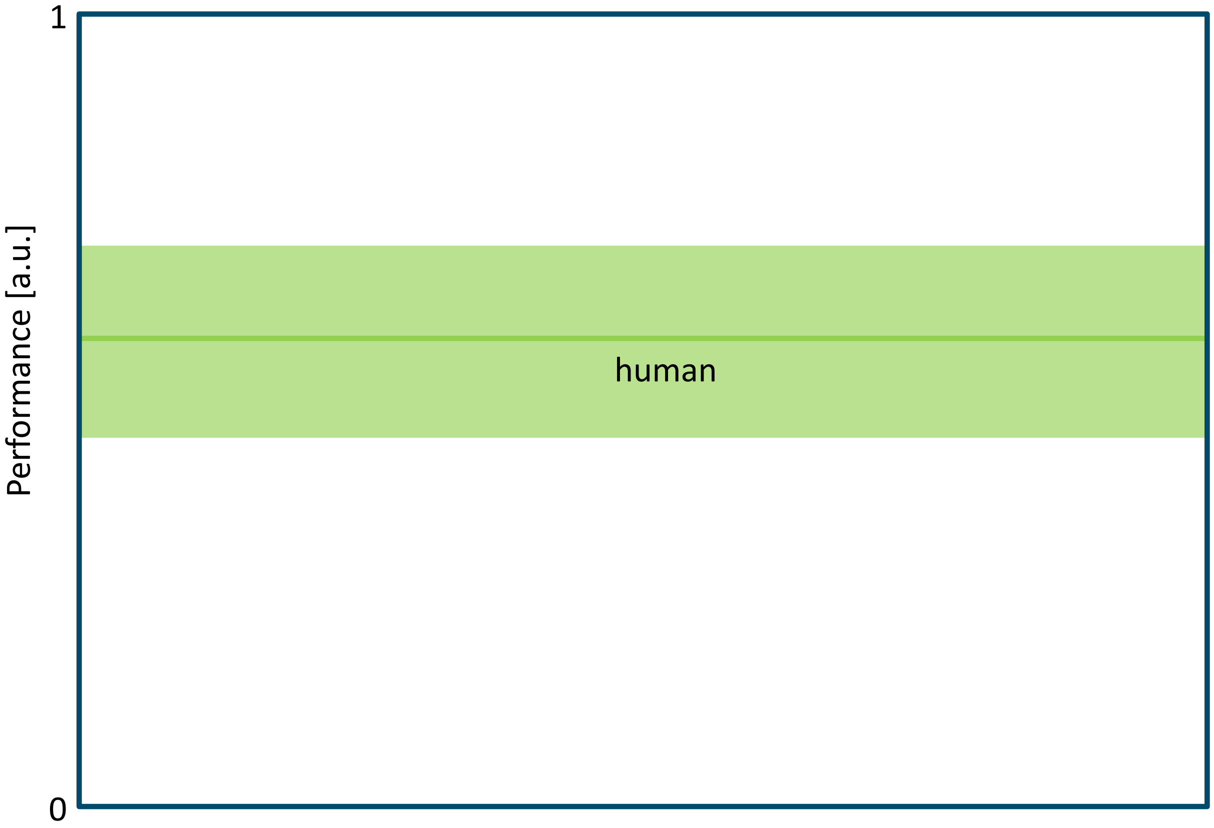
Computer-aided diagnosis

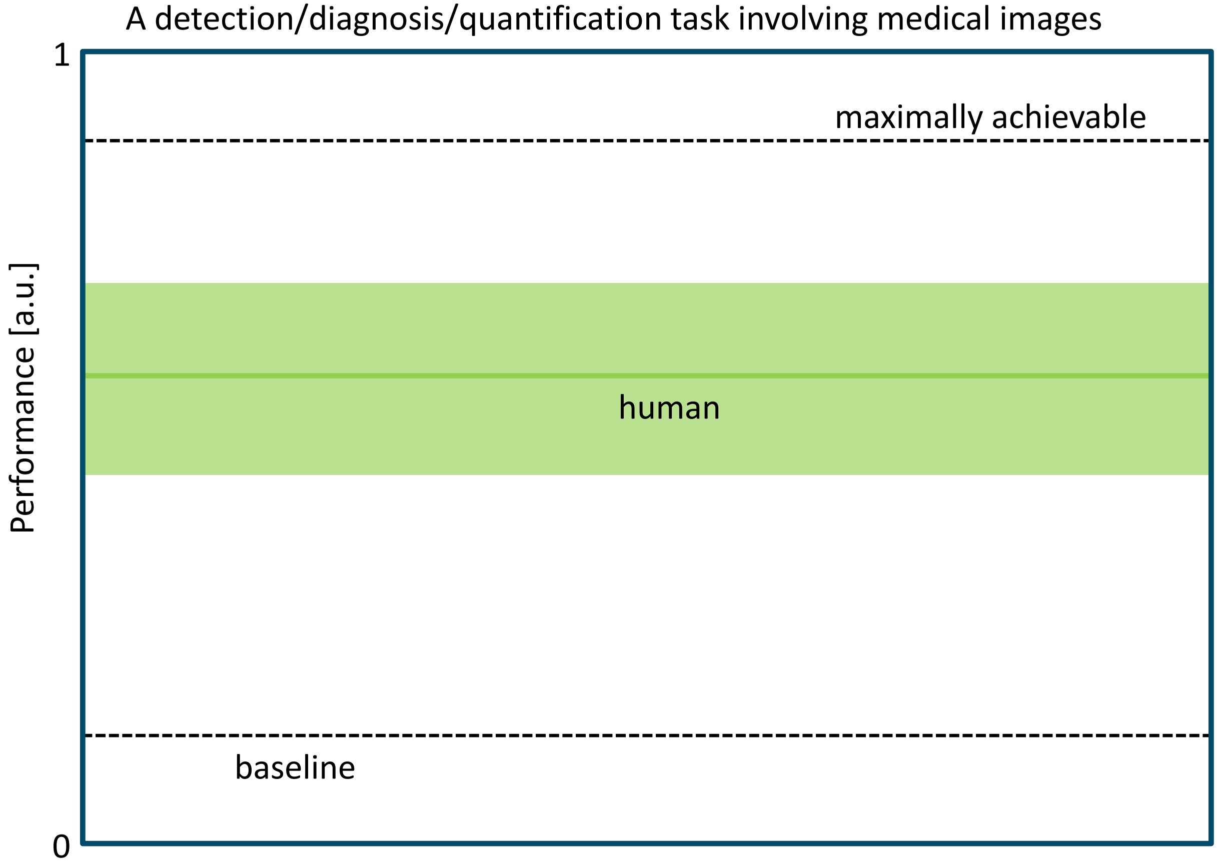


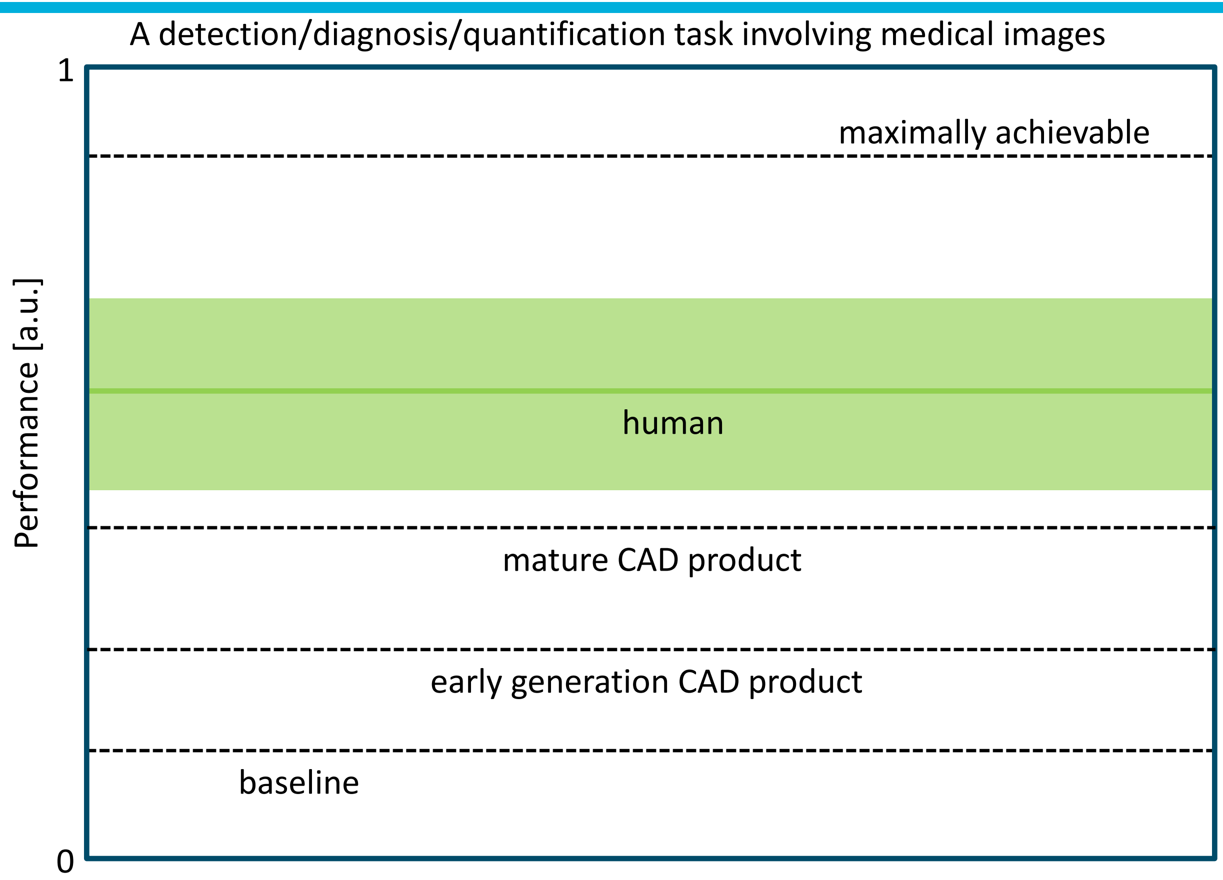
A detection/diagnosis/quantification task involving medical images



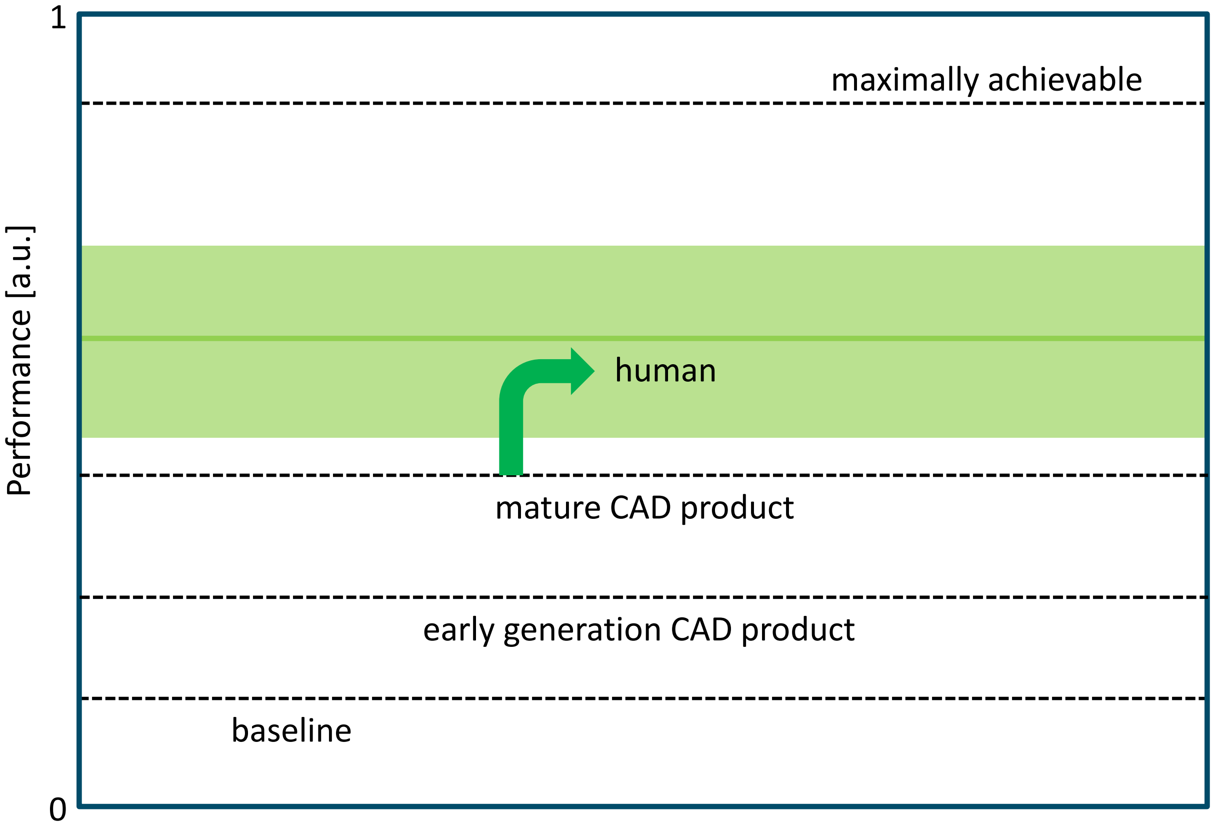
A detection/diagnosis/quantification task involving medical images



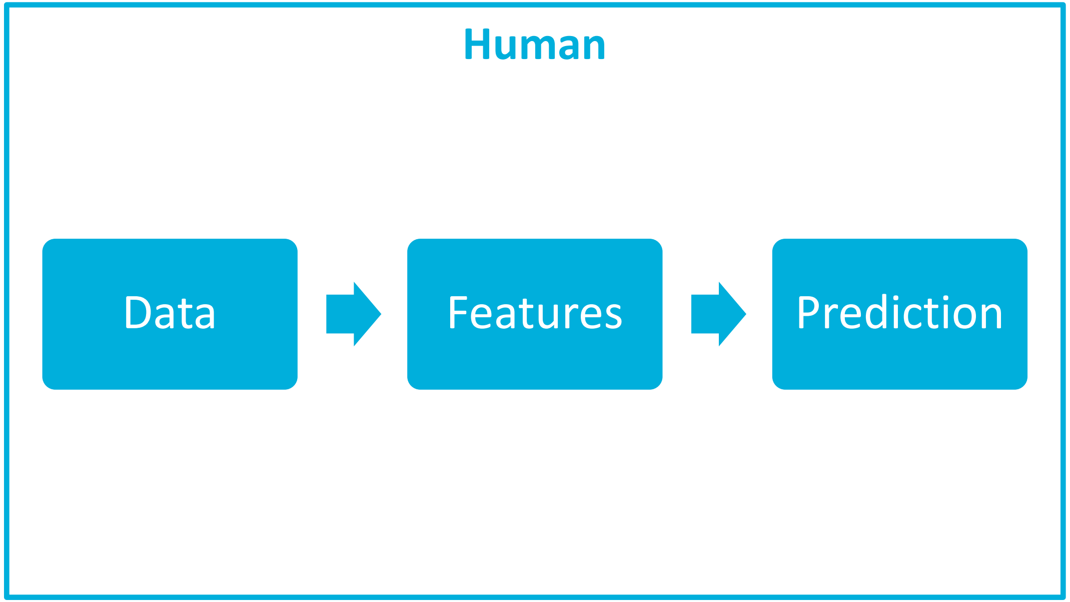

A detection/diagnosis/quantification task involving medical images
1
maximally achievable
Performance [a.u.]
human
baseline
0


A detection/diagnosis/quantification task involving medical images
1
maximally achievable
Performance [a.u.]
human
mature CAD product
early generation CAD product
baseline
0

