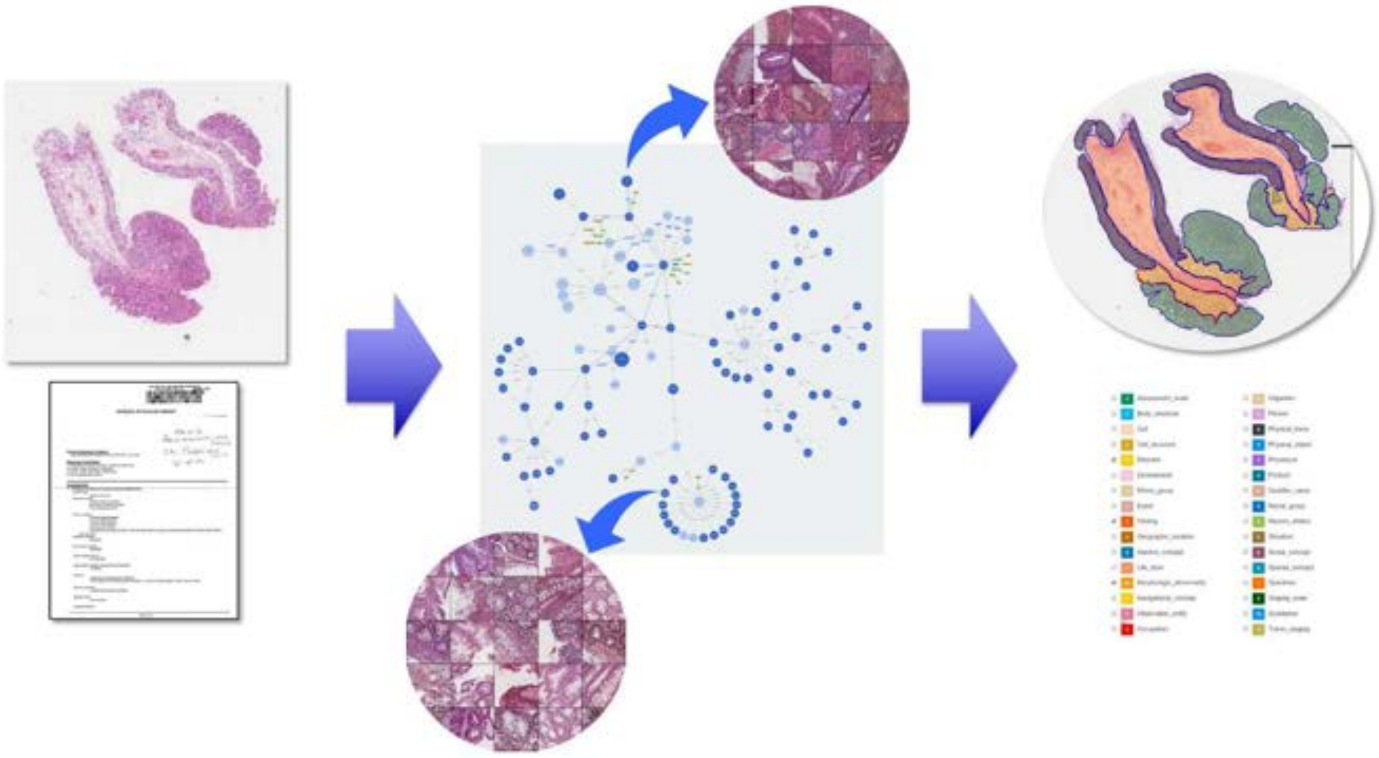
Hes·so // VALAIS WALLIS
Haute Ecole de Gestion & Tourisme
Hochschule für Wirtschaft & Tourismus

Challenges for machine learning using medical data



Henning Müller

AIDA lecture

BETA AIDA
ARTIFICIAL INTELLIGENCE DOCTORAL ACADEMY

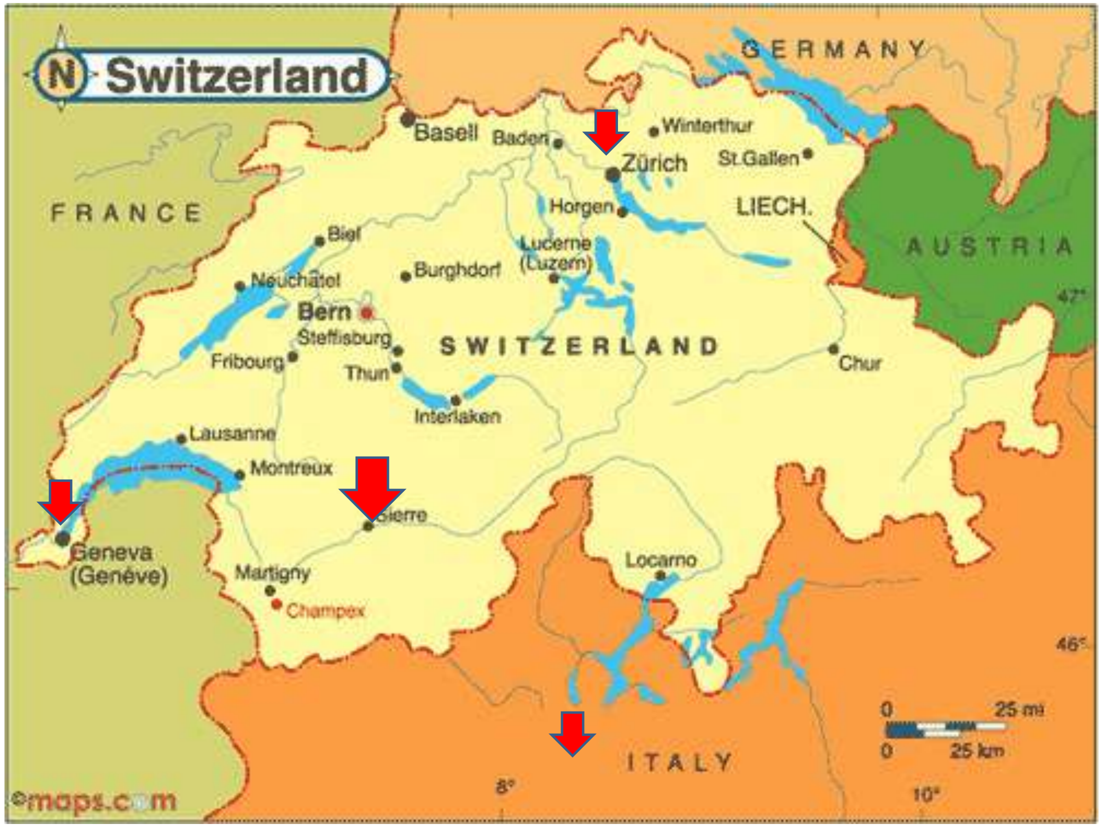
AI4media

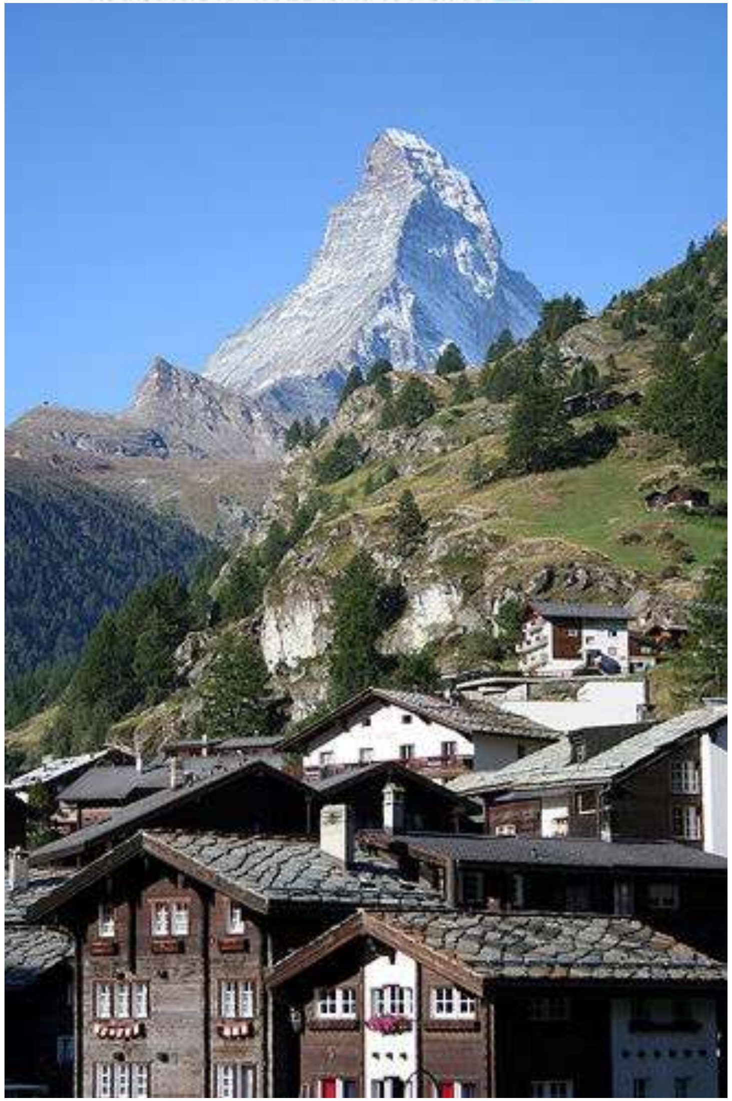
ExaMode

Henning Müller

- Medical informatics studies in Heidelberg, Germany
  - Exchange with Daimler Benz research, USA
- PhD in CBIR, computer vision, Geneva, Switzerland (1998-2002)
  - Exchange with Monash University, Melbourne, AUS
- Professor in radiology and medical informatics at the University of Geneva (2014-)
- Professor in Computer Science at the HES-SO, Sierre, Switzerland (2007-)
  - Visiting faculty at Martinos Center (2015-2016)
- Member of the Swiss National Research Council

Where we are


Switzerland
N
GERMANY
FRANCE
AUSTRIA
LIECH.
SWITZERLAND
ITALY
Basel
Baden
Zürich
Winterthur
St.Gallen
Horgen
Biel
Lucerne
(Luzern)
Burghdorf
Neuchâtel
Bern
Steffisburg
Fribourg
Thun
Interlaken
Chur
Lausanne
Montreux
Sierre
Geneva
(Genève)
Martigny
Champex
Locarno
47°
46°
8°
10°
0
25 mi
0
25 km
©maps.com