A detection/diagnosis/quantification task involving medical images



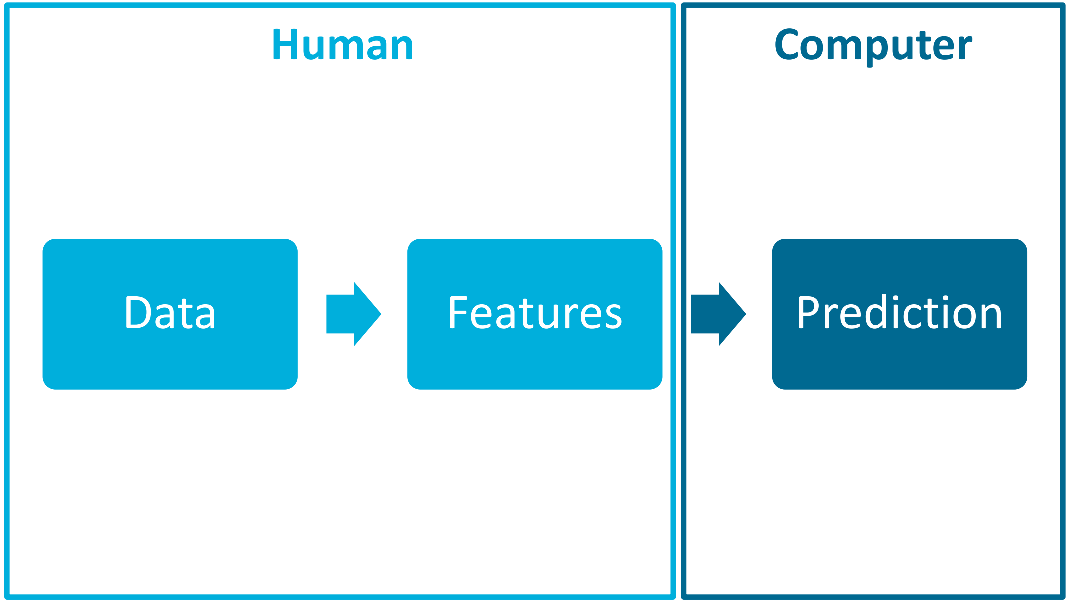
Timeline of computer-aided diagnosis



Prostate cancer risk models

Risk	Stage	Prostate-specific antigen (PSA)	Gleason score
Low	T1c – T2a	< 10 ng/mL	< 7
Medium	T2b-c	10 – 20 ng/mL	7
High	T3	>20 ng/mL	>7

Timeline of computer-aided diagnosis



Lung cancer risk models

Calculator: Solitary pulmonary nodule malignancy risk in adults (Brock University cancer prediction equation)

Logodds = (0.0287 * (Age - 62)) + Sex + FamilyHistoryLungCa + Emphysema - (5.3854 * ((Nodulesize/10)$^{-0.5}$ - 1.58113883)) + Noduletype + NoduleUpperLung - (0.0824 * (Nodulecount - 4)) + Spiculation - 6.7892

Cancerprobability = 100 * (e$^{(Logodds)}$ / (1 + e$^{(Logodds)}$))

Input:

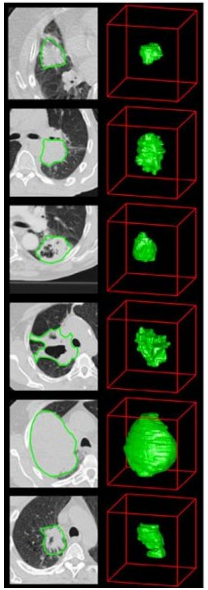
Age		years
Sex	Female (0.6011)	
	Male (0)	
Family history of lung cancer	(0.2961)	
Emphysema	(0.2953)	
Nodule size		mm
Nodule type	Nonsolid or ground-glass (-0.1276)	
	Partially solid (0.377)	
	Solid (0)	
Nodule in upper lung	(0.6581)	
Nodule count		#
Spiculation	(0.7729)	

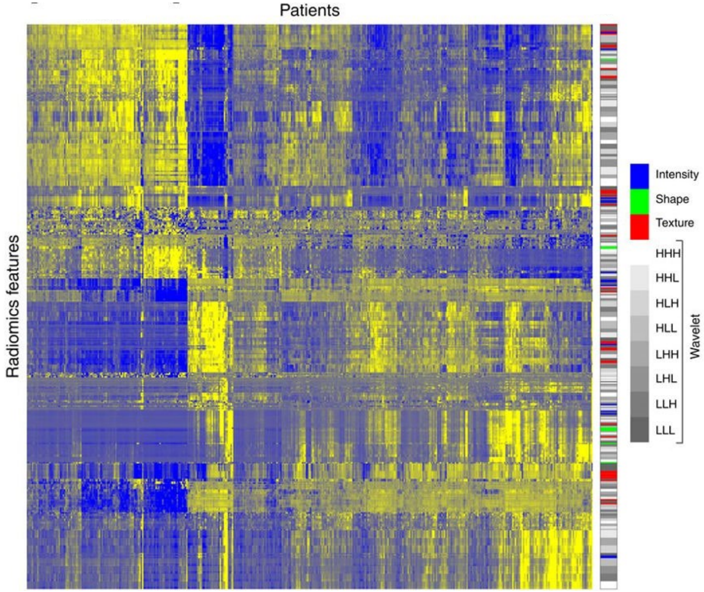
Results:

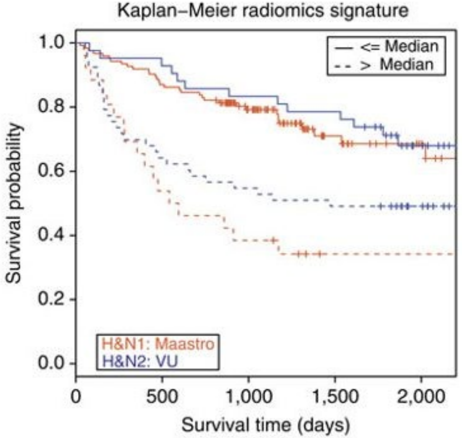
Log odds		
Cancer probability		%
Decimal precision	2	

Reset form

Radiomics

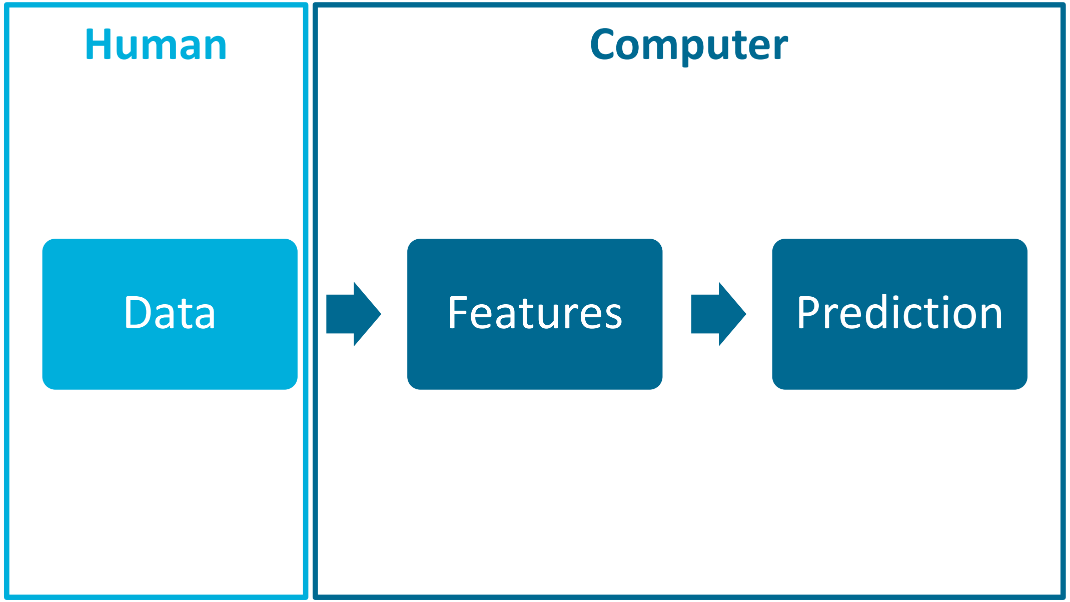






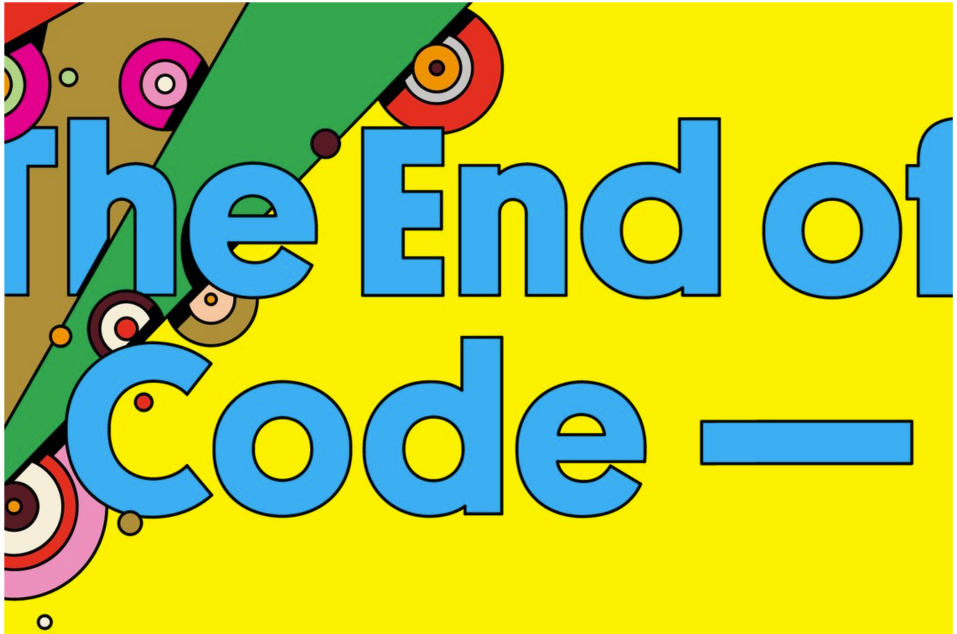
Aerts et al. *Nature Communications*, 2014

Timeline of computer-aided diagnosis



JASON TANZ IDEAS 05.17.16 06:50 AM

SOON WE WON'T PROGRAM COMPUTERS. WE'LL TRAIN THEM LIKE DOGS



EDWARD C. MONAGHAN

BEFORE THE INVENTION of the computer, most experimental psychologists thought the brain was an unknowable black box. You could analyze a subject's behavior—*ring bell, dog salivates*—but thoughts, memories, emotions? That stuff was obscure and inscrutable, beyond the reach of science. So these behaviorists, as they called themselves, confined their work to the study of stimulus and response, feedback and reinforcement, bells and saliva. They gave up trying to

MOST POPULAR

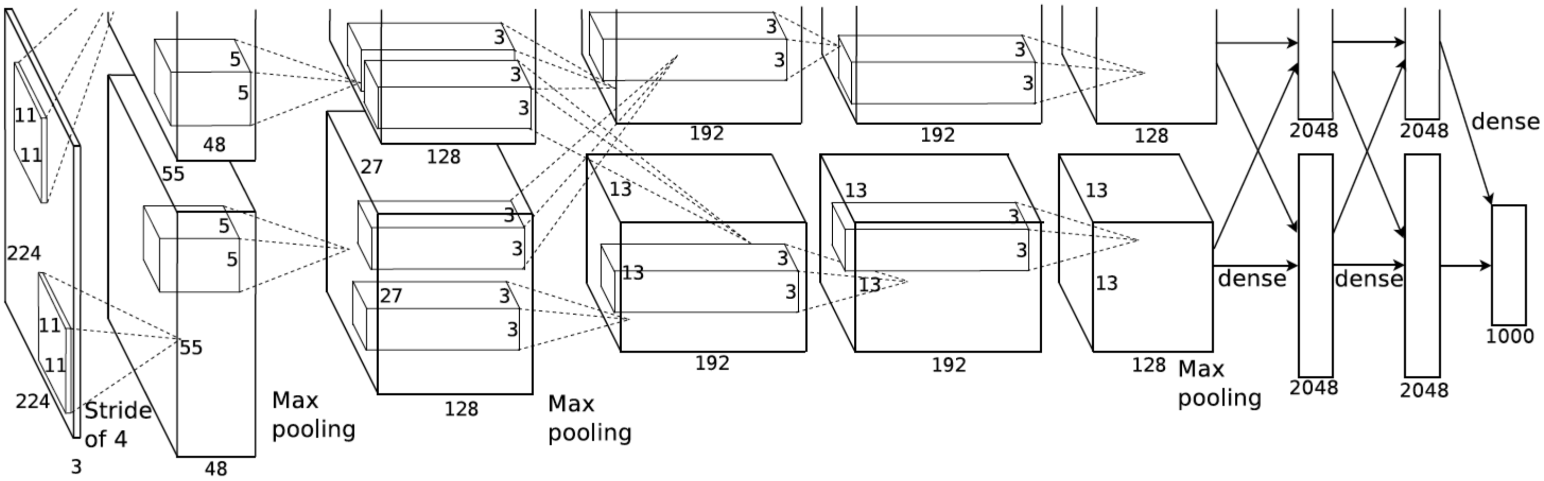
BUSINESS
SpaceX's President is Thinking Even Bigger Than Elon Musk
ERIN GRIFFITH

TRANSPORTATION

JASON TANZ IDEAS 05.17.16 06:50 AM

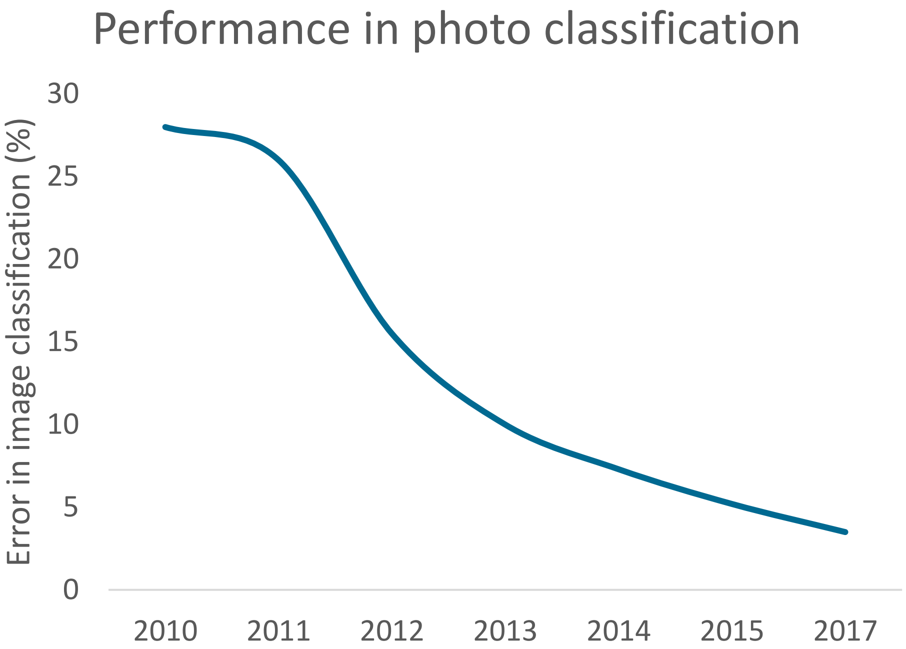
SOON WE WON'T PROGRAM COMPUTERS. WE'LL TRAIN THEM LIKE DOGS

The breakthrough



Krizhevsky et al. *NeurIPS*, 2012

The breakthrough



The breakthrough

Alex Krizhevsky

Unknown affiliation
Verified email at cs.toronto.edu
Machine Learning

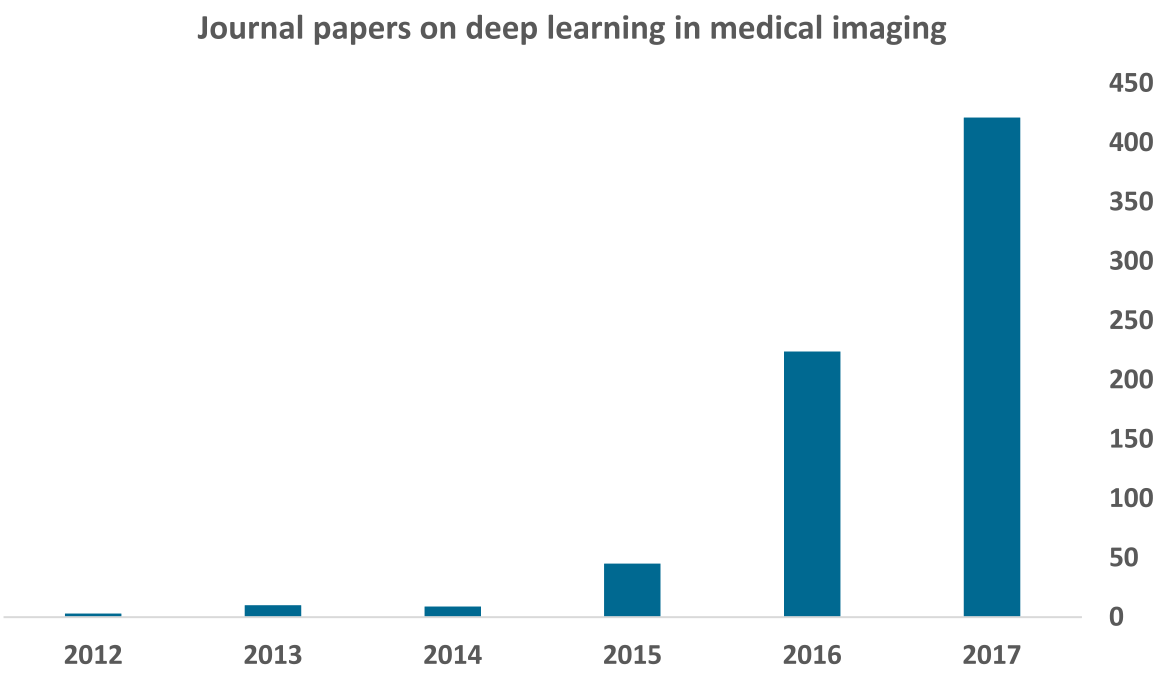
FOLLOW

TITLE	CITED BY	YEAR
Imagenet classification with deep convolutional neural networks A Krizhevsky, I Sutskever, GE Hinton Advances in neural information processing systems, 1097-1105	28245	2012
N Srivastava, G Hinton, A Krizhevsky, I Sutskever, R Salakhutdinov The Journal of Machine Learning Research 15 (1), 1929-1958		
Improving neural networks by preventing co-adaptation of feature detectors GE Hinton, N Srivastava, A Krizhevsky, I Sutskever, RR Salakhutdinov arXiv preprint arXiv:1207.0580	3020	2012
Learning multiple layers of features from tiny images A Krizhevsky, G Hinton Technical report, University of Toronto 1 (4), 7	2611	2009
Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection S Levine, P Pastor, A Krizhevsky, J Ibarz, D Quillen The International Journal of Robotics Research 37 (4-5), 421-436	352	2018


INNOVATIONS IN
The microbiome
nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
LEARNING CURVE
Self-taught AI software attains human-level performance in video games
PAGES 486 & 529
EPIDEMIOLOGY
SHARE DATA IN OUTBREAKS
Forge open access to sequences and more
PAGE 477
COSMOLOGY
A GIANT IN THE EARLY UNIVERSE
A supermassive black hole at a redshift of 6.3
PAGES 490 & 512
QUANTUM PHYSICS
TELEPORTATION FOR TWO
Transferring two properties of a single photon
PAGES 491 & 516
NATURE.COM/NATURE
26 February 2015 £10
Vol. 518, No. 7540


nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card 40 years on
PAGE 462
NATUREASIA.COM
28 January 2016
Vol. 529, No. 7587

Queue the hype…



JAMA | Original Investigation

Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer

Babak Ehteshami Bejnordi, MS; Mitko Veta, PhD; Paul Johannes van Diest, MD, PhD; Bram van Ginneken, PhD; Nico Karssemeijer, PhD; Geert Litjens, PhD; Jeroen A. W. M. van der Laak, PhD; and the CAMELYON16 Consortium

IMPORTANCE Application of deep learning algorithms to whole-slide pathology images can potentially improve diagnostic accuracy and efficiency.

OBJECTIVE Assess the performance of automated deep learning algorithms at detecting metastases in hematoxylin and eosin–stained tissue sections of lymph nodes of women with breast cancer and compare it with pathologists' diagnoses in a diagnostic setting.

DESIGN, SETTING, AND PARTICIPANTS Researcher challenge competition (CAMELYON16) to develop automated solutions for detecting lymph node metastases (November 2015-November 2016). A training data set of whole-slide images from 2 centers in the Netherlands with (n = 110) and without (n = 160) nodal metastases verified by immunohistochemical staining were provided to challenge participants to build algorithms. Algorithm performance was evaluated in an independent test set of 129 whole-slide images (49 with and 80 without metastases). The same test set of corresponding glass slides was also evaluated by a panel of 11 pathologists with time constraint (WTC) from the Netherlands to ascertain likelihood of nodal metastases for each slide in a flexible 2-hour session, simulating routine pathology workflow, and by 1 pathologist without time constraint (WOTC).

EXPOSURES Deep learning algorithms submitted as part of a challenge competition or pathologist interpretation.

MAIN OUTCOMES AND MEASURES The presence of specific metastatic foci and the absence vs presence of lymph node metastasis in a slide or image using receiver operating characteristic curve analysis. The 11 pathologists participating in the simulation exercise rated their diagnostic confidence as definitely normal, probably normal, equivocal, probably tumor, or definitely tumor.

RESULTS The area under the receiver operating characteristic curve (AUC) for the algorithms ranged from 0.556 to 0.994. The top-performing algorithm achieved a lesion-level, true-positive fraction comparable with that of the pathologist WOTC (72.4% [95% CI, 64.3%-80.4%]) at a mean of 0.0125 false-positives per normal whole-slide image. For the whole-slide image classification task, the best algorithm (AUC, 0.994 [95% CI, 0.983-0.999]) performed significantly better than the pathologists WTC in a diagnostic simulation (mean AUC, 0.810 [range, 0.738-0.884]; $P < .001$). The top 5 algorithms had a mean AUC that was comparable with the pathologist interpreting the slides in the absence of time constraints (mean AUC, 0.960 [range, 0.923-0.994] for the top 5 algorithms vs 0.966 [95% CI, 0.927-0.998] for the pathologist WOTC).

CONCLUSIONS AND RELEVANCE In the setting of a challenge competition, some deep learning algorithms achieved better diagnostic performance than a panel of 11 pathologists participating in a simulation exercise designed to mimic routine pathology workflow; algorithm performance was comparable with an expert pathologist interpreting whole-slide images without time constraints. Whether this approach has clinical utility will require evaluation in a clinical setting.

JAMA. 2017;318(22):2199-2210. doi:10.1001/jama.2017.14585

Editorial page 2184

Related articles page 2211 and page 2250

Supplemental content

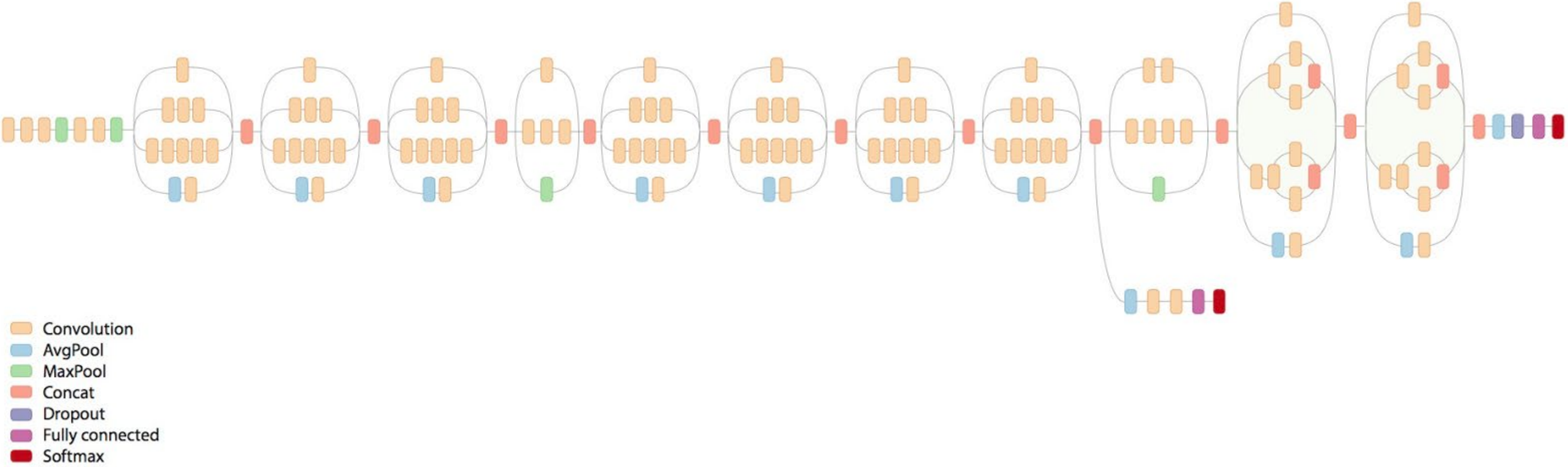
CME Quiz at jamanetwork.com/learning and **CME Questions** page 2252

Author Affiliations: Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, the Netherlands (Ehteshami Bejnordi, van Ginneken, Karssemeijer); Medical Image Analysis Group, Eindhoven University of Technology, Eindhoven, the Netherlands (Veta); Department of Pathology, University Medical Center Utrecht, Utrecht, the Netherlands (Johannes van Diest); Department of Pathology, Radboud University Medical Center, Nijmegen, the Netherlands (Litjens, van der Laak).

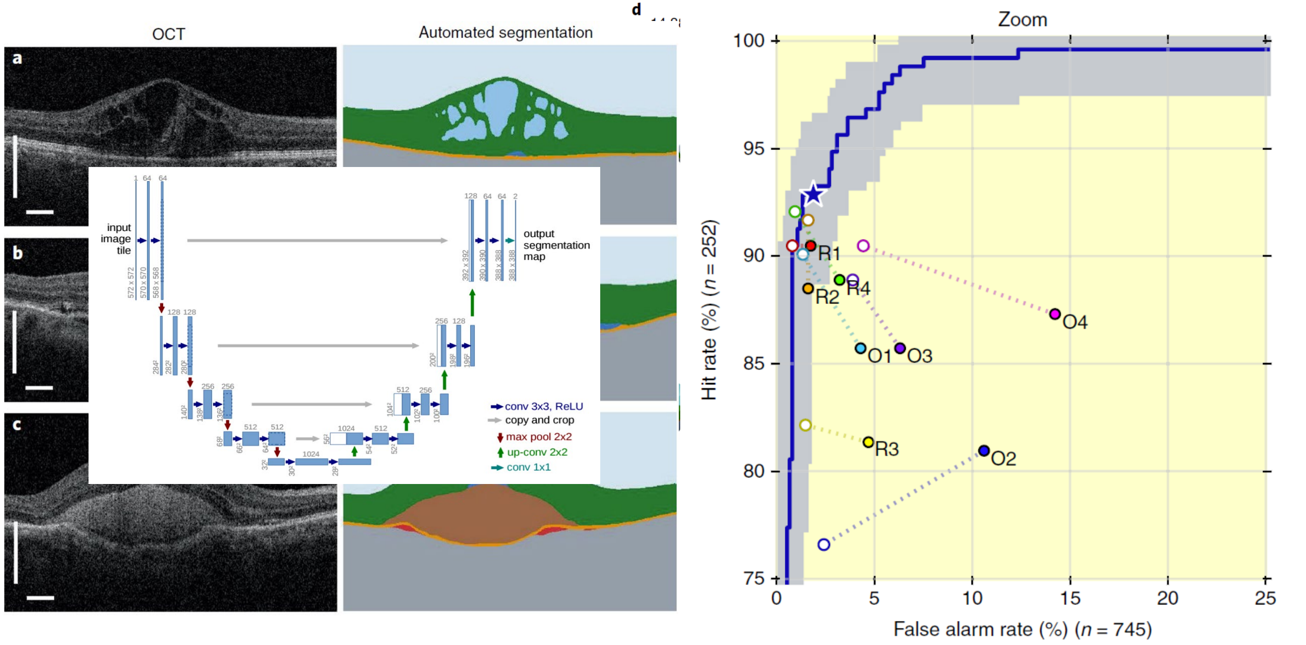
Group Information: The CAMELYON16 Consortium authors and collaborators are listed at the end of this article.

Corresponding Author: Babak Ehteshami Bejnordi, MS, Radboud University Medical Center, Postbus 9101, 6500 HB Nijmegen (ehteshami@babakint.com).


nature
LESIONS LEARNT

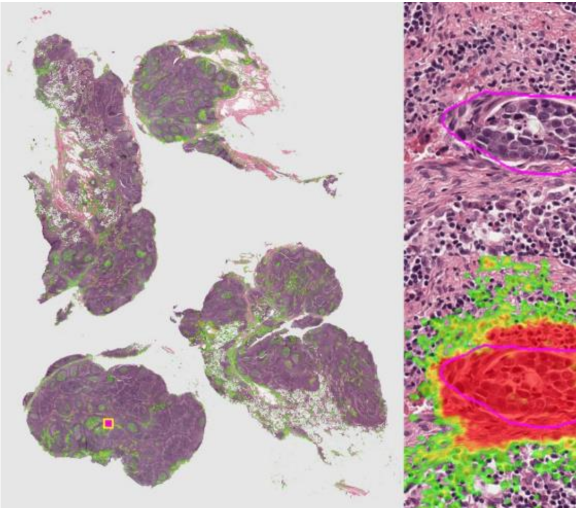

Convolution
AvgPool
MaxPool
Concat
Dropout
Fully connected
Softmax

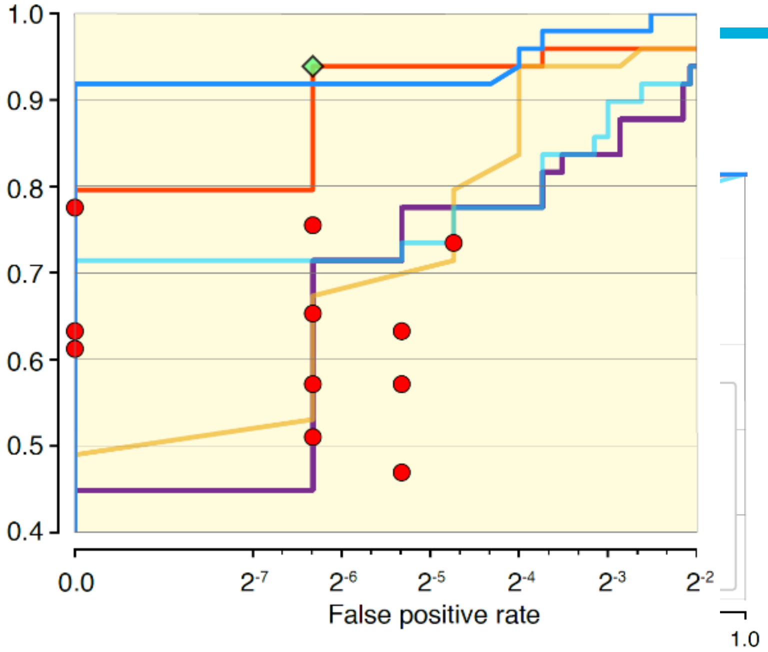

nature medicine
AI accelerates diagnosis
NAD+ biosynthesis and high-risk hospitalizations
Targeted microbiome therapy for thrombosis

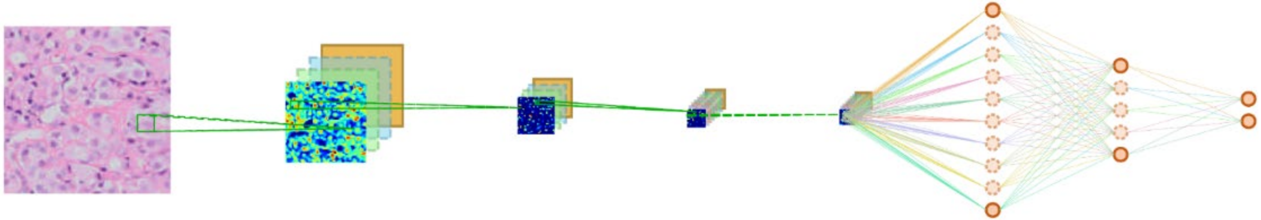

OCT
Automated segmentation
a
b
c
d
input image tile
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1
Zoom
Hit rate (%) (n = 252)
False alarm rate (%) (n = 745)
R1
R2
R3
R4
O1
O2
O3
O4