Overview

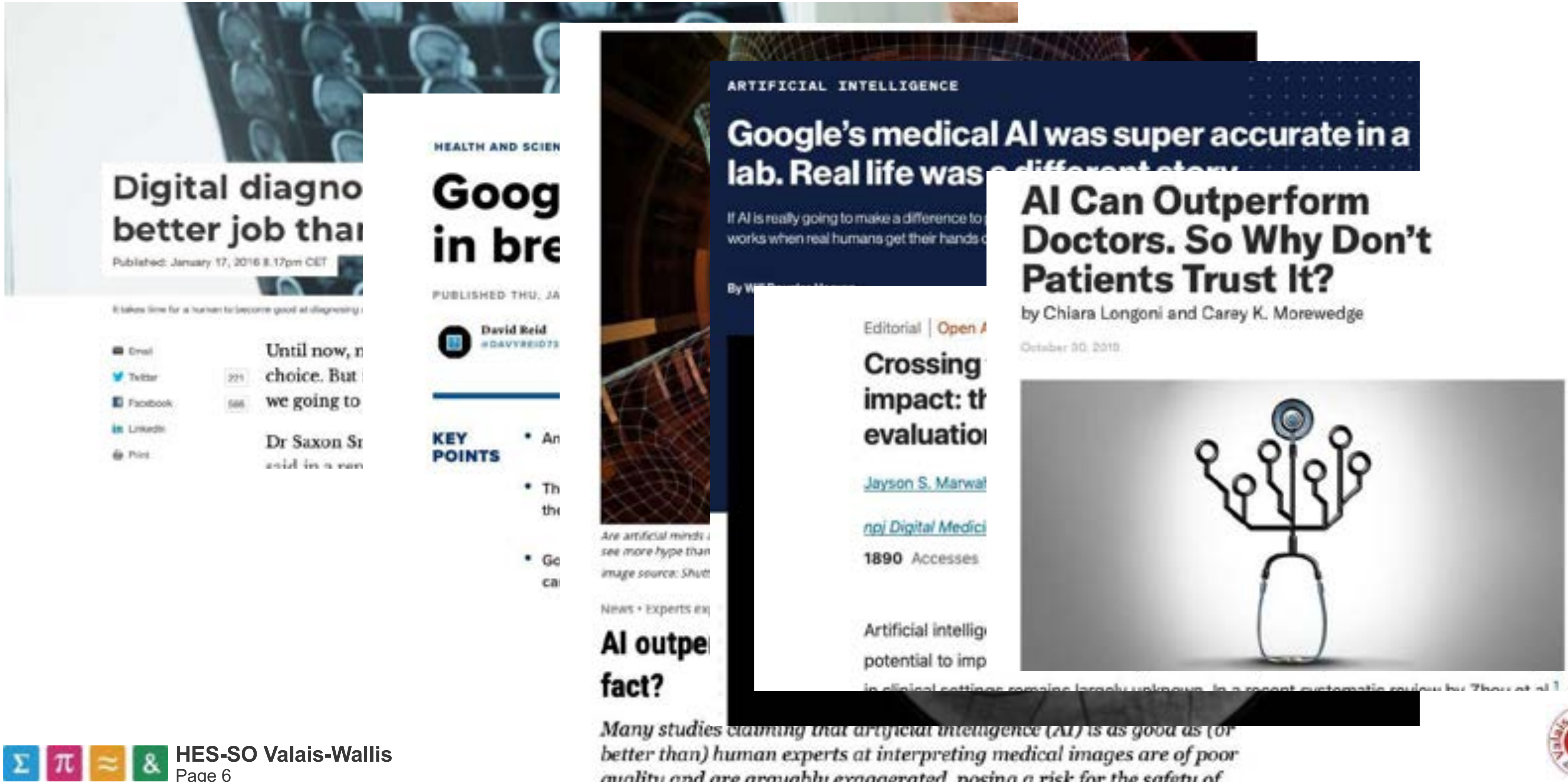
- Status of medical AI
  - And its challenges
- Projects addressing the challenges
  - With a bias towards our work
- Open challenges
- Conclusions and discussion

The promise of medical AI

Geoff Hinton: On Radiology

https://www.youtube.com/watch?v=2HMPRXstSvQ

Medical AI in the media



Digital diagno
better job thar
Published: January 17, 2016 8.17pm CET

It takes time for a human to become good at diagnosing

Email
Twitter 221
Facebook 586
LinkedIn
Print

Until now, n
choice. But
we going to

Dr Saxon Sr
said in a ren

HEALTH AND SCIEN

Goog
in bre

PUBLISHED THU, JA

David Reid
@DAVYREID73

KEY POINTS

- An
- Th
the
- Go
ca

Are artificial minds
see more hype than
image source: Shutt

News • Experts ex

AI outpe
fact?

Many studies claiming that artificial intelligence (AI) is as good as (or better than) human experts at interpreting medical images are of poor quality and are arguably exaggerated, posing a risk for the safety of

ARTIFICIAL INTELLIGENCE

Google's medical AI was super accurate in a lab. Real life was a different story.

If AI is really going to make a difference to
works when real humans get their hands o

By W

Editorial | Open A

Crossing
impact: th
evaluation

Jayson S. Marwah

npj Digital Medici

1890 Accesses

Artificial intellig
potential to imp
in clinical settings remains largely unknown. In a recent systematic review by Zhou et al.1, the

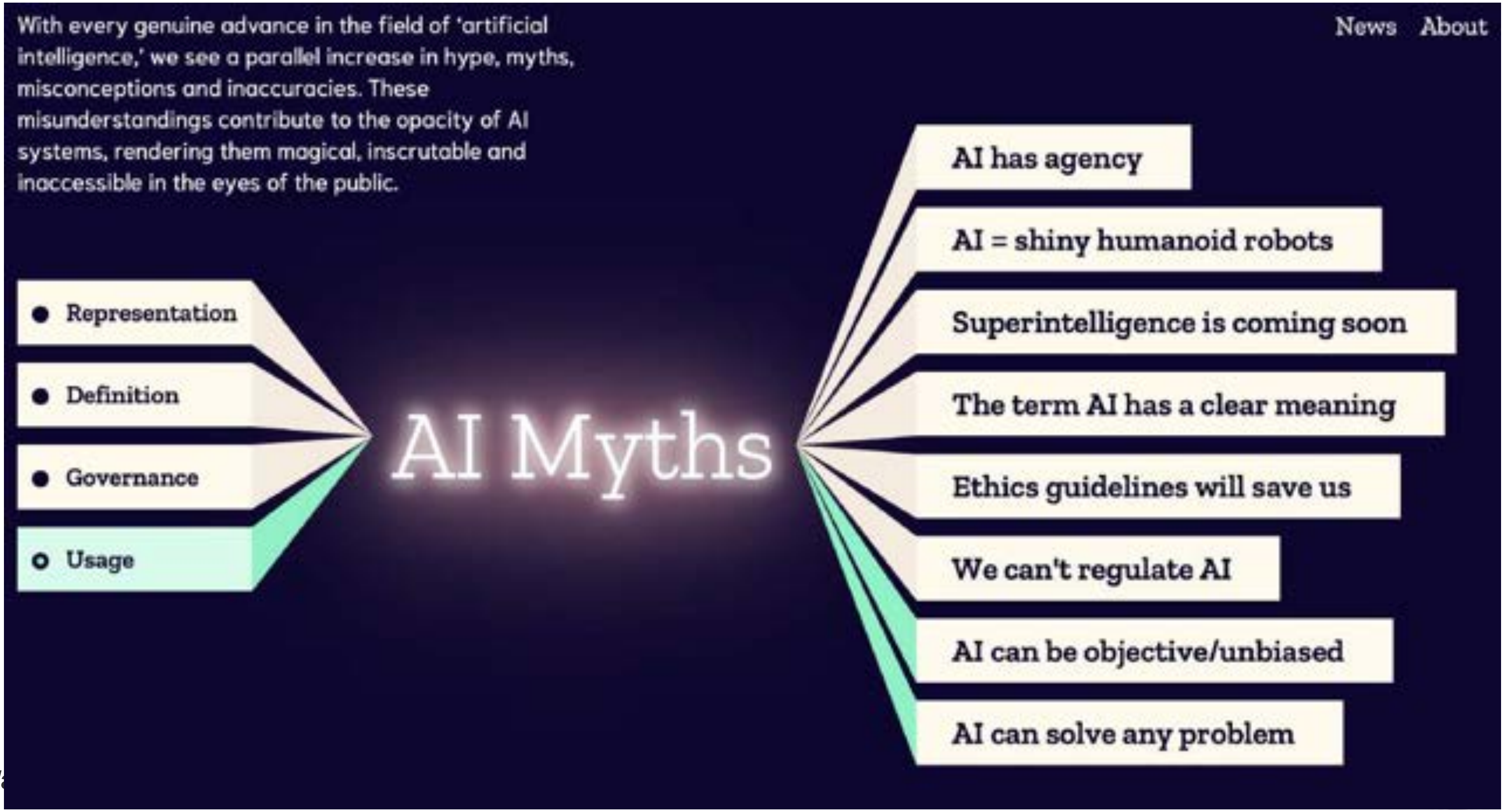
AI Can Outperform Doctors. So Why Don't Patients Trust It?

by Chiara Longoni and Carey K. Morewedge

October 30, 2019

Realistic expectations

- https://www.aimyths.org/

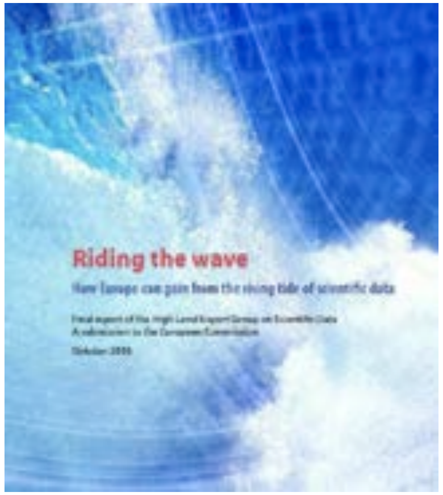


With every genuine advance in the field of 'artificial intelligence,' we see a parallel increase in hype, myths, misconceptions and inaccuracies. These misunderstandings contribute to the opacity of AI systems, rendering them magical, inscrutable and inaccessible in the eyes of the public.

News
About
AI Myths
● Representation
● Definition
● Governance
○ Usage
AI has agency
AI = shiny humanoid robots
Superintelligence is coming soon
The term AI has a clear meaning
Ethics guidelines will save us
We can't regulate AI
AI can be objective/unbiased
AI can solve any problem

Advantages of medical data

- Images created under standardized conditions
- Images are always attached to a case and a report describing it, plus a reason for producing the images
  - Metadata exist, and other data on the same patient
    - We know the context of the images
- Much medical knowledge is available
  - Coded and maintained in ontologies
- Much clinical research is done
  - Medical imaging is estimated to occupy 30% of world storage



Riding the wave

How Europe can gain from the rising tide of scientific data

Final report of the High Level Expert Group on Scientific Data
A submission to the European Commission

October 2010

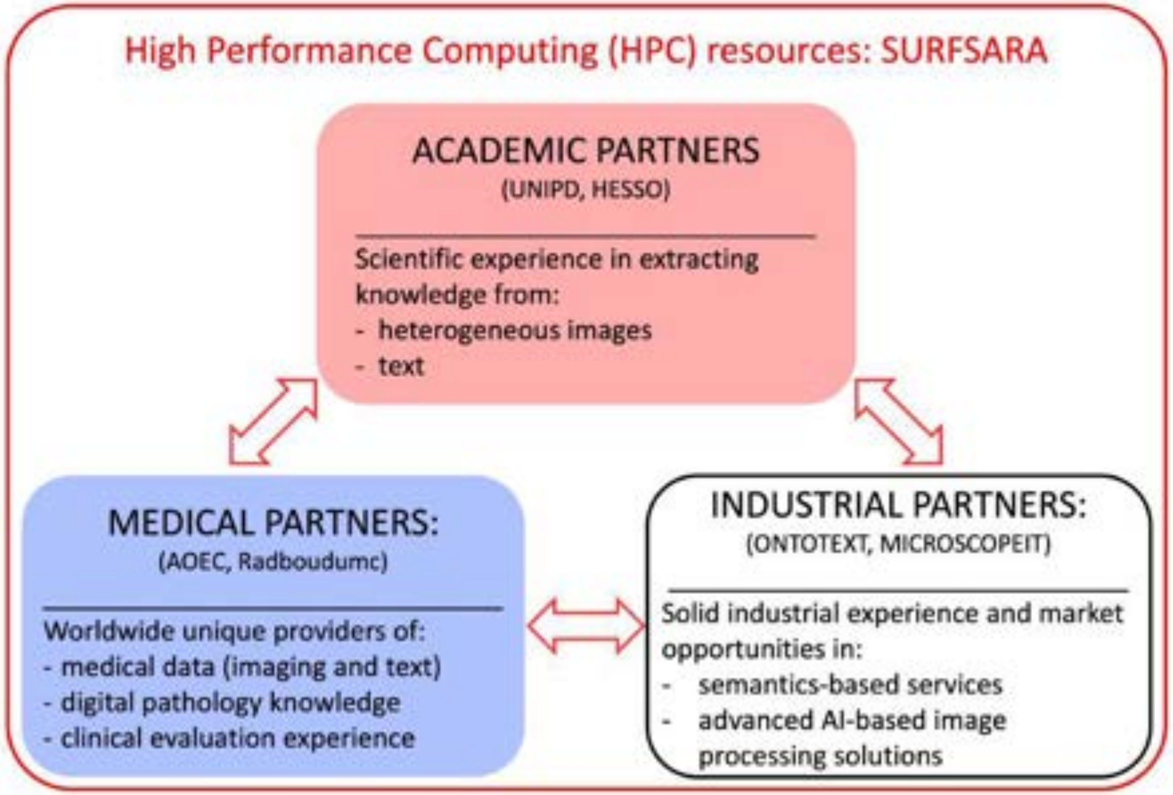
Challenges with medical data

- Data privacy and ethics can make sharing data hard
- Medical equipment and procedures vary across hospitals
  - And equipment changes frequently
- Pixel-level annotations are extremely expensive
  - Specialists are needed and often not available (too busy)
- An image is only a very small part in a case with patient history, temporal data, genetics, text, structured data etc.
  - Regions determining a decision are often extremely small
    - Needle in a haystack