JAMA | Original Investigation

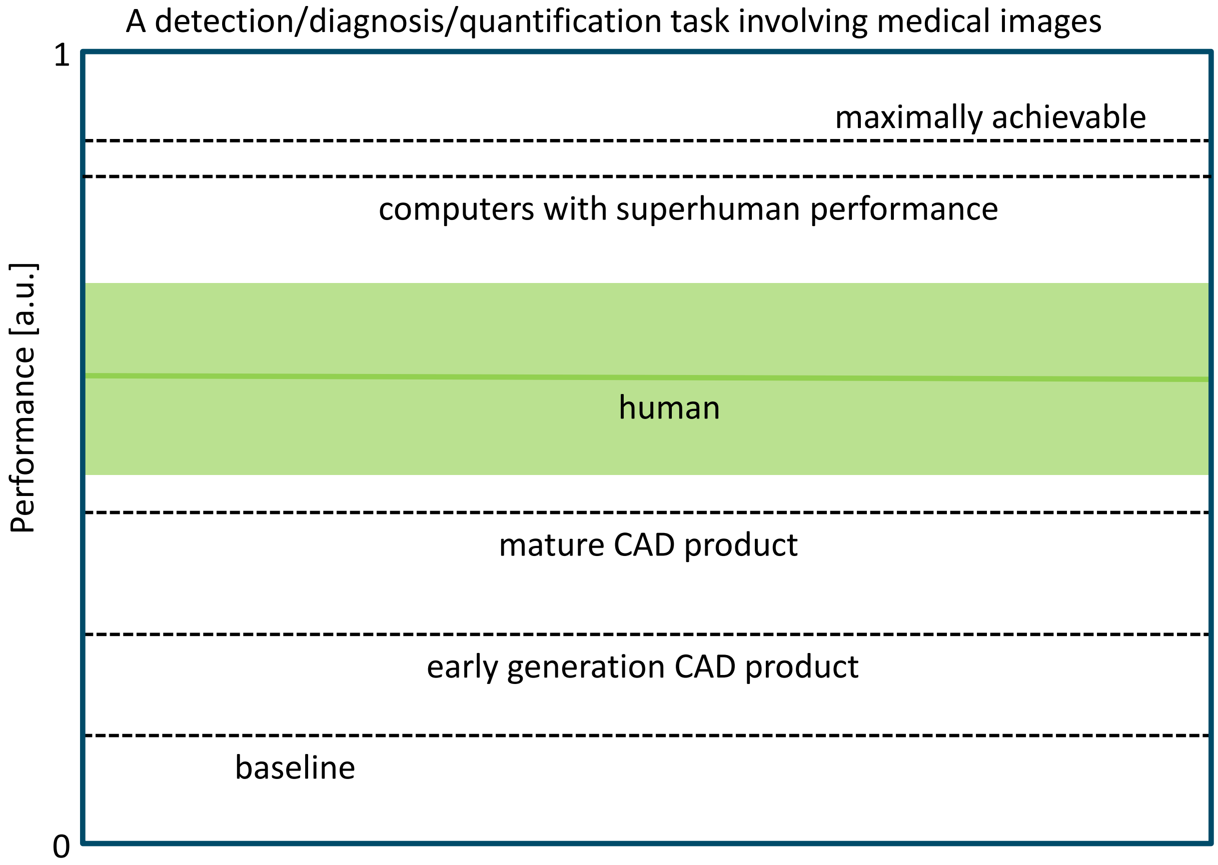
Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer



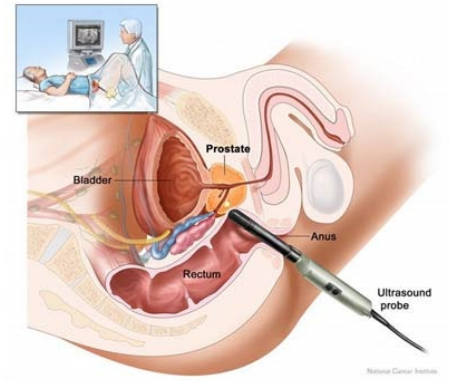


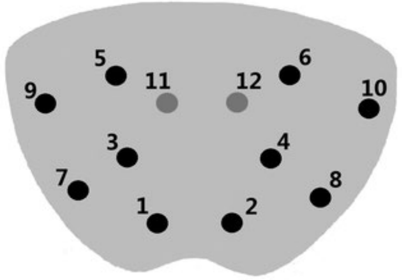


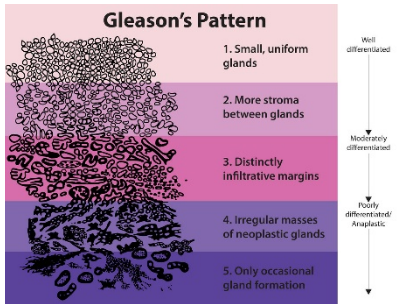
Radboudumc


A detection/diagnosis/quantification task involving medical images
1
maximally achievable
computers with superhuman performance
Performance [a.u.]
human
mature CAD product
early generation CAD product
baseline
0

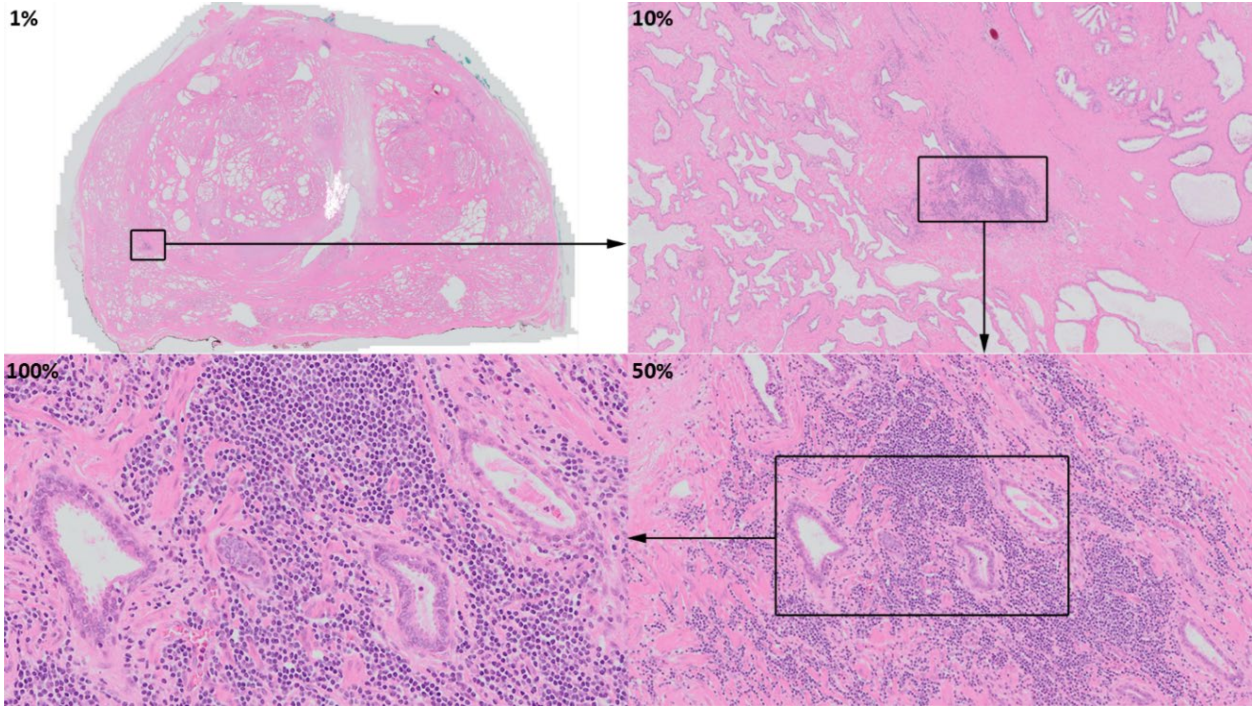
Prostate Cancer Diagnosis





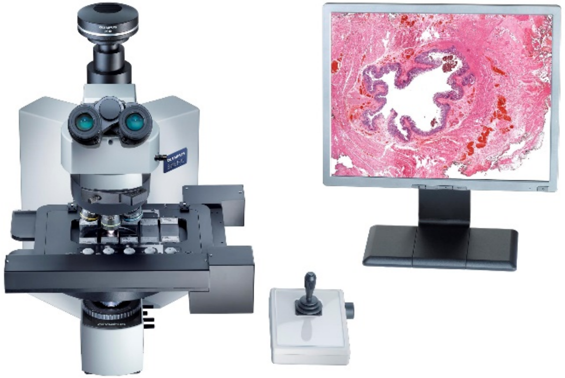


Histopathology data not digital

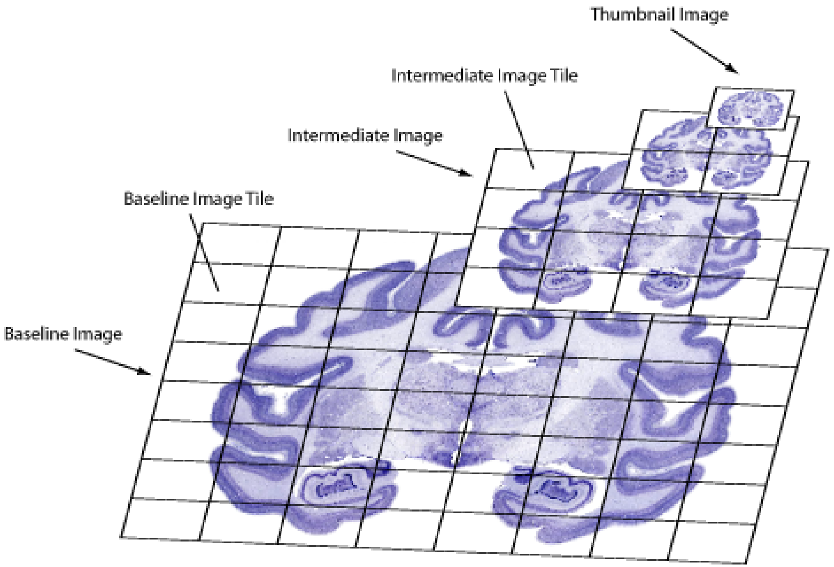


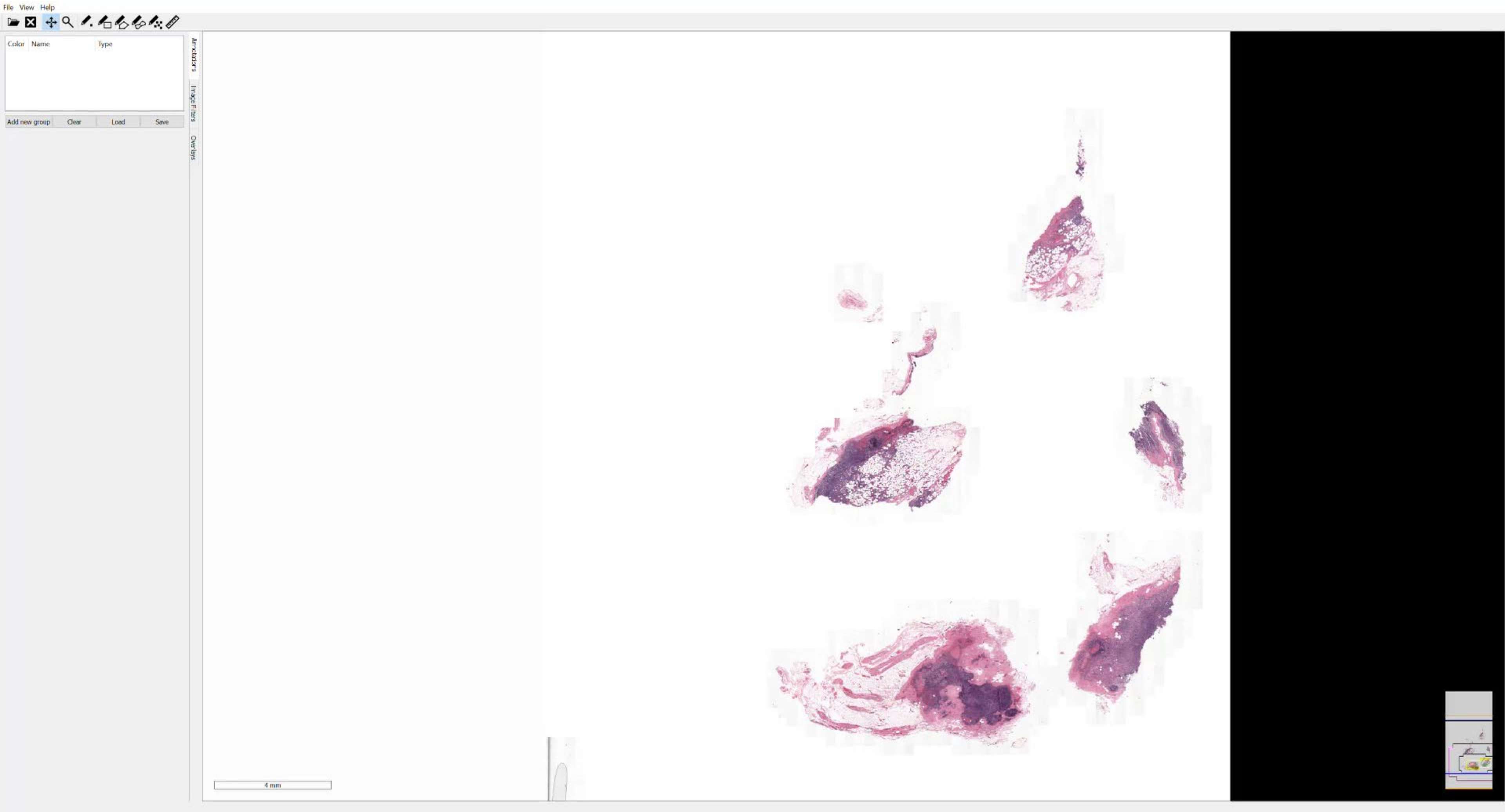
Whole-slide imaging

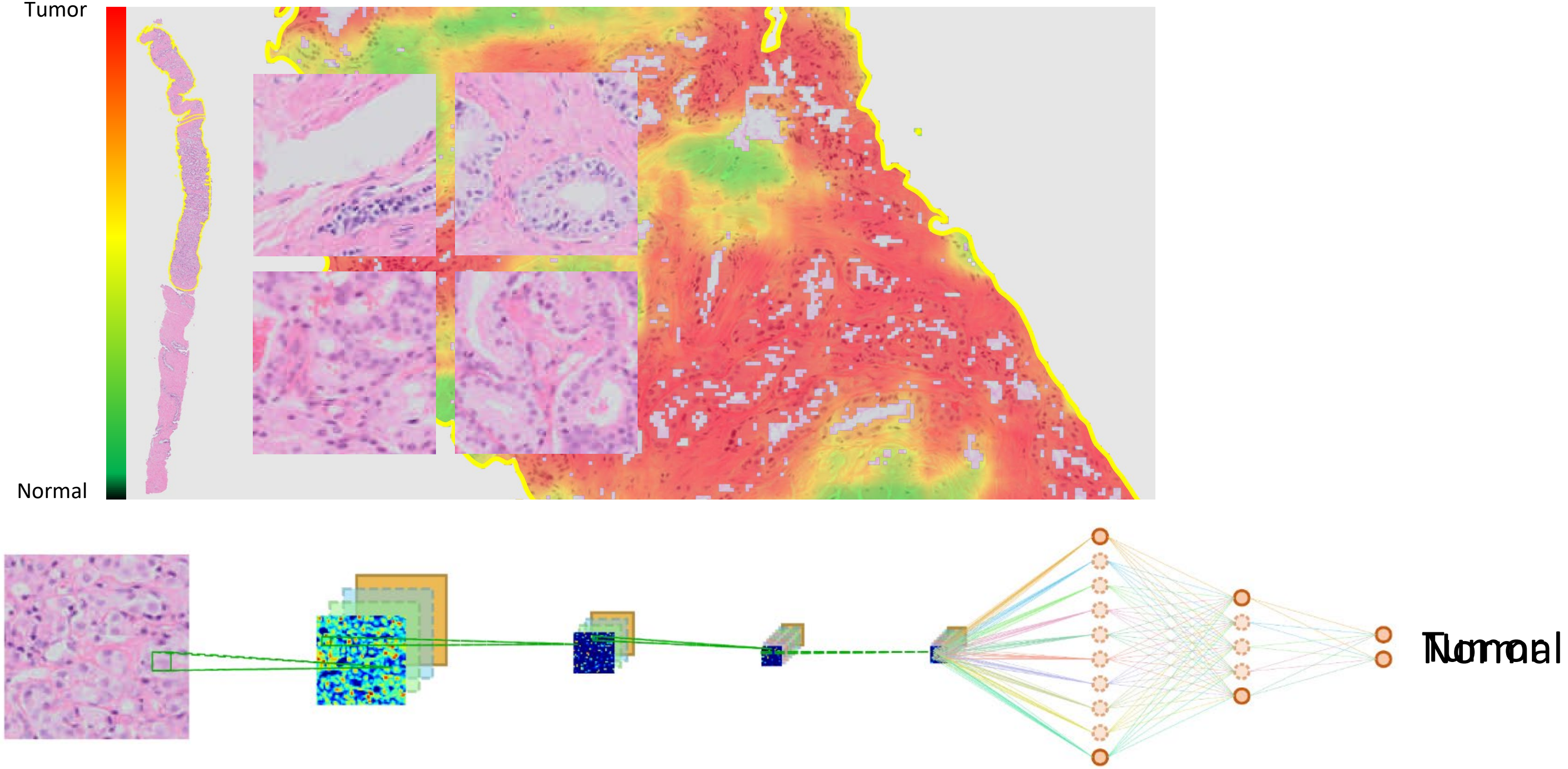
Digital acquisition of an entire histopathology slide



Whole-slide imaging

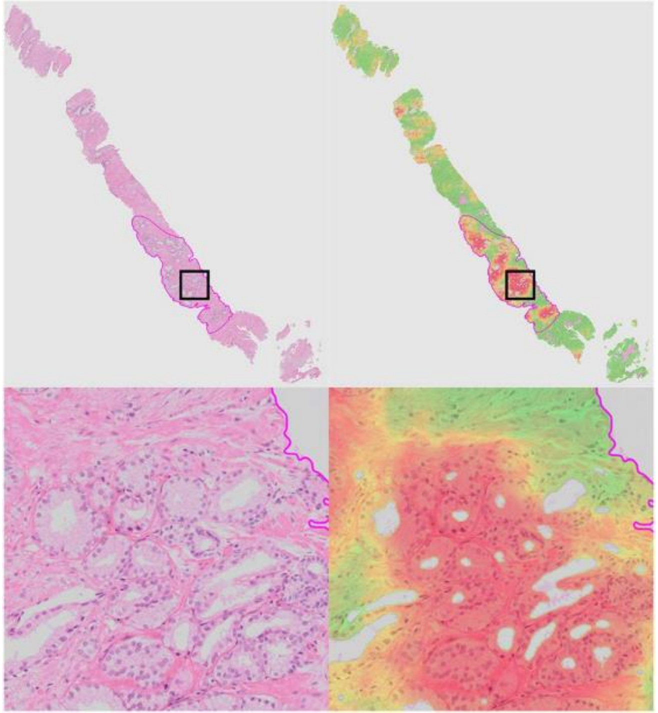


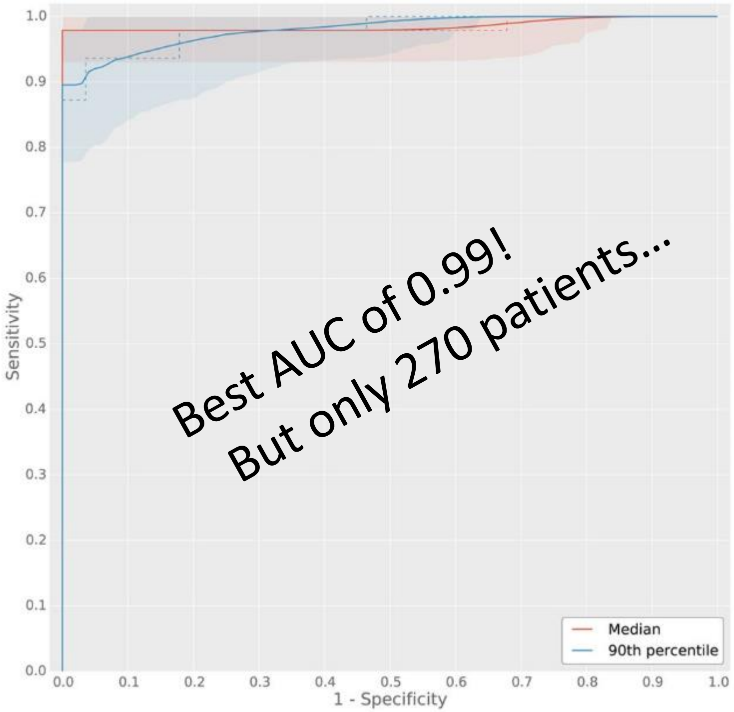

File View Help
Color Name Type
Add new group
Clear
Load
Save
Annotations
Image Filters
Overlays
4 mm



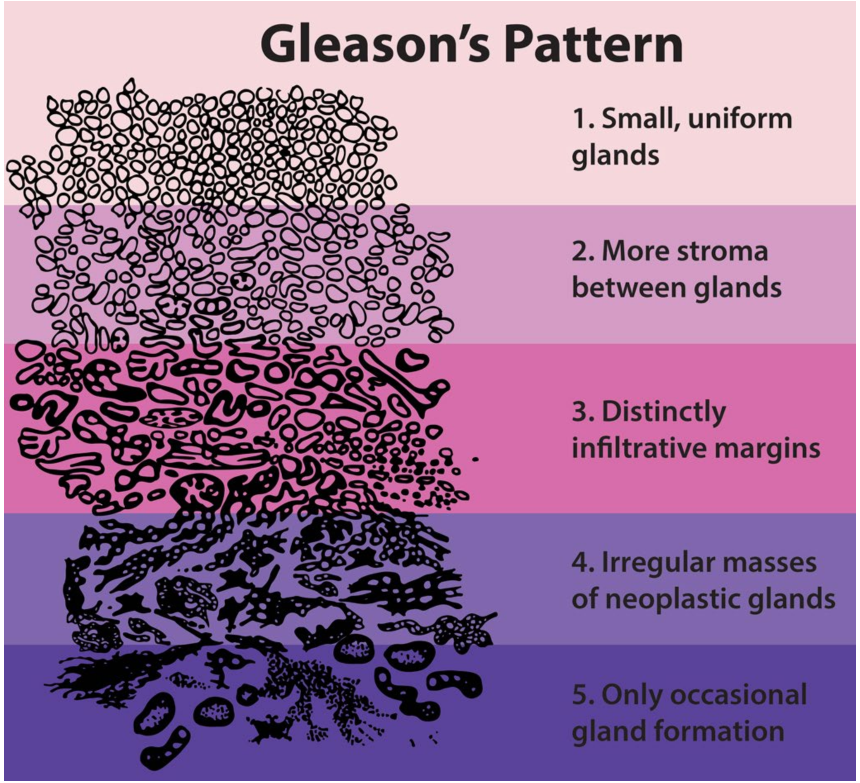
Bulten et al. *Sci Rep.* (2016)

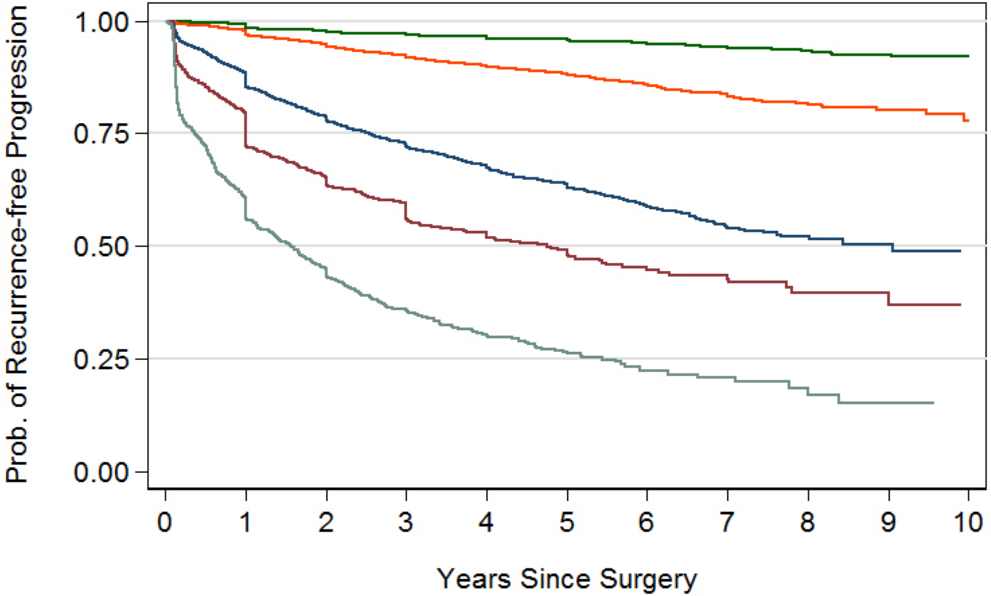
Cancer detection





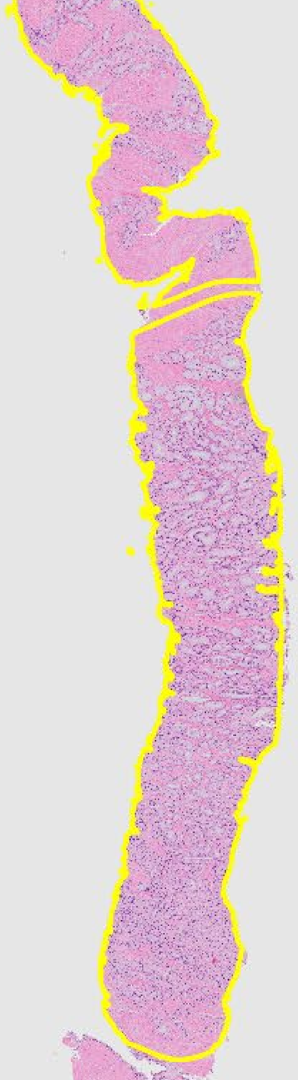

Tumor

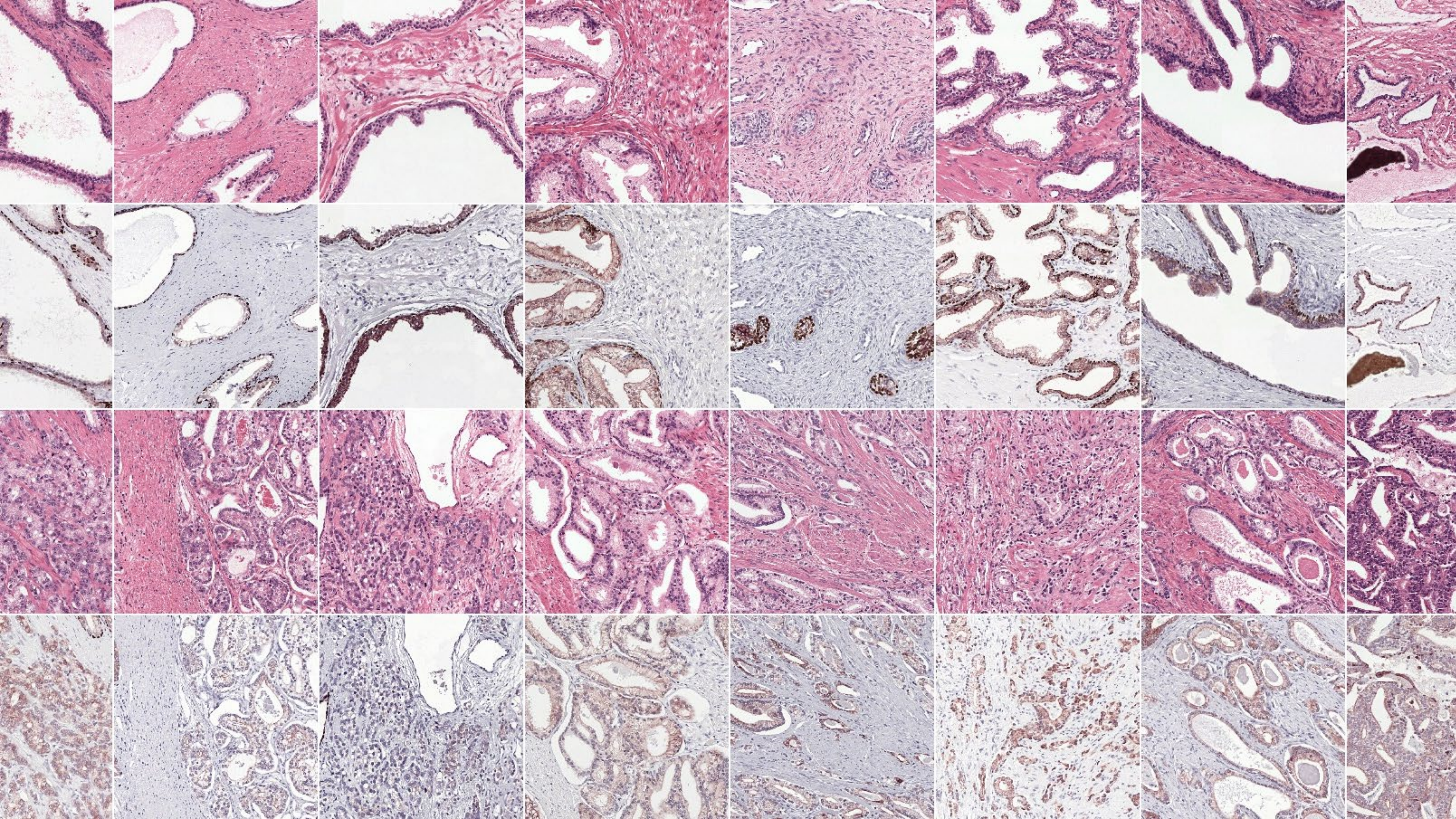


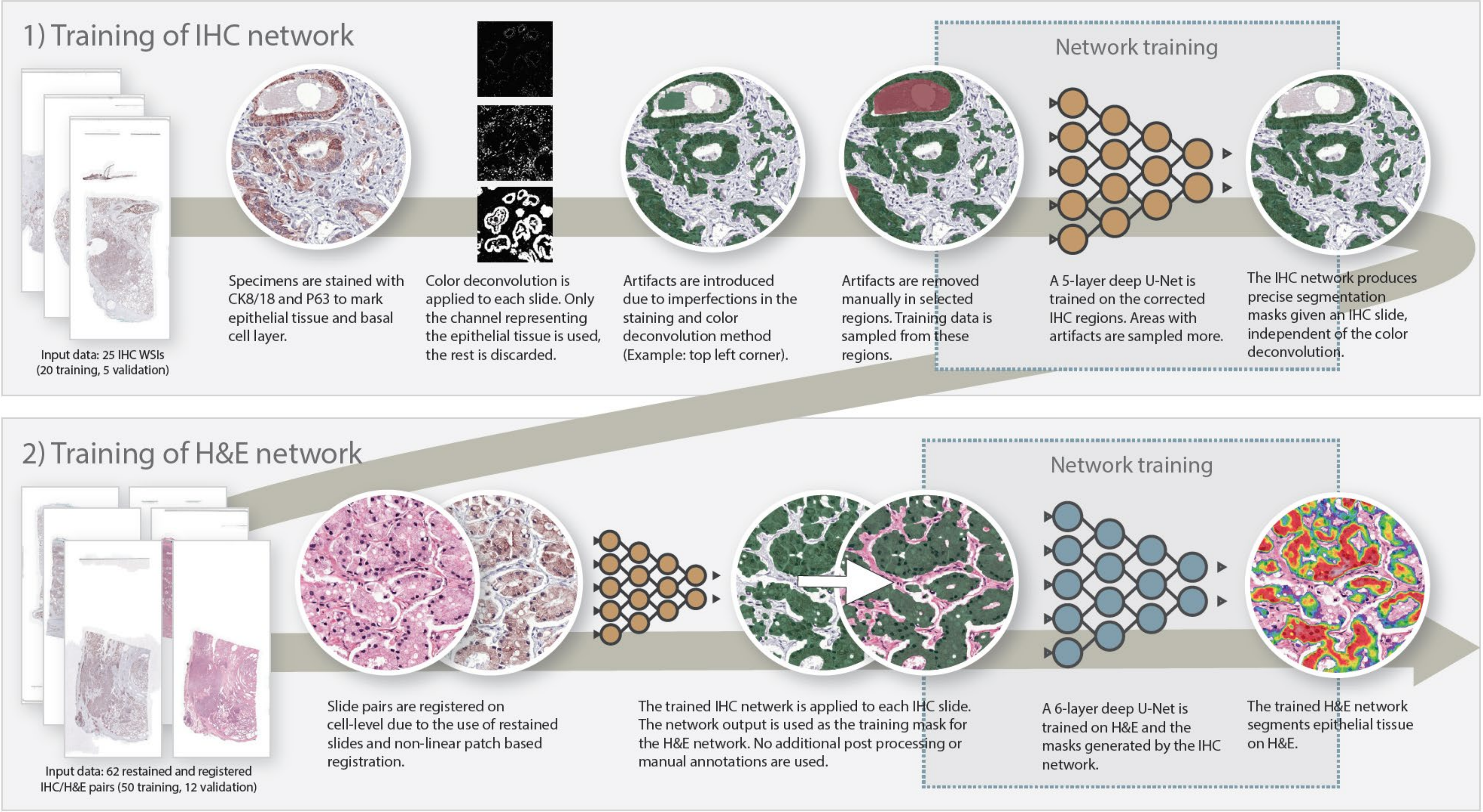


Number at risk

6	7397	6973	5104	4064	3226	2461	1768	1186	670	278	108
3+4	8353	7202	5298	3983	2955	2091	1299	778	393	135	45
4+3	3106	2452	1605	1152	839	569	350	199	90	38	15
8	917	678	412	280	191	129	86	59	35	14	7
>=9	1051	578	325	194	118	73	41	24	12	4	2