Challenges for medical AI

- Much data are needed, so solutions need to be scalable
  - Diversity is required for generalization
  - Data sets are very unbalanced
- Continuous learning is required due to changing equipment and clinical guidelines (half-life of knowledge)
- Pixel level annotations are not available, as expensive
- Combining multiple sources is needed for proper learning
- Results need to be explainable for workflow integration
  - Deep learning is a priori a black box

Examode consortium



High Performance Computing (HPC) resources: SURFSARA

ACADEMIC PARTNERS
(UNIPD, HESSO)

Scientific experience in extracting knowledge from:
- heterogeneous images
- text

MEDICAL PARTNERS:
(AOEC, Radboudumc)

Worldwide unique providers of:
- medical data (imaging and text)
- digital pathology knowledge
- clinical evaluation experience

INDUSTRIAL PARTNERS:
(ONTOTEXT, MICROSCOPEIT)

Solid industrial experience and market opportunities in:
- semantics-based services
- advanced AI-based image processing solutions

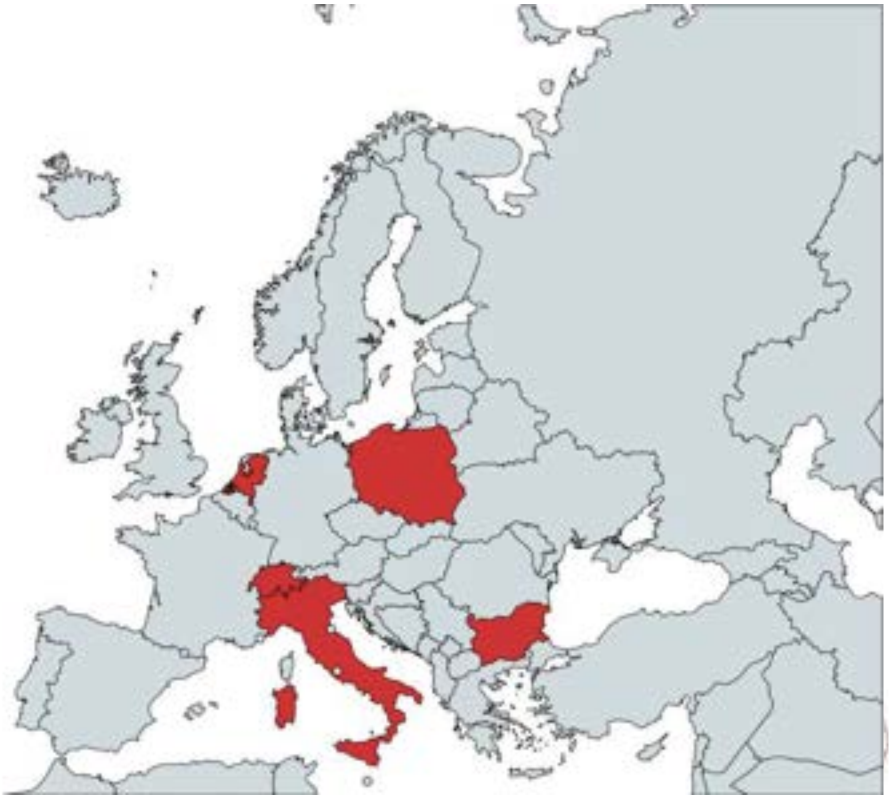
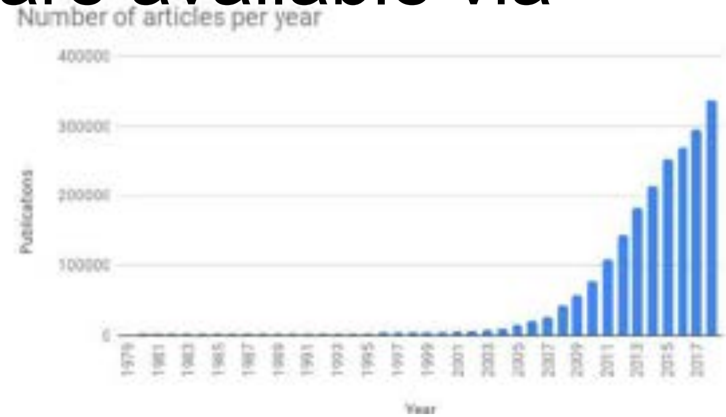


Image accessibility

- Open data policies of funding agencies make large medical data sets available
  - Particularly NIH is pushing towards this
- TCIA and TCGA are very large repositories
  - There are many scientific challenges
- Images from the Biomedical literature are available via PubMedCentral
  - Exponentially increasing
  - Extremely varied, hard to use

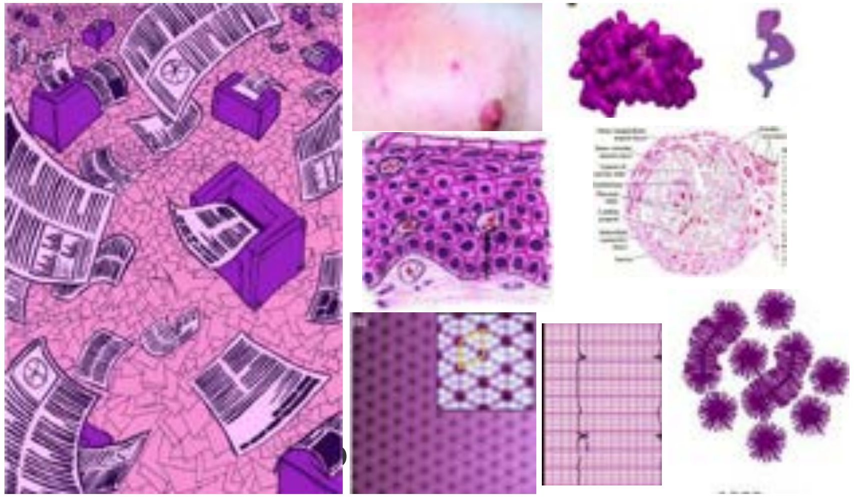


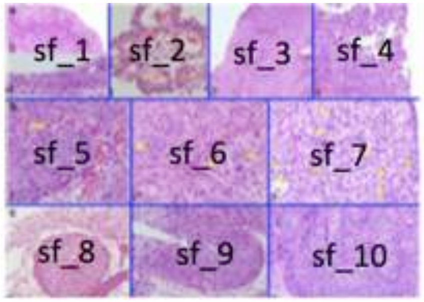
Unbalanced data sets

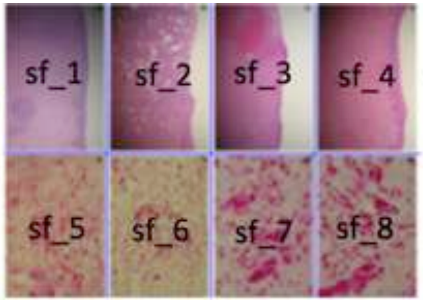
- Differing frequencies of the relevant classes need to be taken into account
  - At cancer screen even high-risk people are ~1% positive
  - Sensitivity and specificity as measures, not accuracy
    - Weight between false positives and false negatives varies
  - Some cases may occur once/twice per year in large hospitals
- Rare cases is what is more commonly described in articles
  - Images from articles can thus help (at least in theory)
  - Variety of imaging parameters and laboratories is very high

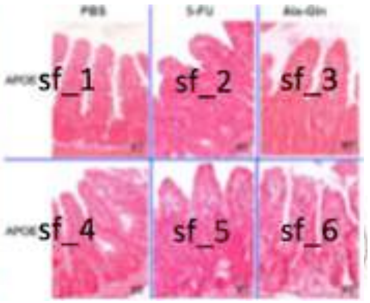
Challenges with PubMed

- >20’000’000 images in 2022, many graphs, charts
- Look-alikes is a problem, and compound figures
  - Very varied and sometimes strange content needs removal
- Compound figures need to be separated
  - Cutting sub figures apart makes content accessible









Making the images usable

- Removing very small images & strange aspect ratios
- Classify figures into figure types
  - Using image data and also text, remove non-relevant images
- Detect and cut compound figures into their parts
  - Classify these into figure types again
- Filter human vs. animal tissue and specific organs
- Check diseases or grading/staging images
  - Classes for machine learning

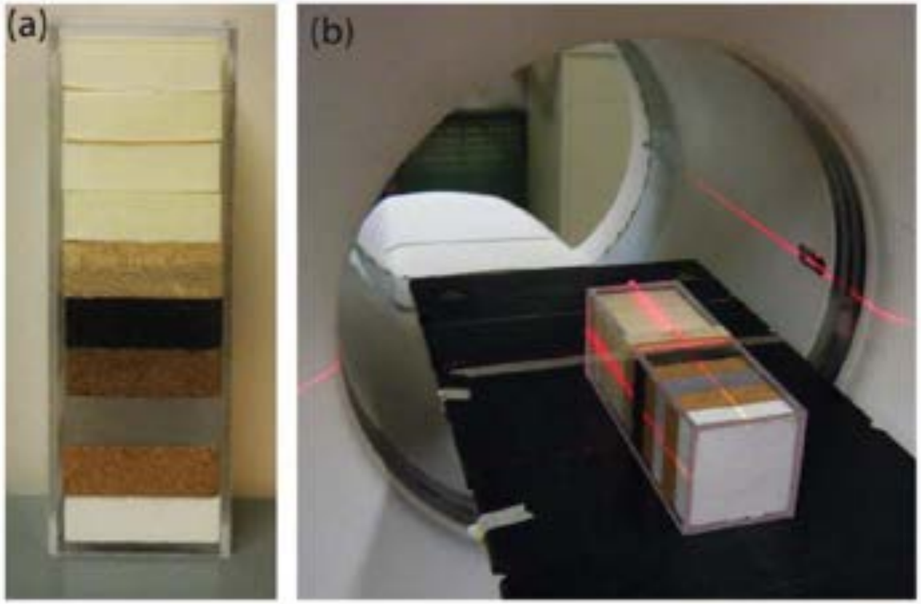
Advantages of literature images

- Rare images (unusual, untypical) are generally used for articles and case descriptions
  - A few typical cases but mainly extreme cases
  - Creates critical mass for rare diseases
- Images are from many laboratories and thus contain many image variations (staining, scanners)
  - Increase generalizability of learned models thanks to this diversity
- Exponentially increasing content

Image harmonization for radiomics

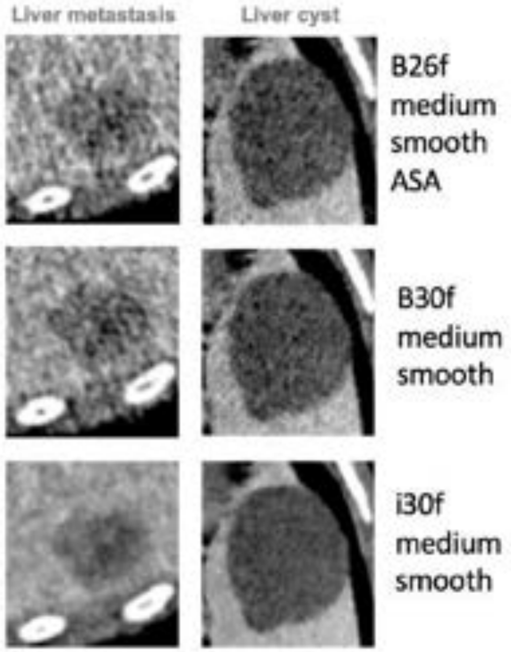
- Different scanners produce different images
  - Many protocols, construction kernels, producers, voxel sizes, …
    - Strong influence on features extracted
- How can we harmonize this?
  - Deep learning!
- Phantom study with 17 scanners
  - 10 solid textures
  - Features invariant to scanner

Vincent Andrearczyk, Adrien Depeursinge, and Henning Müller, Neural Network Training for Cross-Protocol Radiomic Feature Standardization in Computed Tomography, Journal of Medical Imaging, 2019.