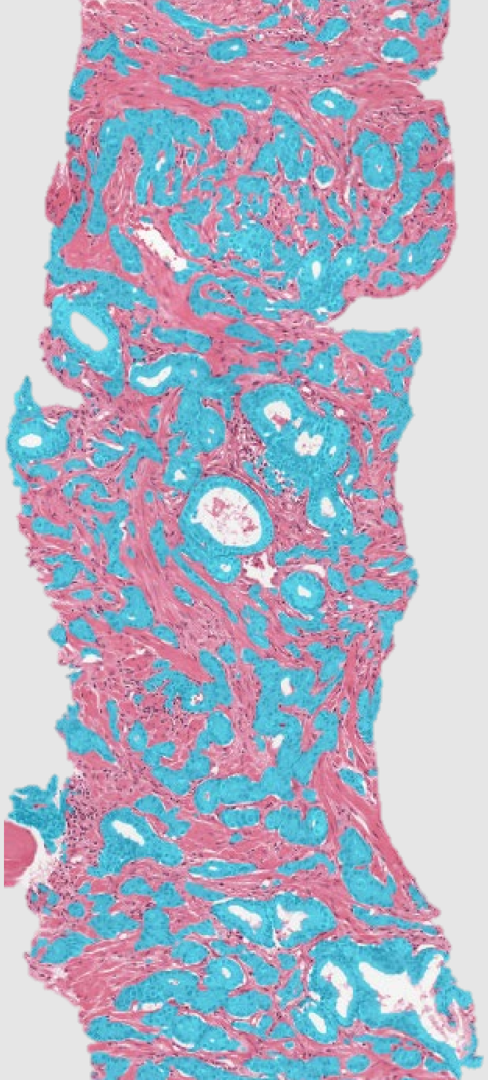




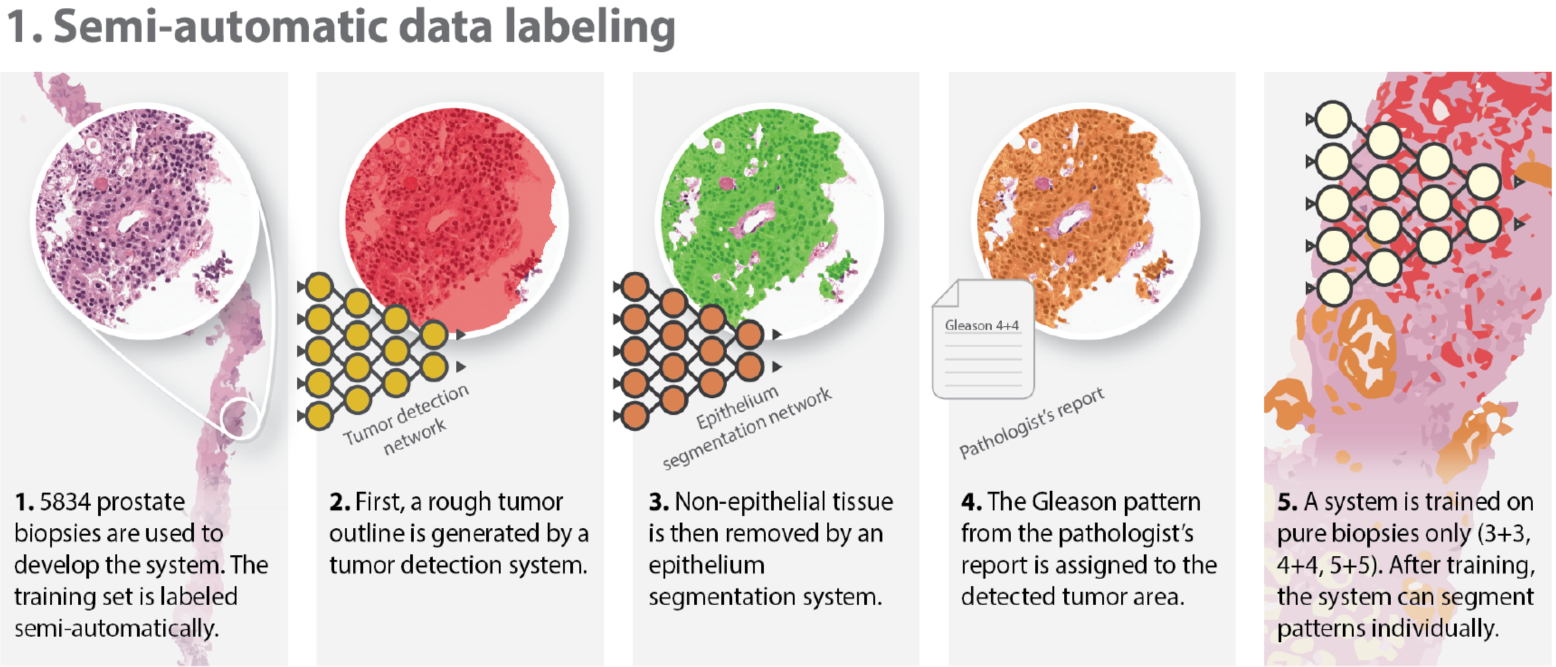


Bulten et al. *Sci Rep.* (2019)

Radboudumc



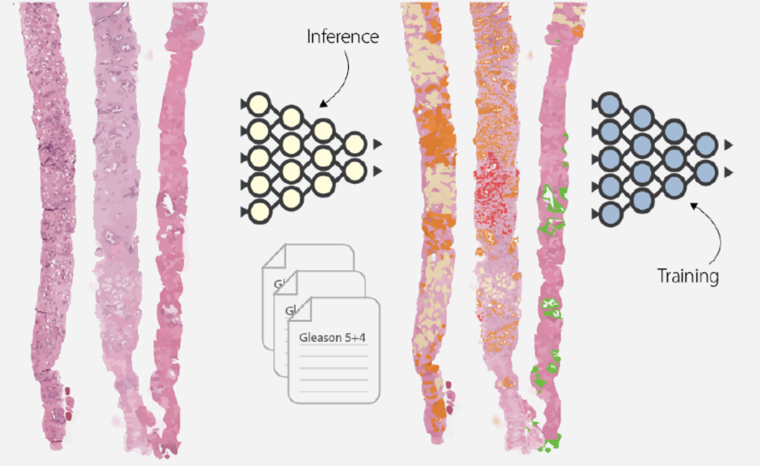
Gleason grading



Bulten et al. Automated Gleason Grading of Prostate Biopsies using Deep Learning. *Lancet Oncology.* Accepted

Gleason grading

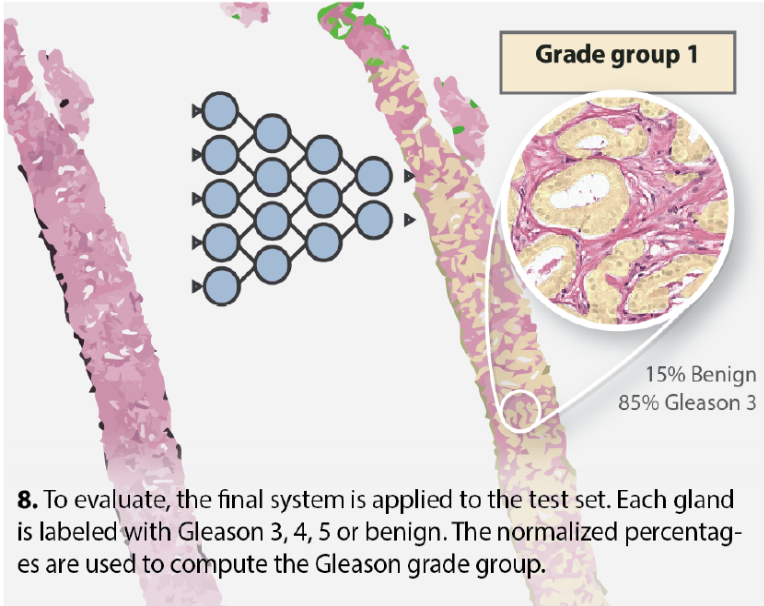
2. Refinement & training



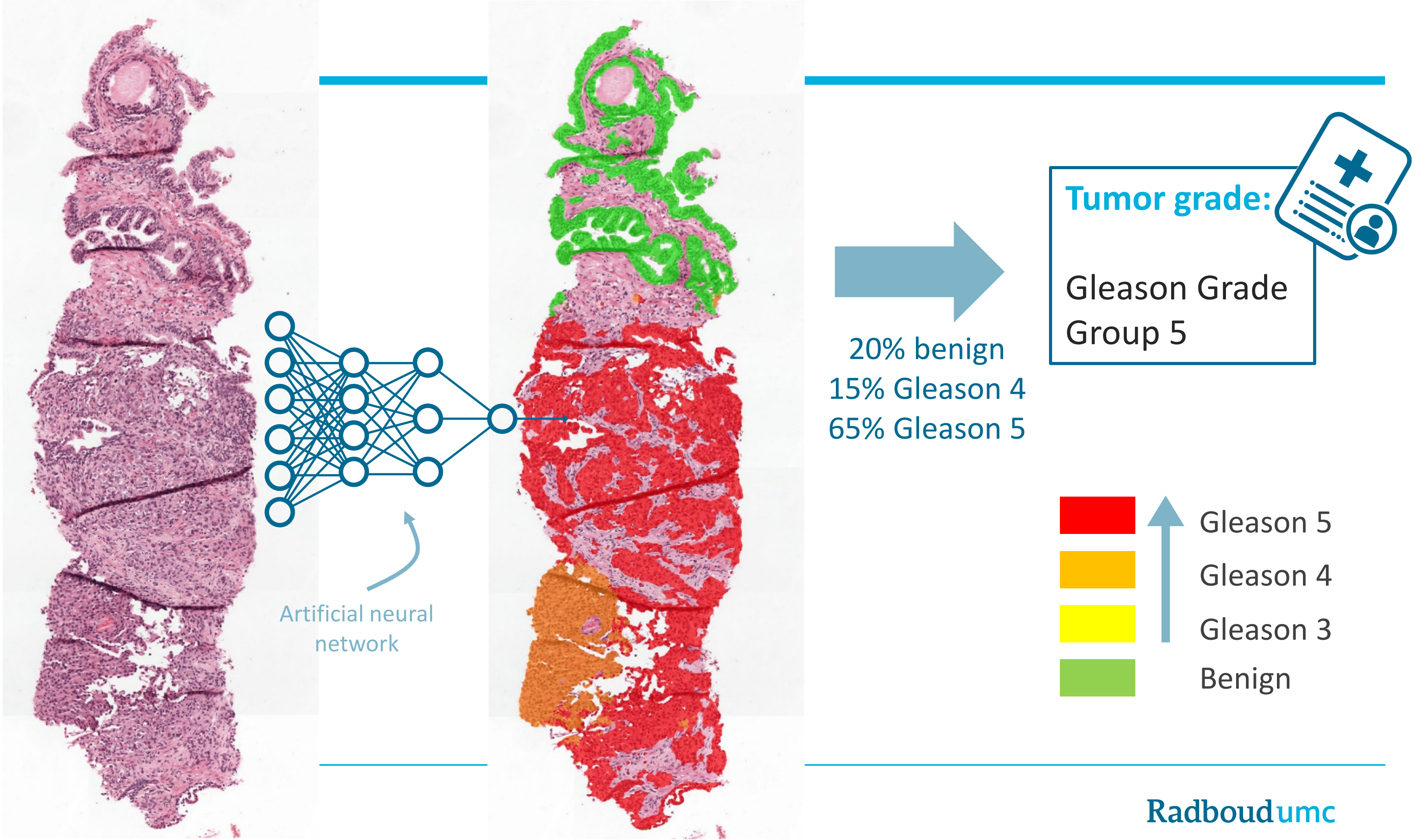
6. The full training set is labeled using the network trained on pure biopsies. Reports are used to further refine the labels.