Measuring CT variability

- Many CT parameter variations stemming from:
  - Acquisition protocols (radiation dose, …)
  - Image reconstruction parameters
  - Image resolution (slice thickness, overlap, …)
- Variability has a strong influence on the analysis & comparison of radiomics features
- Patient studies evaluating image/feature stability entail ethical concerns with multiple exposures to radiation

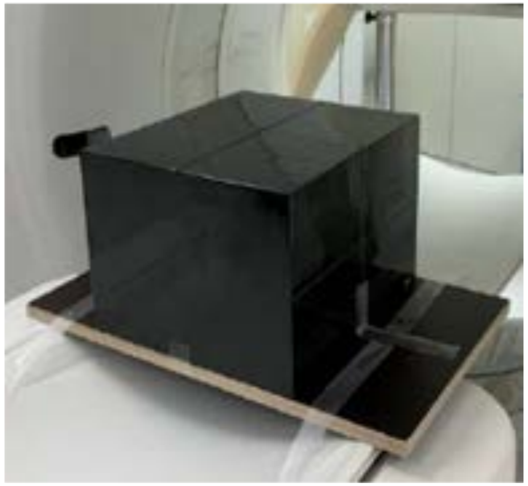


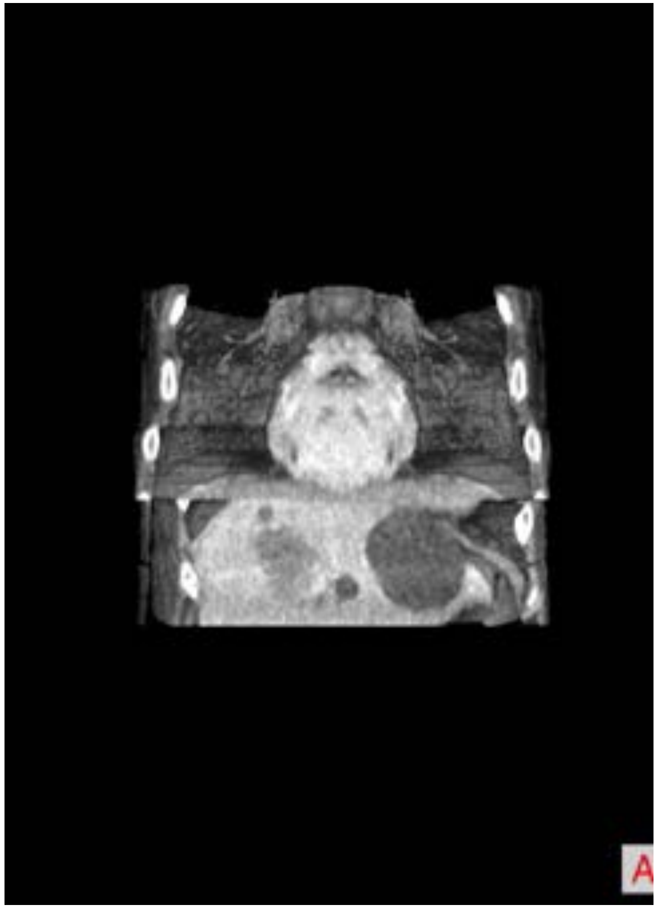
Traverso et al. (2018) *Repeatability and reproducibility of radiomic features: a systematic review.* International Journal of Radiation Oncology Physics **102**.

Solomon et al. (2014) Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. Medical Physics 41, 091908.

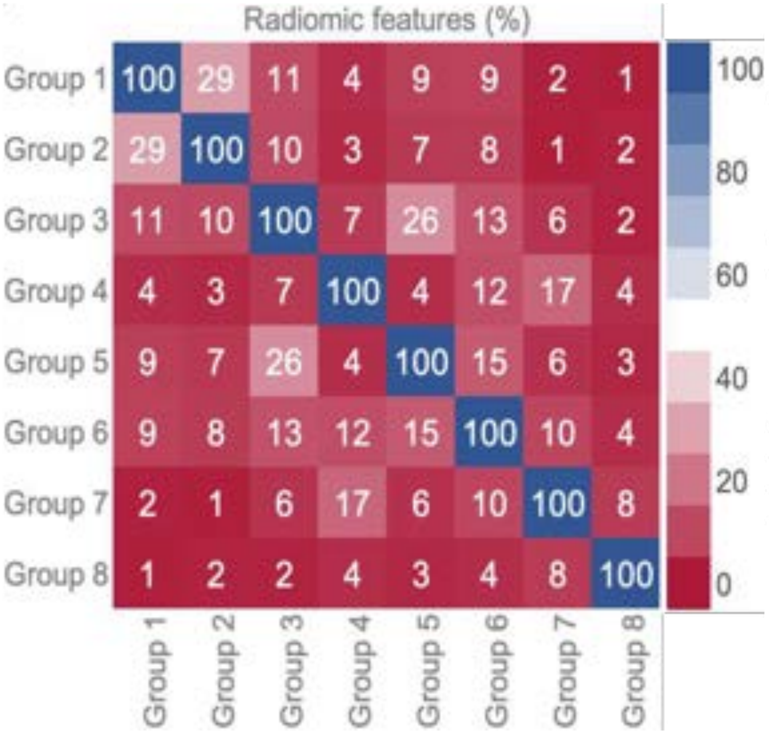
3D printed phantom

- Phantoms allow repeated radiation exposure
- Highly controlled acquisitions
  - No patient movement
  - No breathing
  - Precise positioning
- Limitations
  - In density (-100 HU to1000 HU)
  - Small blocks are glued (artifacts)



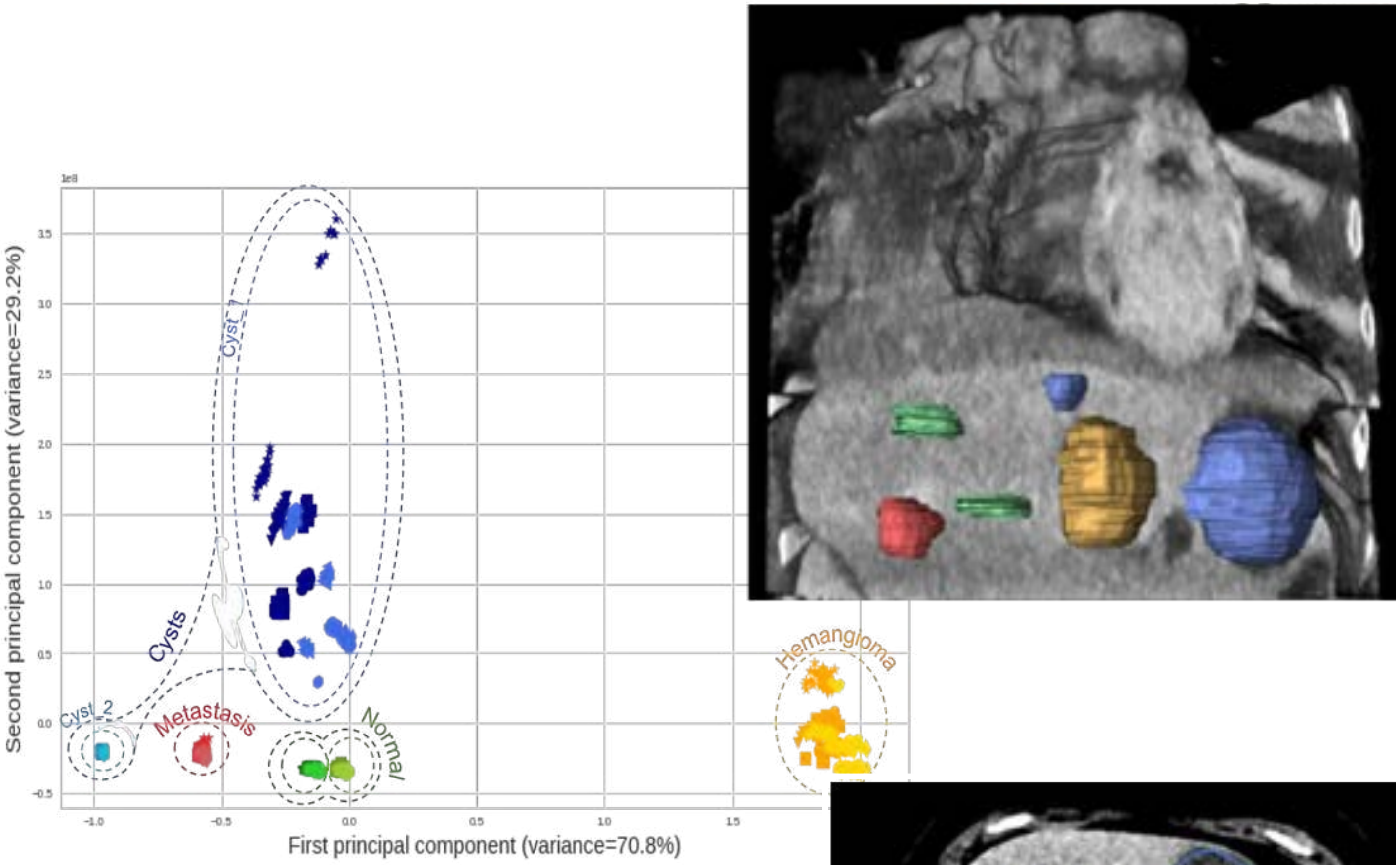


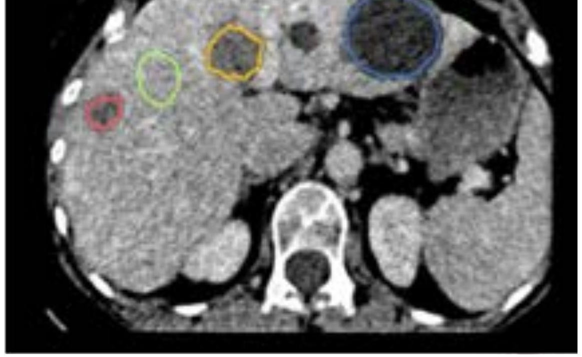
First results



Radiomic features (%)

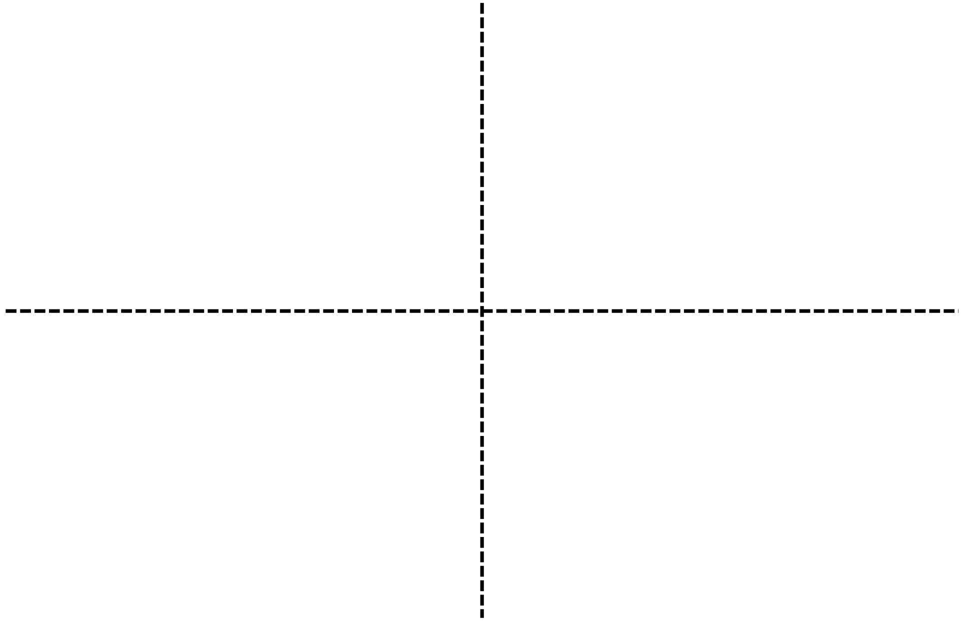
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
Group 1	100	29	11	4	9	9	2	1
Group 2	29	100	10	3	7	8	1	2
Group 3	11	10	100	7	26	13	6	2
Group 4	4	3	7	100	4	12	17	4
Group 5	9	7	26	4	100	15	6	3
Group 6	9	8	13	12	15	100	10	4
Group 7	2	1	6	17	6	10	100	8
Group 8	1	2	2	4	3	4	8	100





Oscar Jimenez-del-Toro, Christoph Aberle, Michael Bach, Roger Schaer, Markus Obmann, Kyriakos, Ender Konukoglu, Bram Stieltjes, Henning Müller, Adrien Depeursinge, The discriminative power and reproducibility of radiomics features with CT variations: Task-based analysis in a realistic CT liver phantom, Investigative Radiology, 2021.

Stability vs. discriminative power



!"#$%$&'('%")$%*$(*+,-"

🔵 !"#$%&' (&)* +,-./0$1230*4

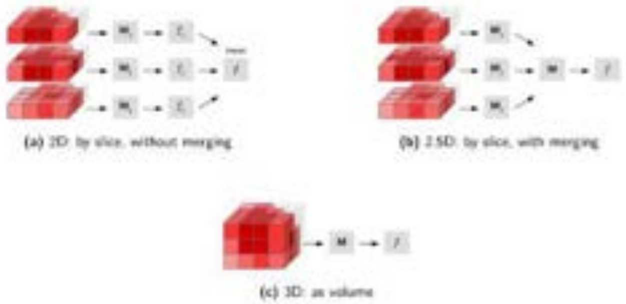
🔴 #&0*5 678.97:.+8,./0$1230*4

🟡 %00!;0$3<+:=./0$1230*4

IBSI
image biomarker standardisation initiative

- Image Biomarker Standardization Initiative
  - Define all visual features used in radiomics
    - And compare implementations on the same data

      Digital phantoms
    - When there are differences then check the implementations
  - Installment 1 is finished,
    - Simple statistical (texture) features
  - Installment 2 is under way
    - Filter banks (Wavelets, Gabor, …)
- All information is available at: https://theibsi.github.io

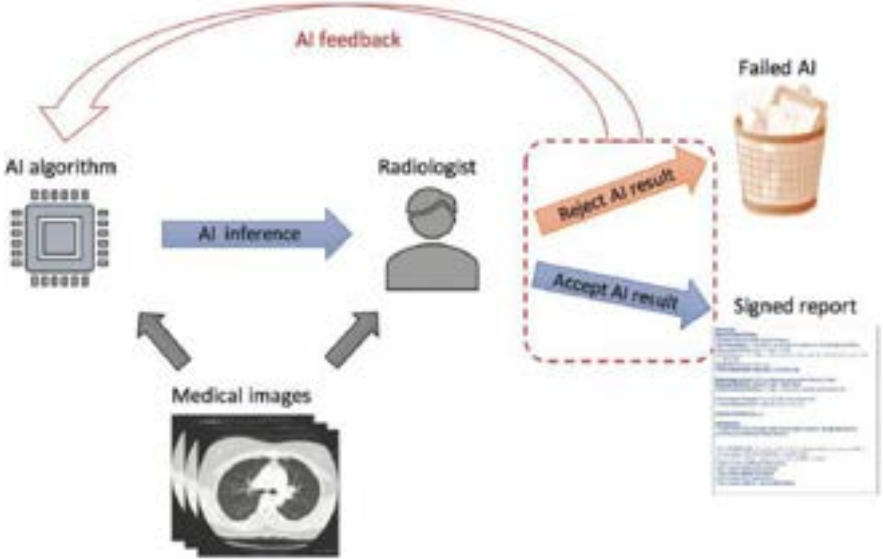


(a) 2D: by slice, without merging
(b) 2.5D: by slice, with merging
(c) 3D: as volume

Zwanenburg, A., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, *295*(2), 328-338.

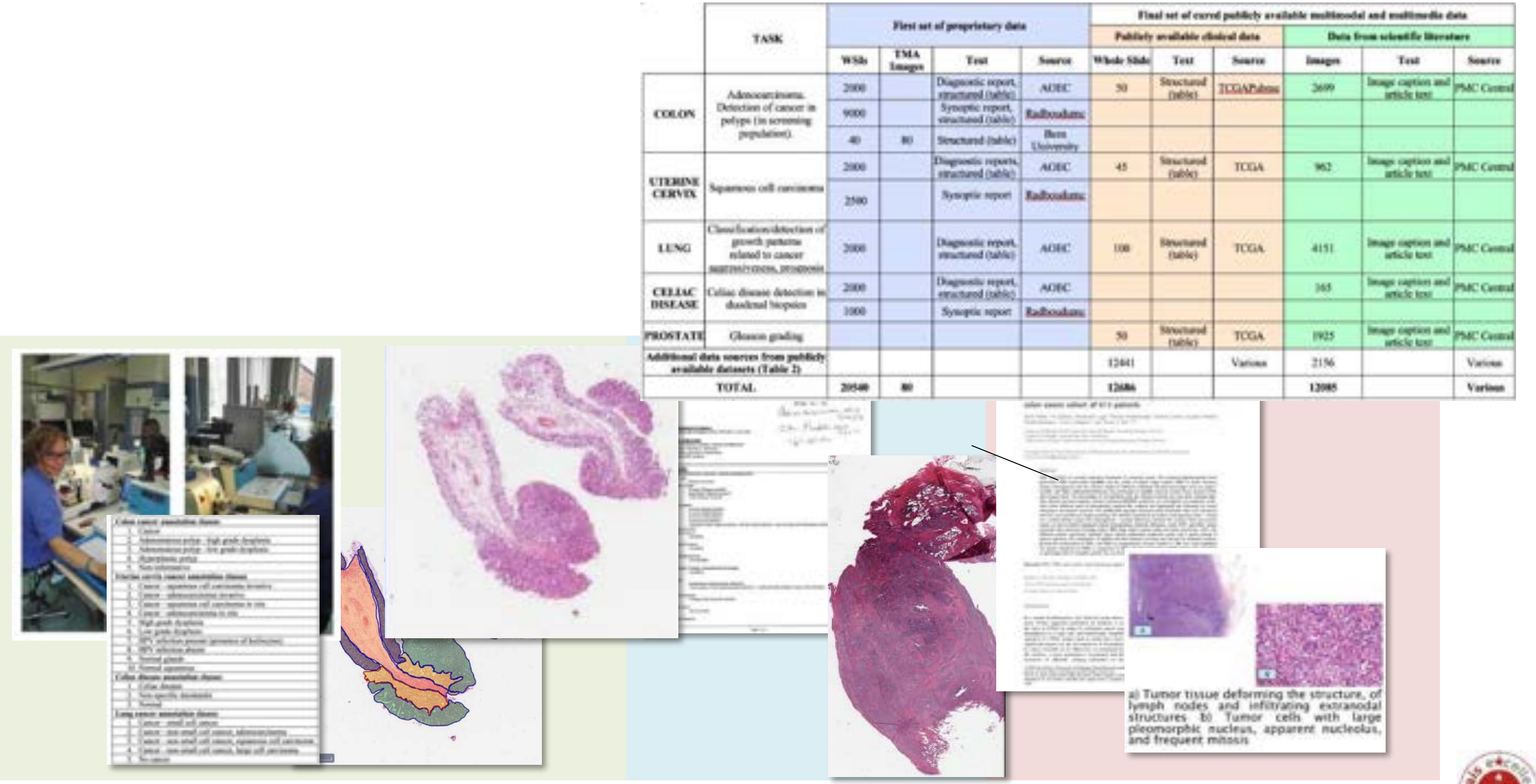
Continuous learning

- Add new annotated samples regularly to update algorithms (for example with new machines)
  - Avoid catastrophic forgetting
    - By adapting to a few specific cases or outliers
- Regular feedback loop
  - With clinicians using AI



Pianykh, O.S., Langs, G., Dewey, M., Enzmann, D.R., Herold, C.J., Schoenberg, S.O. and Brink, J.A., 2020. Continuous learning AI in radiology: implementation principles and early applications. *Radiology*, *297*(1), pp.6-14.

Data used in ExaMode



	TASK	First set of proprietary data				Final set of cured publicly available multimodal and multimedia data					
						Publicly available clinical data			Data from scientific literature		
		WSIs	TMA Images	Text	Source	Whole Slide	Text	Source	Images	Text	Source
COLON	Adenocarcinoma. Detection of cancer in polyps (in screening population).	2000		Diagnostic report, structured (table)	AOEC	50	Structured (table)	TCGAPubmed	2699	Image caption and article text	PMC Central
		9000		Synoptic report, structured (table)	Radboudumc						
		40	80	Structured (table)	Bern University						
UTERINE CERVIX	Squamous cell carcinoma	2000		Diagnostic reports, structured (table)	AOEC	45	Structured (table)	TCGA	962	Image caption and article text	PMC Central
		2500		Synoptic report	Radboudumc						
LUNG	Classification/detection of growth patterns related to cancer aggressiveness, prognosis	2000		Diagnostic report, structured (table)	AOEC	108	Structured (table)	TCGA	4151	Image caption and article text	PMC Central
CELIAC DISEASE	Celiac disease detection in duodenal biopsies	2000		Diagnostic report, structured (table)	AOEC				165	Image caption and article text	PMC Central
		1000		Synoptic report	Radboudumc						
PROSTATE	Gleason grading					50	Structured (table)	TCGA	1925	Image caption and article text	PMC Central
Additional data sources from publicly available datasets (Table 2)						12441		Various	2156		Various
TOTAL		20540	80			12686			12085		Various

Colon cancer annotation classes:
1. Cancer
2. Adenomatous polyp - high grade dysplasia
3. Adenomatous polyp - low grade dysplasia
4. Hyperplastic polyp
5. Non-informative

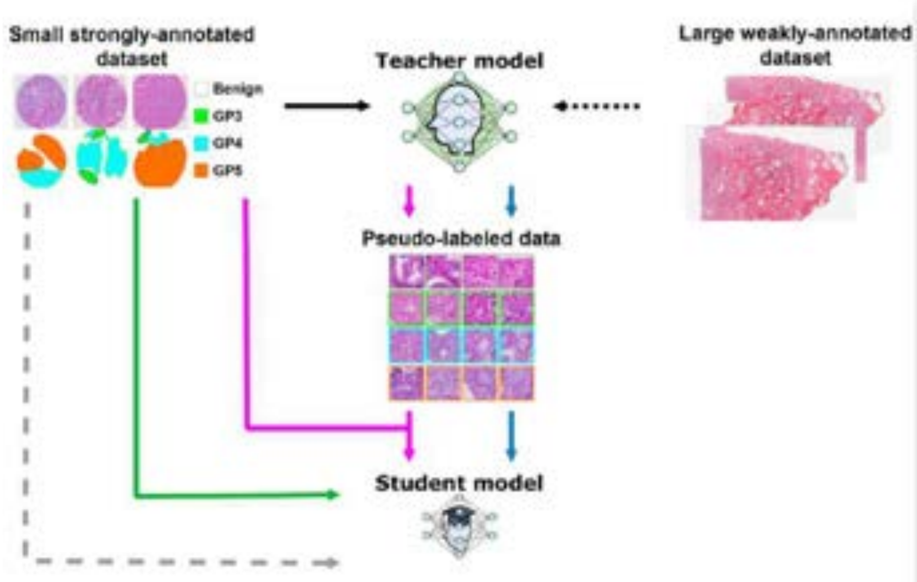
Uterine cervix cancer annotation classes:
1. Cancer - squamous cell carcinoma invasive
2. Cancer - adenocarcinoma invasive
3. Cancer - squamous cell carcinoma in situ
4. Cancer - adenocarcinoma in situ
5. High grade dysplasia
6. Low grade dysplasia
7. HPV infection present (presence of koilocytes)
8. HPV infection absent
9. Normal glands
10. Normal squamous