7. Using the new labels the final system is trained.

3. Grade group prediction

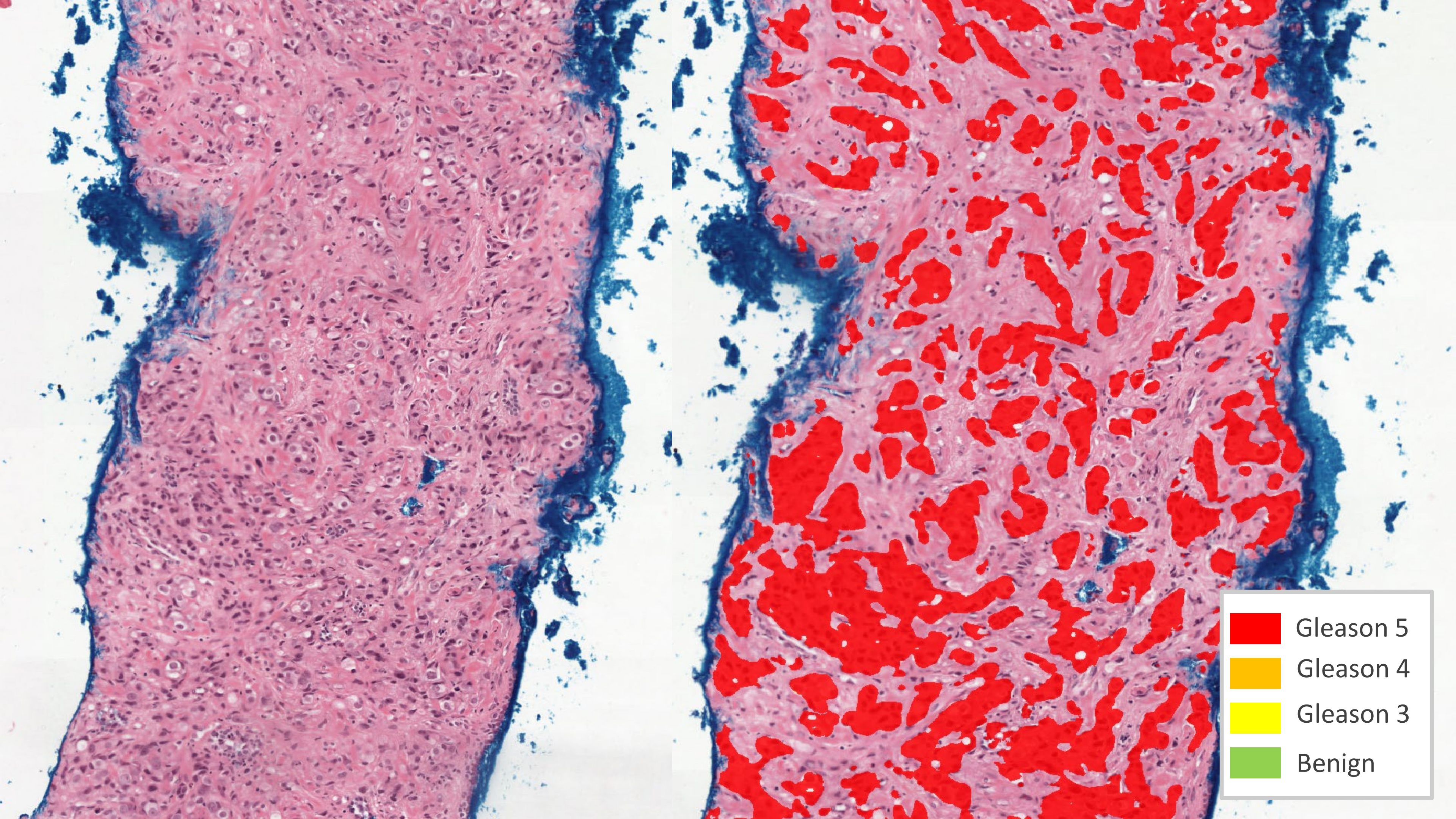


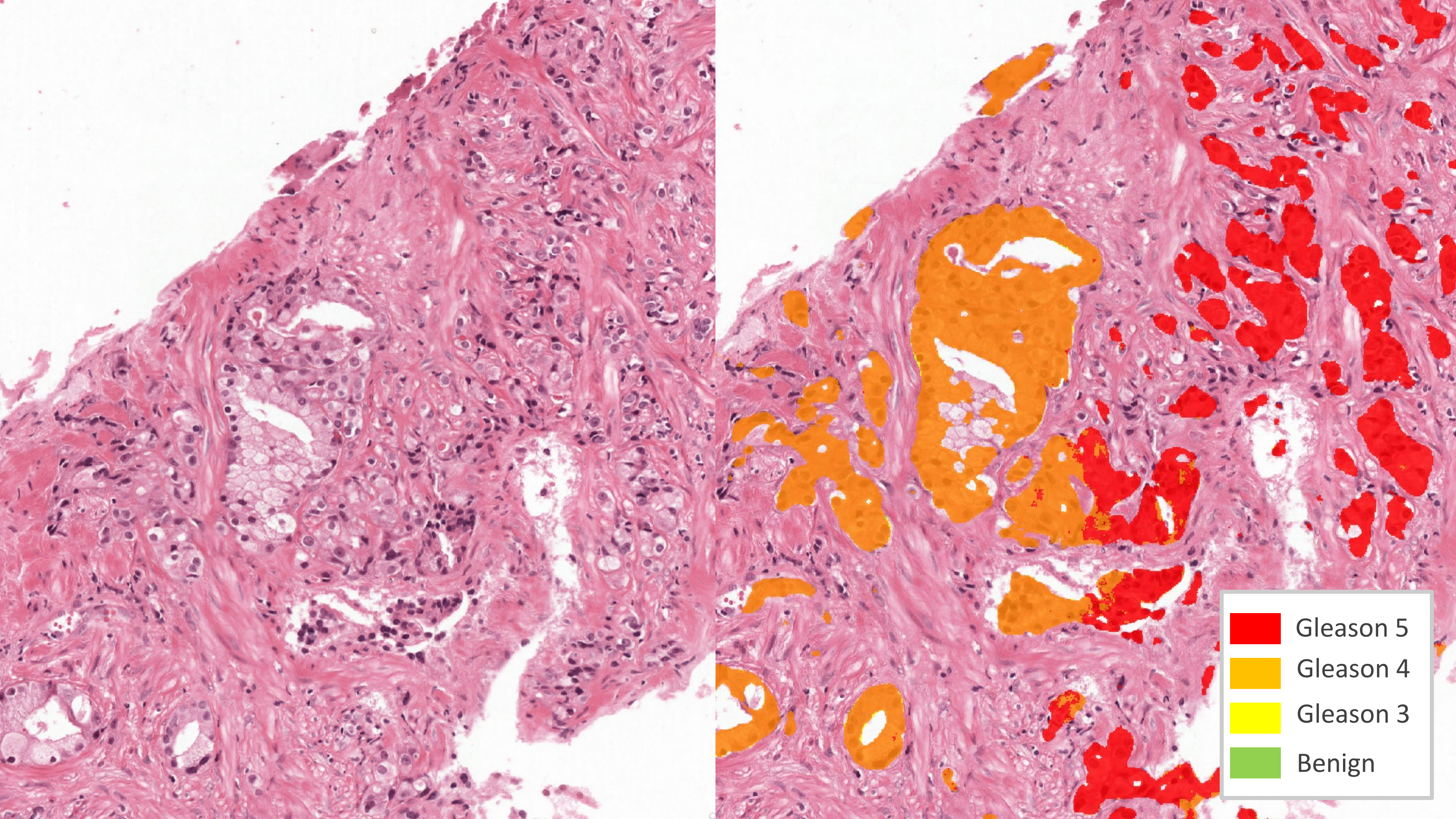
8. To evaluate, the final system is applied to the test set. Each gland is labeled with Gleason 3, 4, 5 or benign. The normalized percentages are used to compute the Gleason grade group.

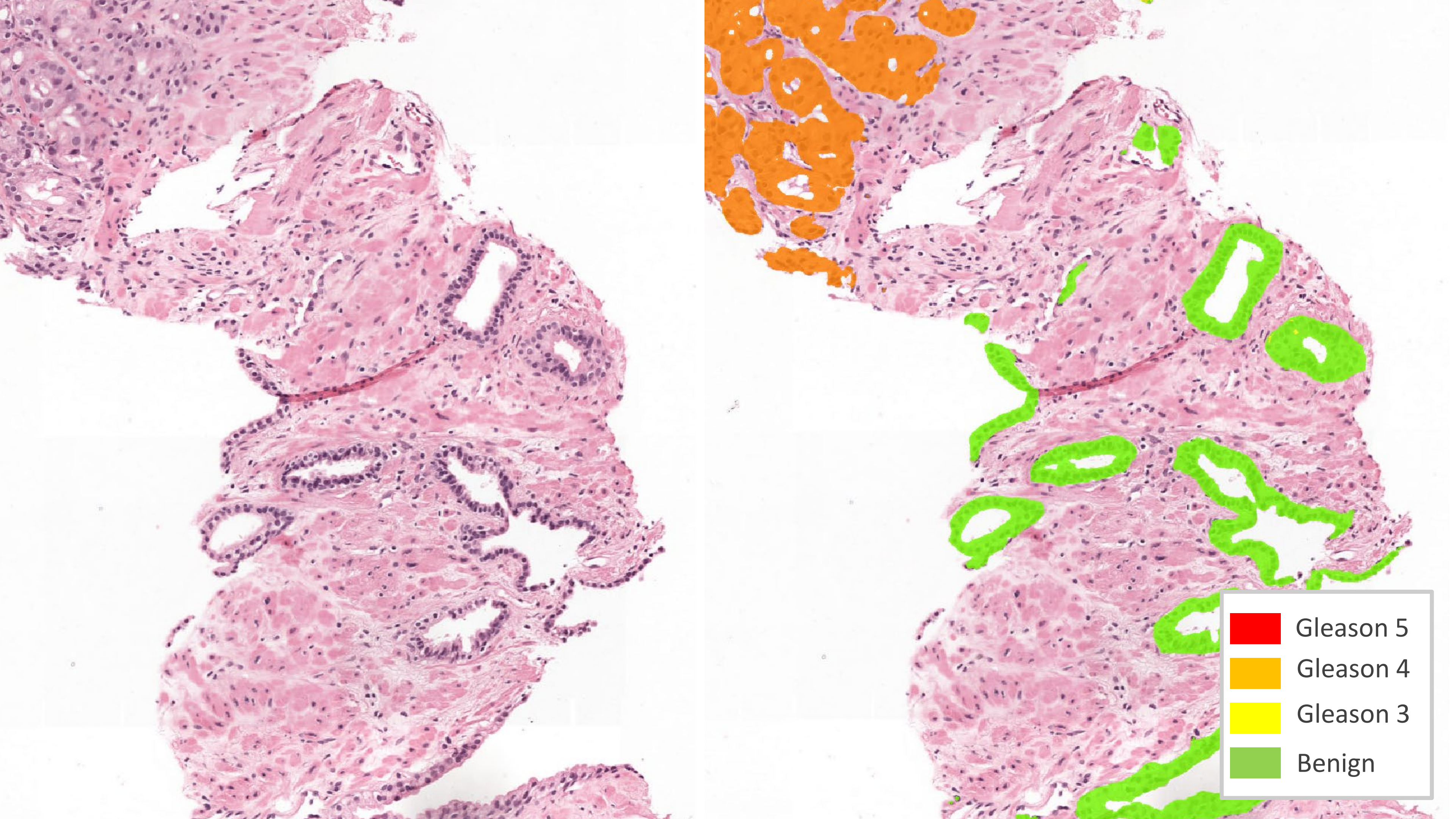


Artificial neural network
20% benign
15% Gleason 4
65% Gleason 5
Tumor grade:
Gleason Grade Group 5
Gleason 5
Gleason 4
Gleason 3
Benign

Radboudumc


Gleason 5
Gleason 4
Gleason 3
Benign

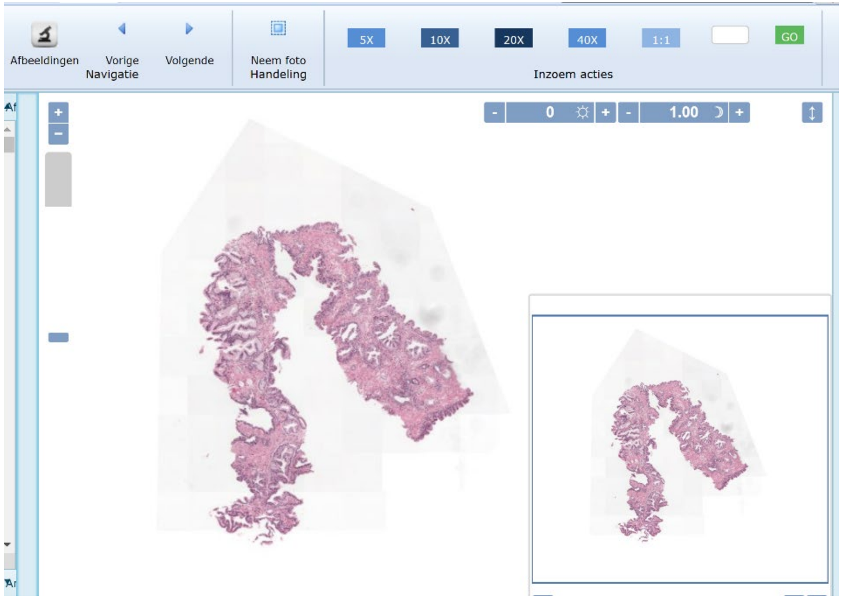

Gleason 5
Gleason 4
Gleason 3
Benign


Gleason 5
Gleason 4
Gleason 3
Benign

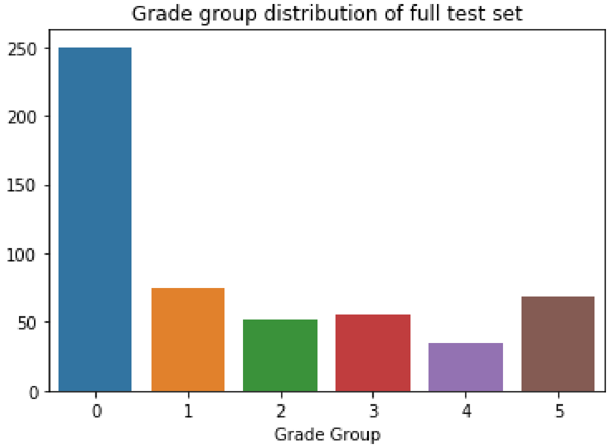
Reference standard

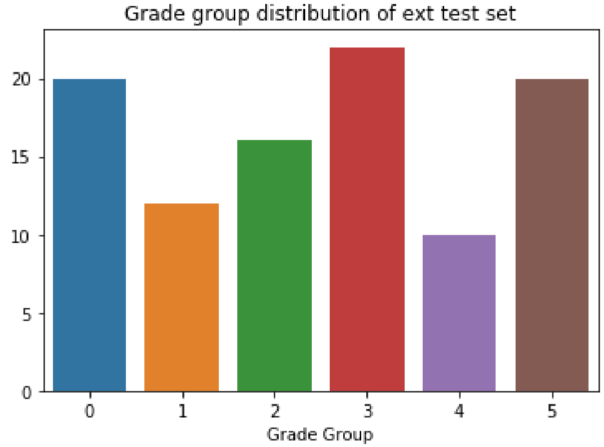
- 3 expert uropathologists
- 550 prostate biopsies