Celiac disease annotation classes:
1. Celiac disease
2. Non-specific duodenitis
3. Normal

Lung cancer annotation classes:
1. Cancer - small cell cancer
2. Cancer - non-small cell cancer, adenocarcinoma
3. Cancer - non-small cell cancer, squamous cell carcinoma
4. Cancer - non-small cell cancer, large cell carcinoma
5. No cancer

a) Tumor tissue deforming the structure, of lymph nodes and infiltrating extranodal structures b) Tumor cells with large pleomorphic nucleus, apparent nucleolus, and frequent mitosis

Weakly supervised learning



Small strongly-annotated dataset
Benign
GP3
GP4
GP5
Teacher model
Large weakly-annotated dataset
Pseudo-labeled data
Student model

Teacher/student paradigm approaches:

Semi-supervised learning

Semi-weakly supervised

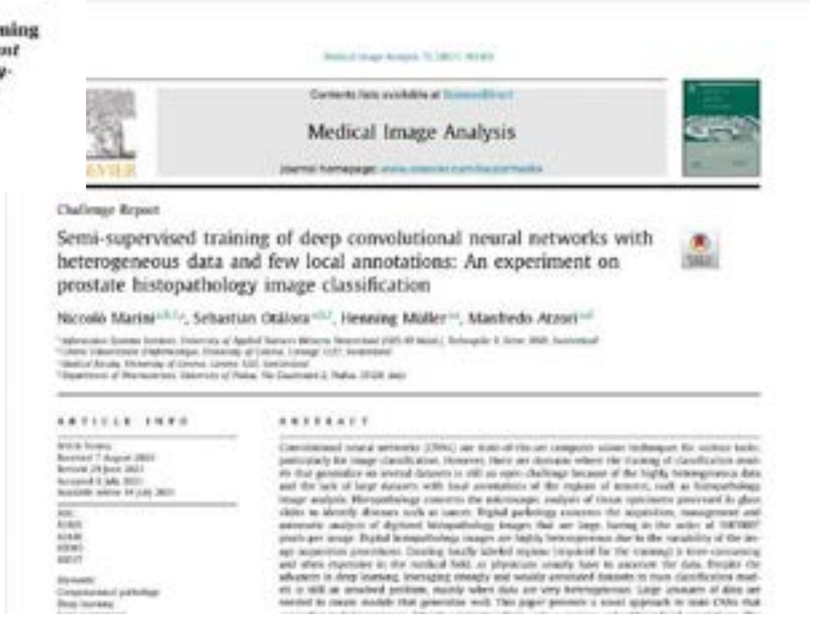
Student training variants:

Student training variant I *(training the student only with the pseudo-labeled data)* 🟦

Student training variant II *(pre-training the student with pseudo-labeled data and fine-tuning it with strongly-annotated data)* 🟩

Student training variant III *(training the student with pseudo-labeled data and strongly-annotated data)* 🟪

Fully-supervised learning *(training the student only with strongly-annotated data)* ▨



Medical Image Analysis 73 (2021) 102165

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Challenge Report

Semi-supervised training of deep convolutional neural networks with heterogeneous data and few local annotations: An experiment on prostate histopathology image classification

Niccolò Marini[a,b,1,*], Sebastian Otálora[a,b,1], Henning Müller[a,c], Manfredo Atzori[a,d]

[a] Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO Valais), Technopôle 3, Sierre 3960, Switzerland
[b] Centre Universitaire d'Informatique, University of Geneva, Carouge 1227, Switzerland
[c] Medical faculty, University of Geneva, Geneva 1211, Switzerland
[d] Department of Neurosciences, University of Padua, Via Giustiniani 2, Padua 35128, Italy

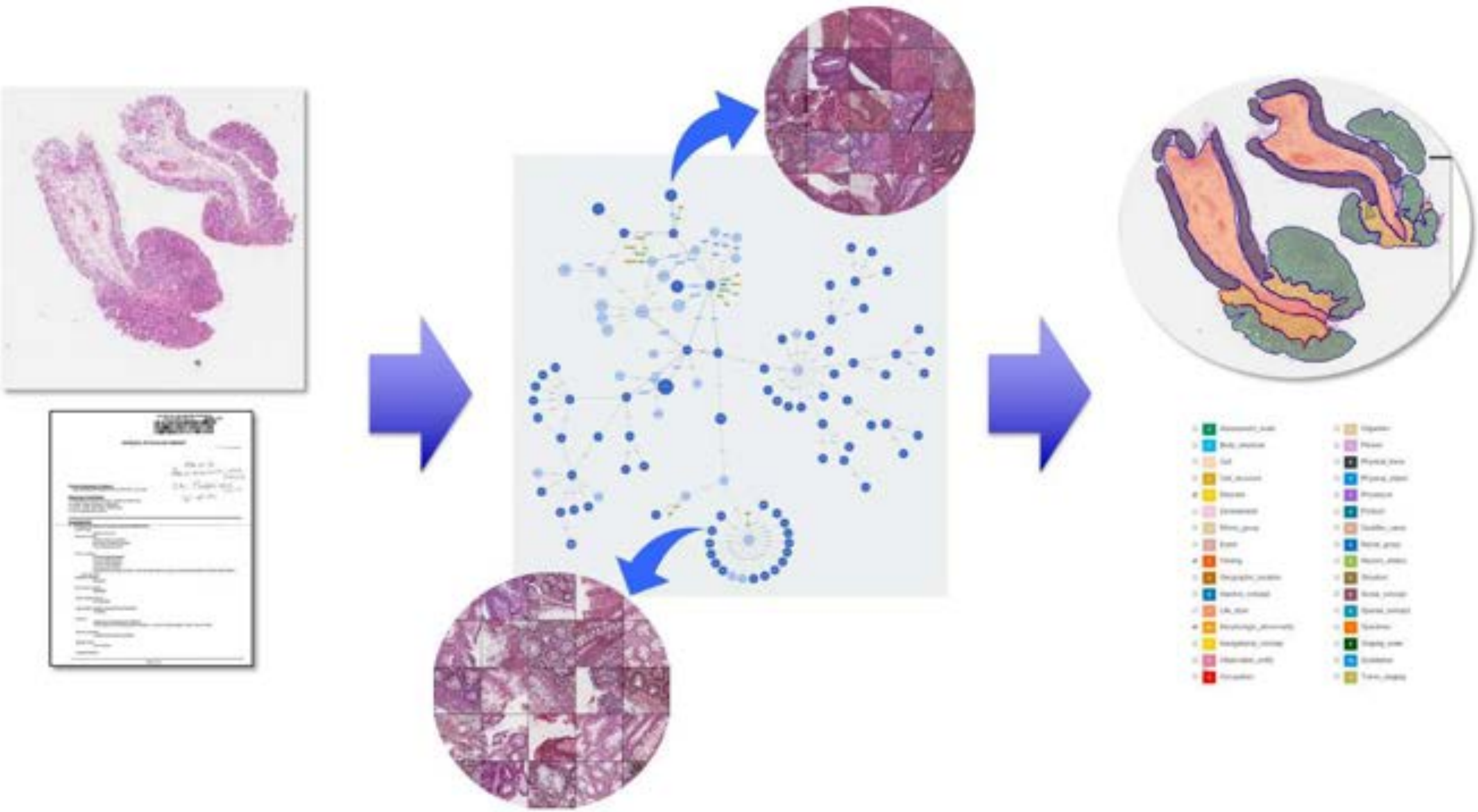
ARTICLE INFO

Article history:
Received 7 August 2020
Revised 29 June 2021
Accepted 5 July 2021
Available online 14 July 2021

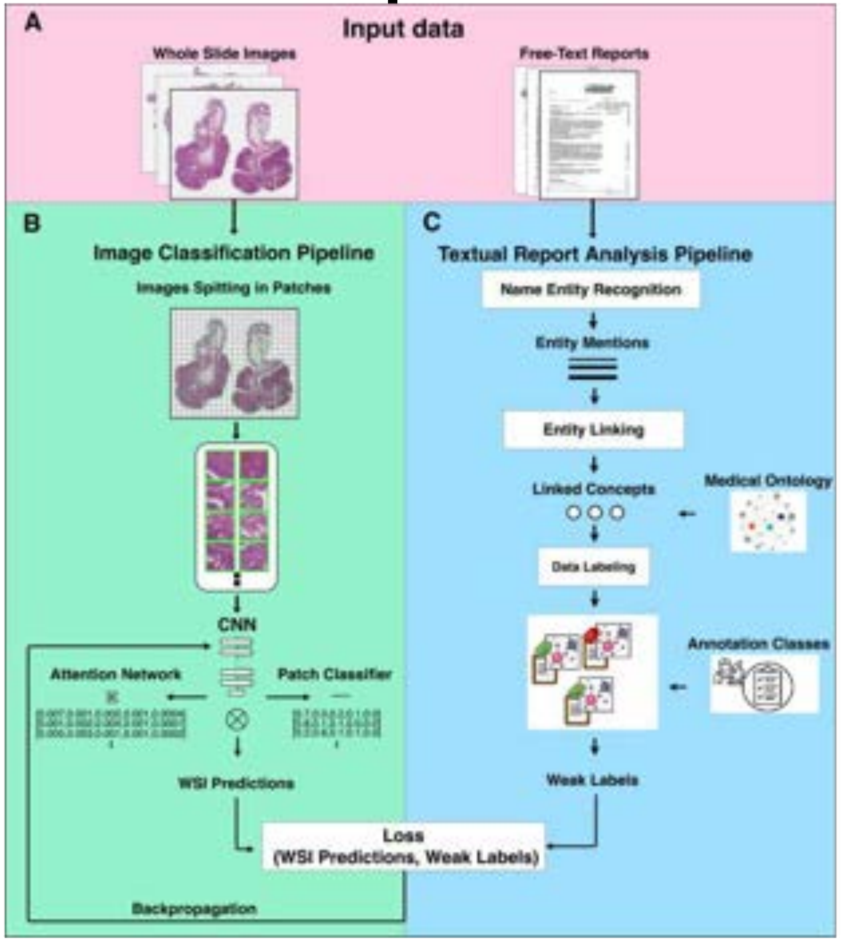
ABSTRACT

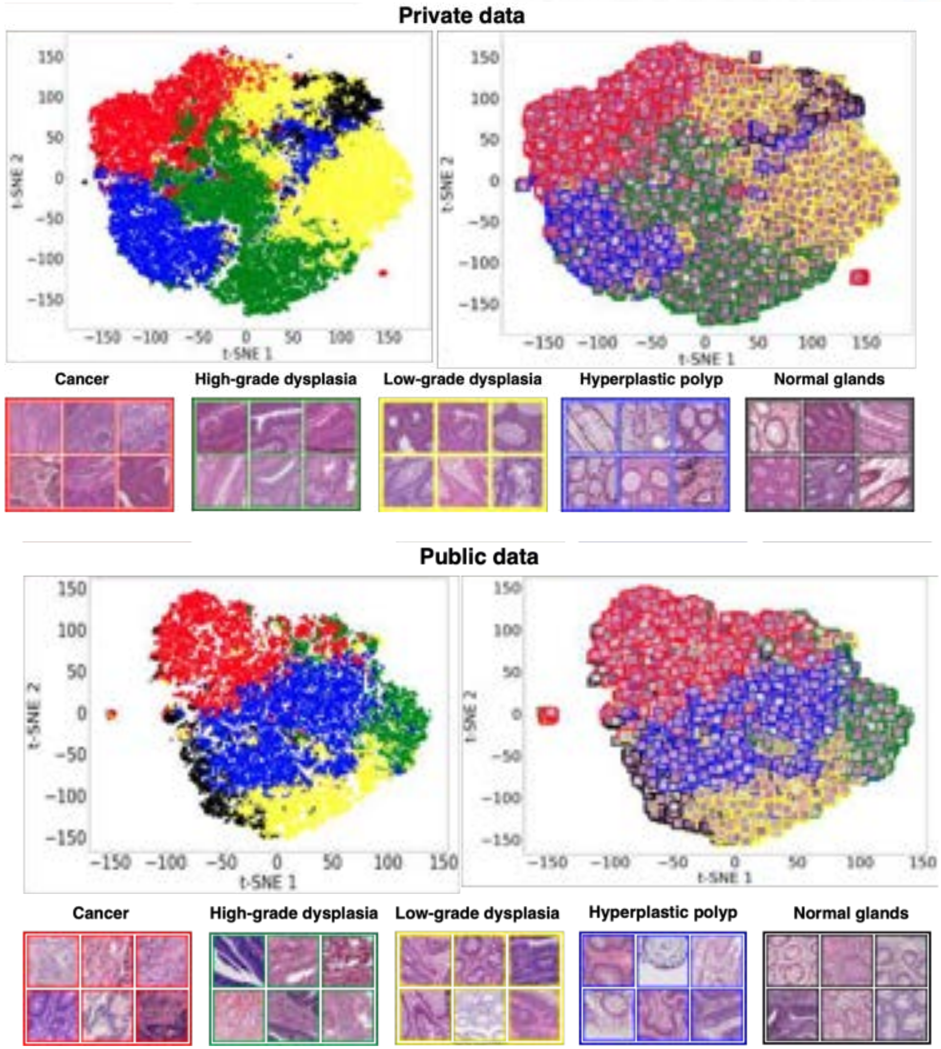
Convolutional neural networks (CNNs) are state-of-the-art computer vision techniques for various tasks, particularly for image classification. However, there are domains where the training of classification models that generalize on several datasets is still an open challenge because of the highly heterogeneous data and the lack of large datasets with local annotations of the regions of interest, such as histopathology image analysis. Histopathology concerns the microscopic analysis of tissue specimens processed in glass slides to identify diseases such as cancer. Digital pathology concerns the acquisition, management and automatic analysis of digitized histopathology images that are large, having in the order of 100'000² pixels per image. Digital histopathology images are highly heterogeneous due to the variability of the image acquisition procedures. Creating locally labeled regions (required for the training) is time-consuming and often expensive in the medical field, as physicians usually have to annotate the data. Despite the advances in deep learning, leveraging strongly and weakly annotated datasets to train classification models is still an unsolved problem, mainly when data are very heterogeneous. Large amounts of data are needed to create models that generalize well. This paper presents a novel approach to train CNNs that

Project status

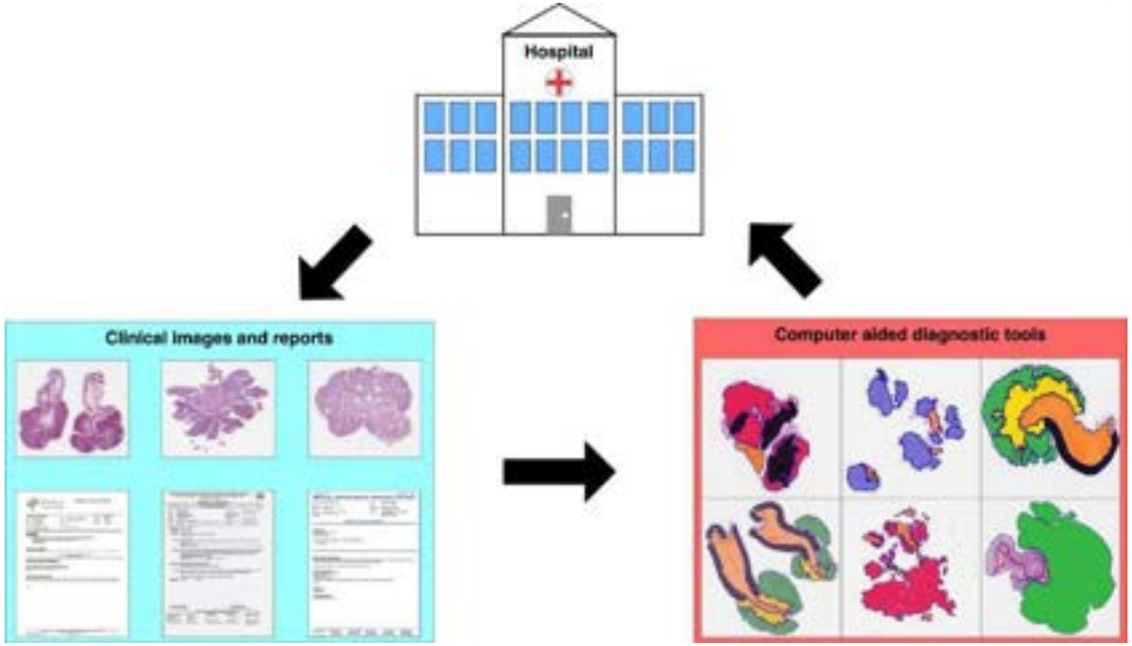


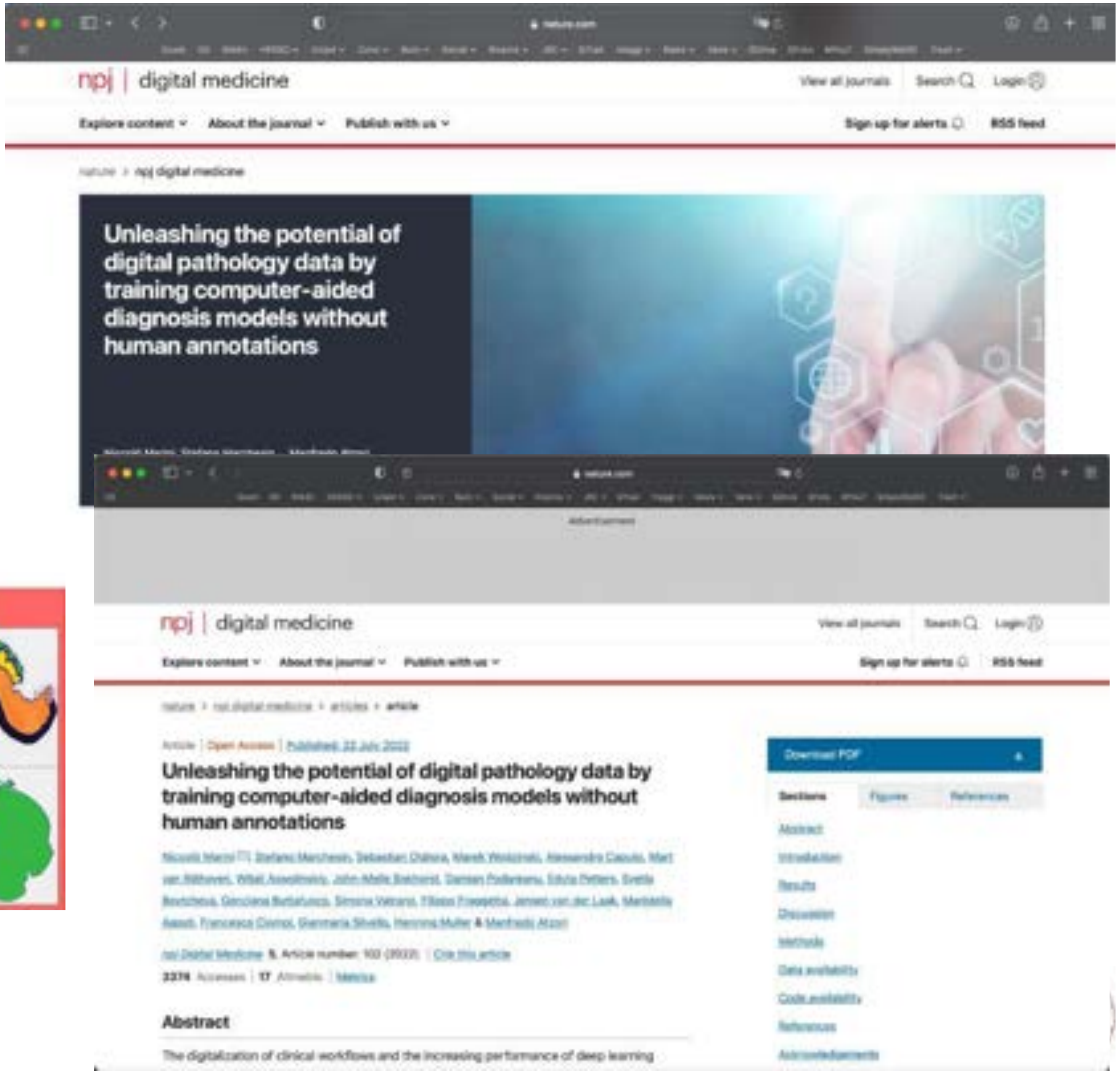
Weakly supervised learning from reports





First results on multimodal data





Unleashing the potential of digital pathology data by training computer-aided diagnosis models without human annotations

Article | Open Access | Published: 22 July 2022

Unleashing the potential of digital pathology data by training computer-aided diagnosis models without human annotations

Niccolò Marini, Stefano Marchesin, Sebastian Otálora, Marek Wodzinski, Alessandro Caputo, Mart van Rijthoven, Witali Aswolinskiy, John-Melle Bokhorst, Damian Podareanu, Edyta Petters, Svetla Boytcheva, Genziana Buttafuoco, Simona Vatrano, Filippo Fraggetta, Jeroen van der Laak, Maristella Agosti, Francesco Ciompi, Gianmaria Silvello, Henning Muller & Manfredo Atzori

npj Digital Medicine 5, Article number: 102 (2022) | Cite this article

3376 Accesses | 17 Altmetric | Metrics

Abstract

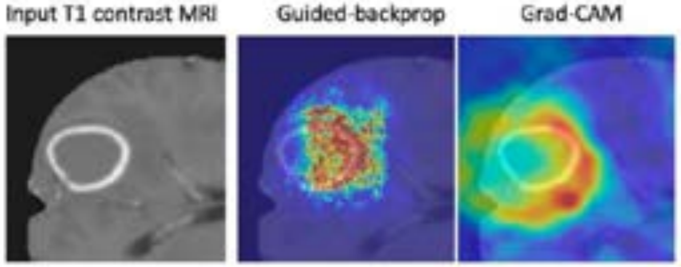
The digitalization of clinical workflows and the increasing performance of deep learning

Clinical workflow and AI

- A clinician orders an image
- A radiologist/pathologists produces and views the image and writes a report based on the question and anamnesis
  - Much data on the patient (environment, prior diseases, genetics, blood tests, development of a condition, …)
  - Differential diagnosis, under much time pressure
- Any AI needs to be integrated into the workflow and tools
  - Adding evidence, identifying bias, uncertainty, …
    - Explaining the decisions and their context

Interpretability of Deep Learning

- Make decisions understandable & remove black box image
- Make sure that decisions are sound
- Explain why things may not be working



- In medicine it is particularly important to make sure that results can be explained & reproduced
  - High impact of wrong decisions
- There are many approaches interpretability
  - 2D projections, PCA, TSNE
  - Class activation maps, saliency, …