- Gleason growth patterns, tumor volumes & grade groups



Case distribution

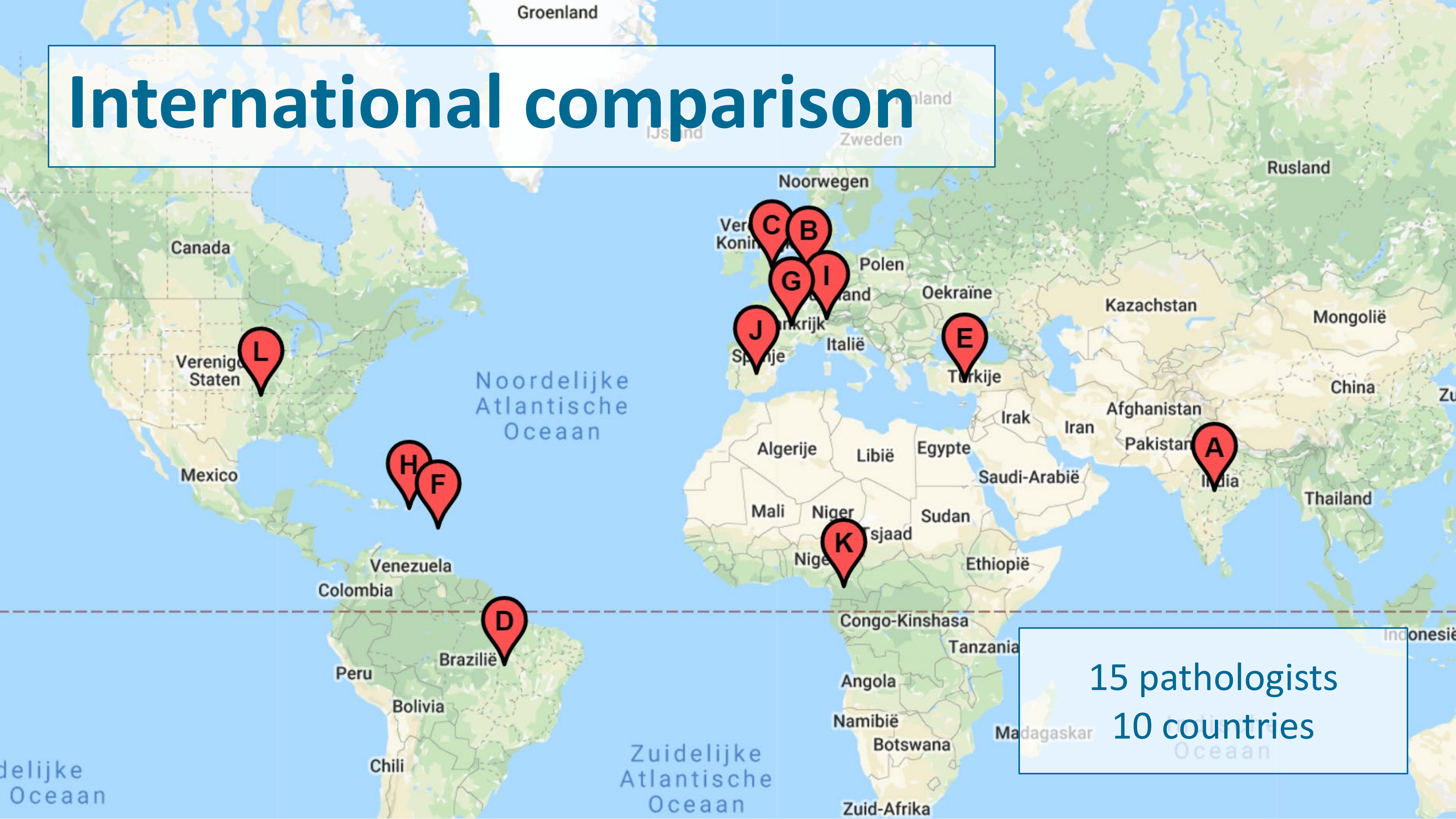




Full test set:
- 550 biopsies
- 3 experts

Subset
- 100 biopsies (selected from the 550)
- 15 external pathologists

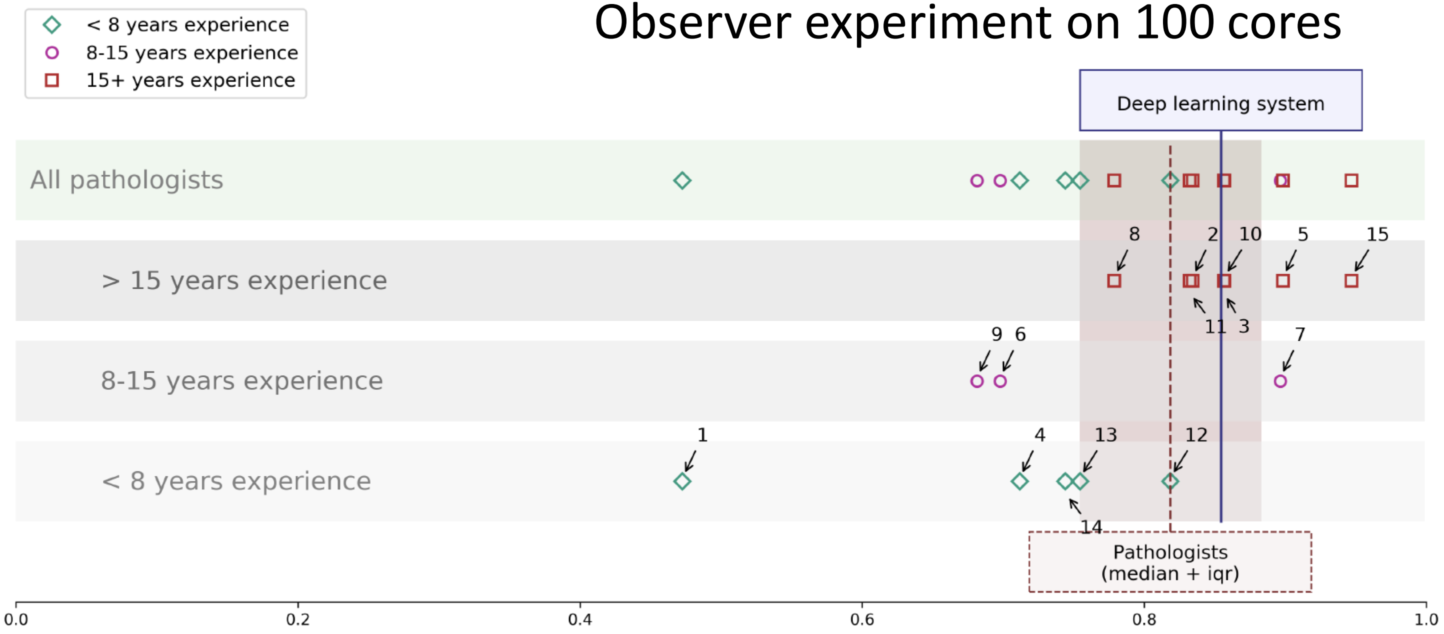
International comparison

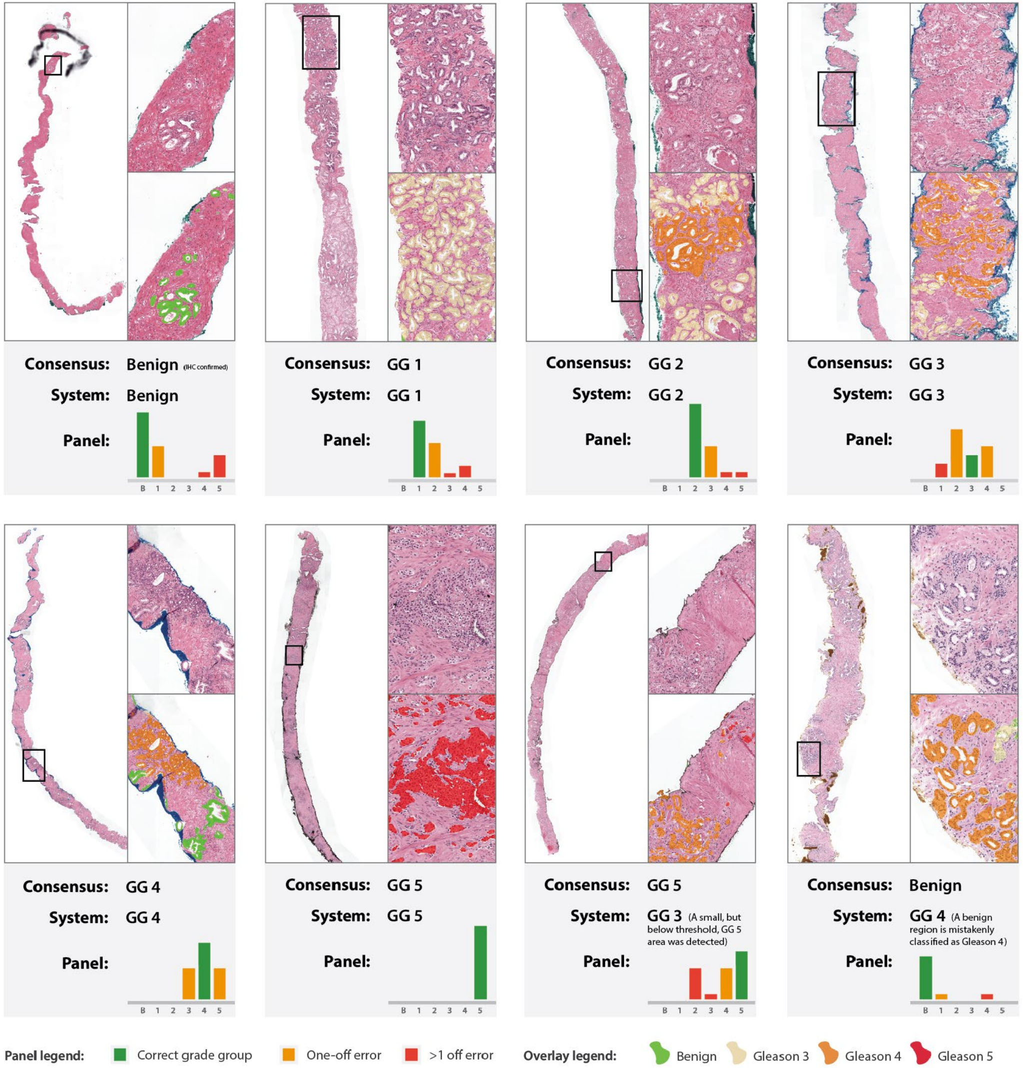


15 pathologists
10 countries

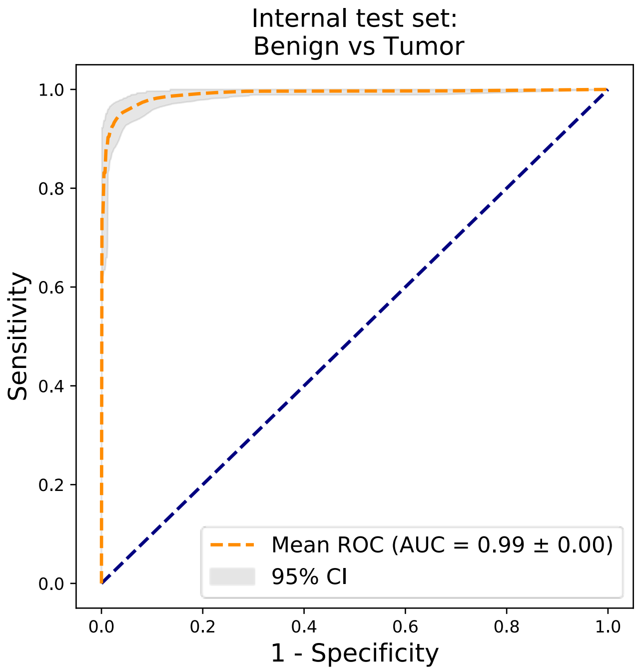
Gleason grading

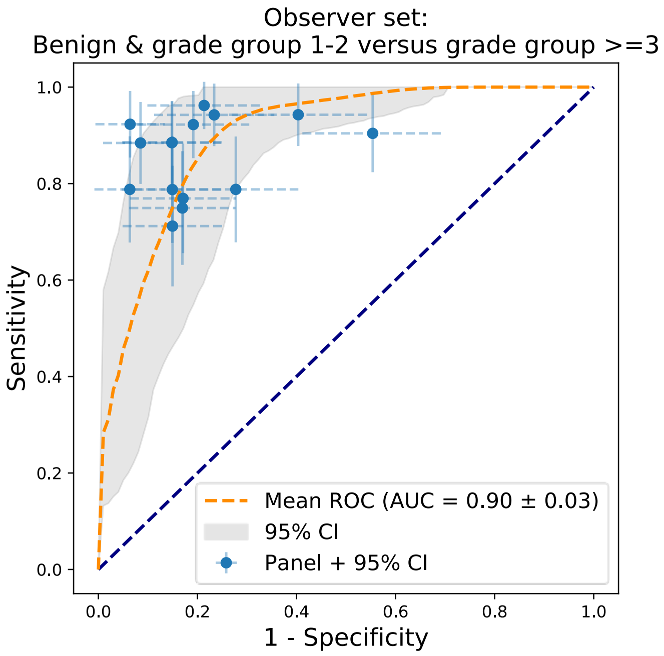
Observer experiment on 100 cores




Consensus: Benign (IHC confirmed)
System: Benign
Panel:
Consensus: GG 1
System: GG 1
Panel:
Consensus: GG 2
System: GG 2
Panel:
Consensus: GG 3
System: GG 3
Panel:
Consensus: GG 4
System: GG 4
Panel:
Consensus: GG 5
System: GG 5
Panel:
Consensus: GG 5
System: GG 3 (A small, but below threshold, GG 5 area was detected)
Panel:
Consensus: Benign
System: GG 4 (A benign region is mistakenly classified as Gleason 4)
Panel:
B 1 2 3 4 5
Panel legend: Correct grade group One-off error >1 off error
Overlay legend: Benign Gleason 3 Gleason 4 Gleason 5

Classification – ROC

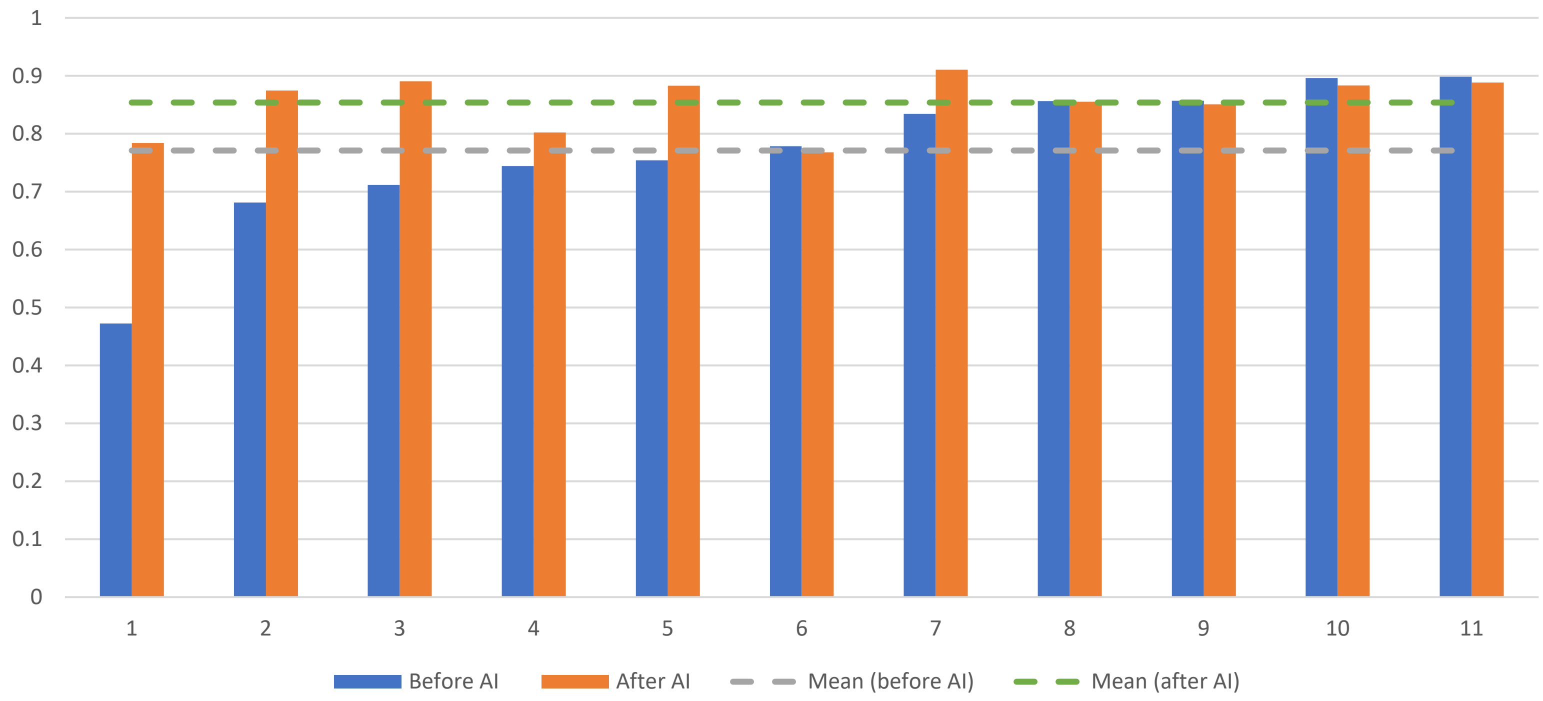




N=535

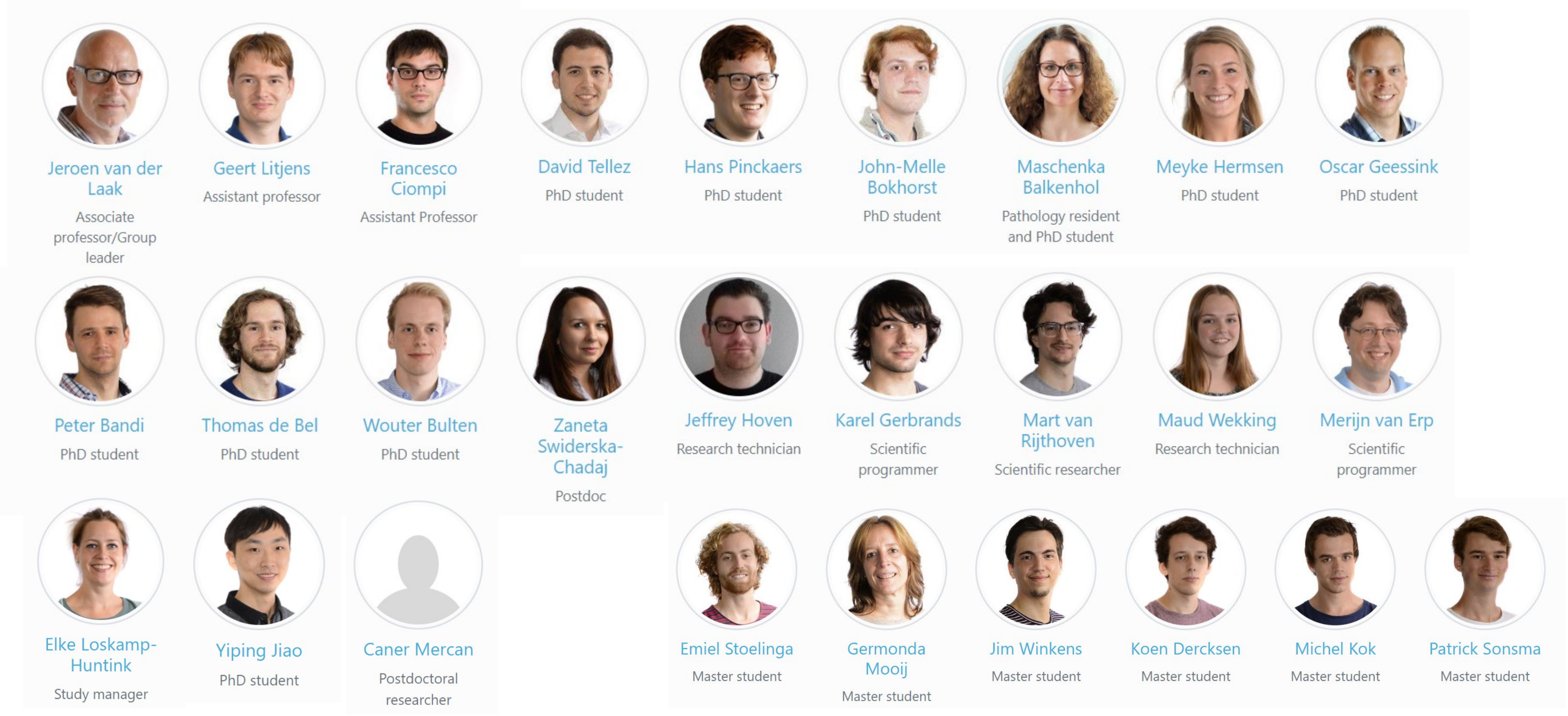
N=100

Usefulness of AI in practice



Future directions

- Algorithm available via grand-challenge.org
- Decision thresholds for grade groups
- Correlation with survival / recurrence
- Direct prediction of survival / recurrence from morphological patterns



Want to learn more? Visit www.computationalpathology.eu

Radboudumc

ST IT PRO
stichting IT projecten

SEVENTH FRAMEWORK PROGRAMME

Alpe d'HuZes

KWF KANKER BESTRIJDING

NWO

ZonMw

RESEARCH & INNOVATION
Horizon 2020

Radboudumc
AMI
Fraunhofer MEVIS

KWF KANKER BESTRIJDING

tEPIS

trait
a ctmm project

Institute for Health Sciences
Radboudumc