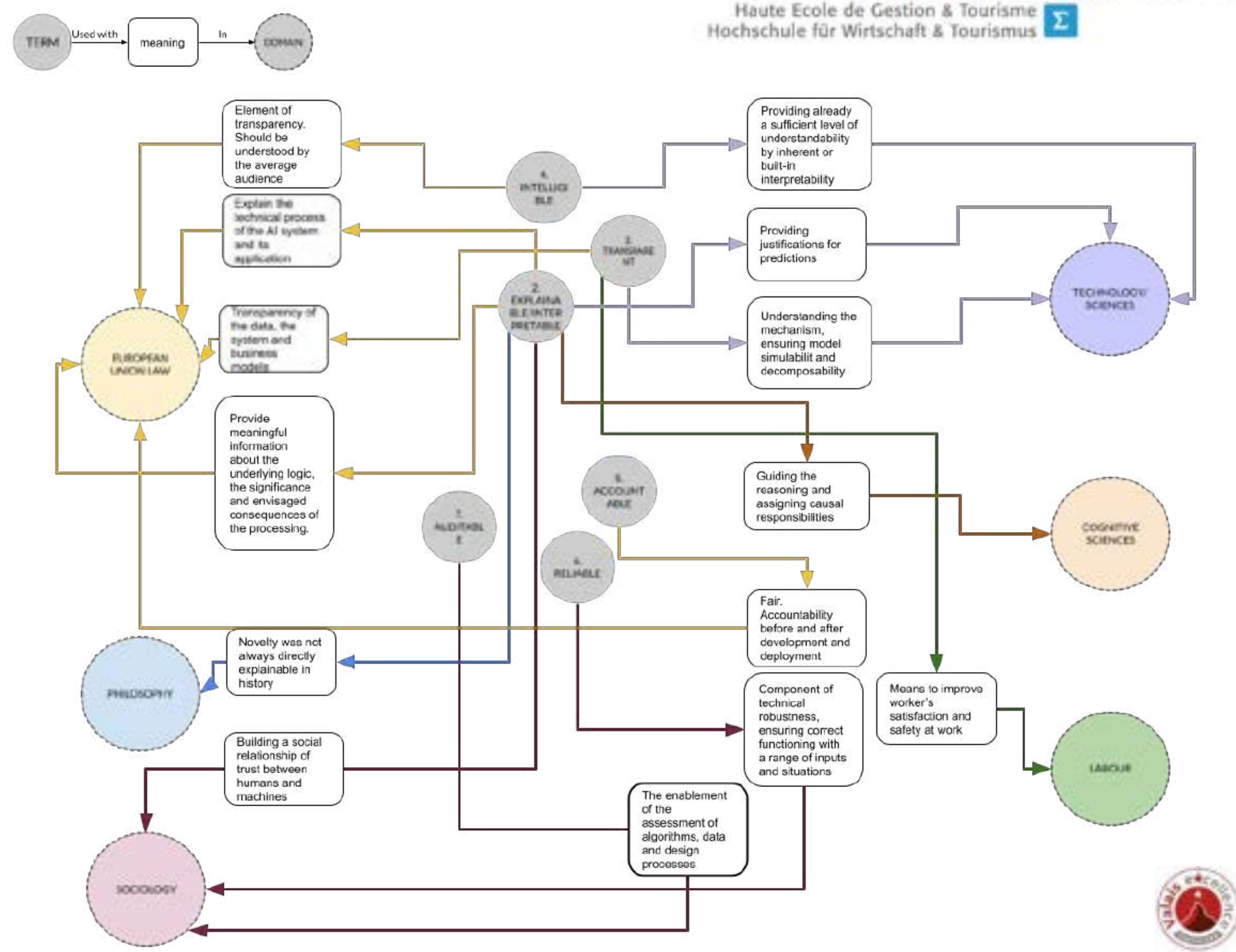
A taxonomy for explainability

- Many terms have been used in slightly different ways for AI: interpretability, explainability, transparency, accountability, fairness, (opacity) …
  - Bias, reliability, robustness, uncertainty, confidence
- A workshop was held in the summer of 2021 on this with views from several domains: legal, technical, philosophical, social, cognitive, ethical, …
  - https://taxonomyinterpretableai.wordpress.com
- EU is preparing the way
  - GDPR on data protection and AI policy
    - Limit the strong risks of AI and its use and abuse

M Graziani, L Dutkiewicz, D Calvaresi, J Pereira Amorim, K Yordanova, M Vered, R Nair, P Henriques Abreu, T Blanke, V Pulignano, JO. Prior, L Lauwaert, W Reijers, A Depeursinge, V Andrearczyk, H Müller, A Global Taxonomy of Interpretable AI: Unifying the Terminology for the Technical and the Social Sciences, Artificial Intelligence Reviews, 2022.

Taxonomy

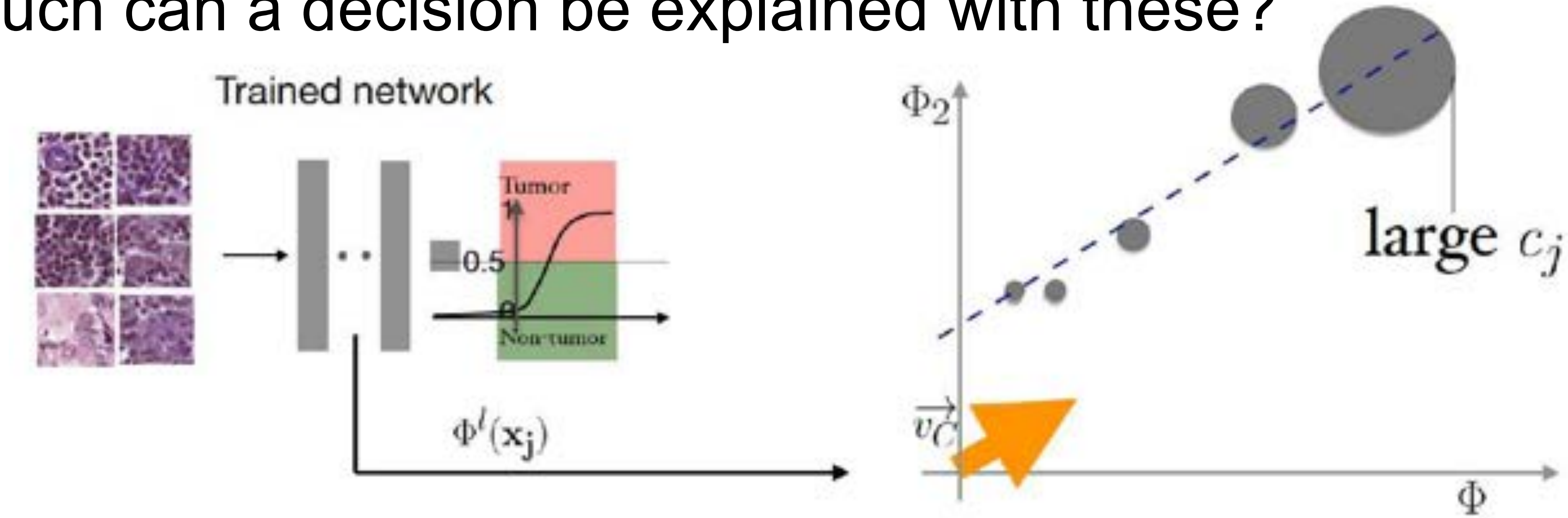
Interpretable AI terminology
Main terms and domains



TERM
Used with
meaning
In
DOMAIN
Element of transparency. Should be understood by the average audience
Explain the technical process of the AI system and its application
Transparency of the data, the system and business models
Provide meaningful information about the underlying logic, the significance and envisaged consequences of the processing.
4. INTELLIGIBLE
1. TRANSPARENT
2. EXPLAINABLE/INTERPRETABLE
Providing already a sufficient level of understandability by inherent or built-in interpretability
Providing justifications for predictions
Understanding the mechanism, ensuring model simulabilit and decomposability
TECHNOLOGY SCIENCES
EUROPEAN UNION LAW
Guiding the reasoning and assigning causal responsibilities
5. ACCOUNTABLE
7. AUDITABLE
6. RELIABLE
COGNITIVE SCIENCES
Fair. Accountability before and after development and deployment
Novelty was not always directly explainable in history
PHILOSOPHY
Component of technical robustness, ensuring correct functioning with a range of inputs and situations
Means to improve worker's satisfaction and safety at work
LABOUR
Building a social relationship of trust between humans and machines
The enablement of the assessment of algorithms, data and design processes
SOCIOLOGY

Regression concept vectors

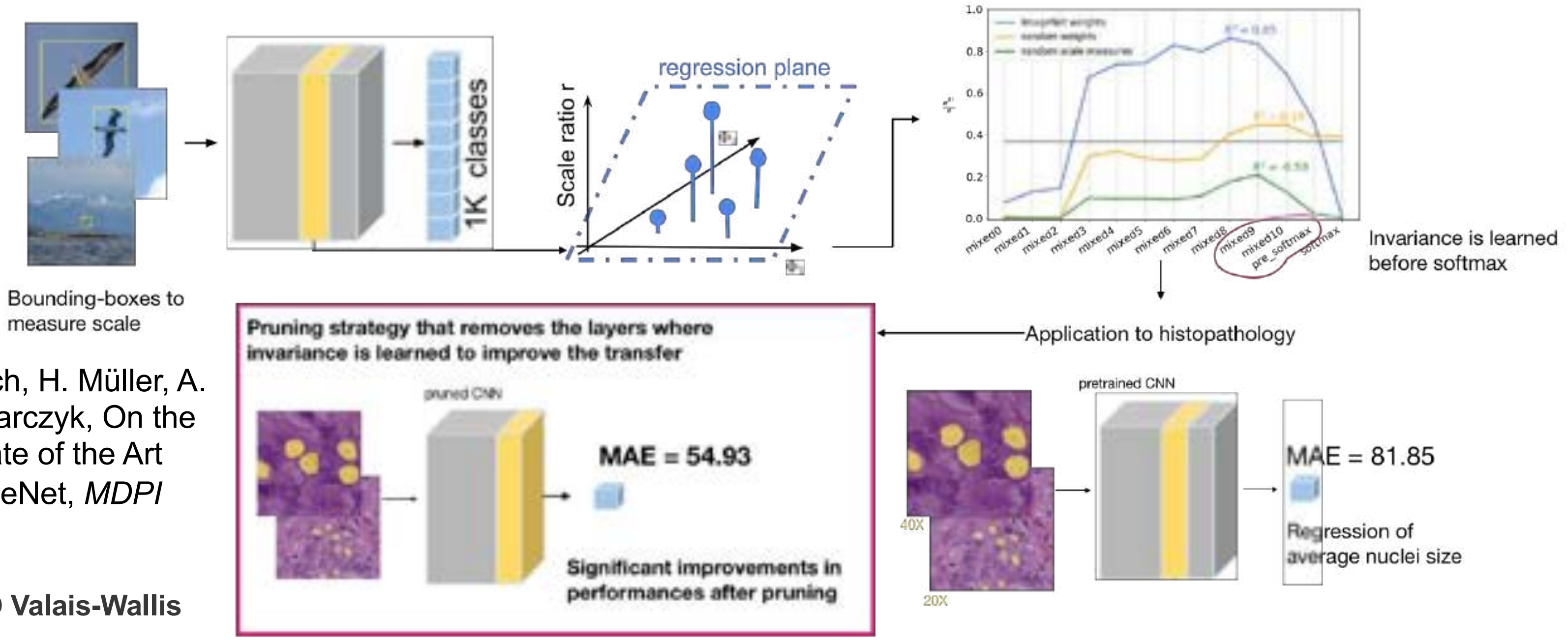
- Identify existing clinical features and check how the decision layers correlate to these features
  - i.e.: nuclei size, internal heterogeneity, borders, …
  - How much can a decision be explained with these?



M Graziani, V Andrearczik, H Müller, Concept attribution: Explaining CNN decisions to physicians, *Computers in Medicine and Biology*, 2020.

Improve with interpretability

- Pre-trained models often include scale invariance
- In medical applications this can be problematic, as scale carries information



Bounding-boxes to measure scale

Invariance is learned before softmax

Application to histopathology

Pruning strategy that removes the layers where invariance is learned to improve the transfer

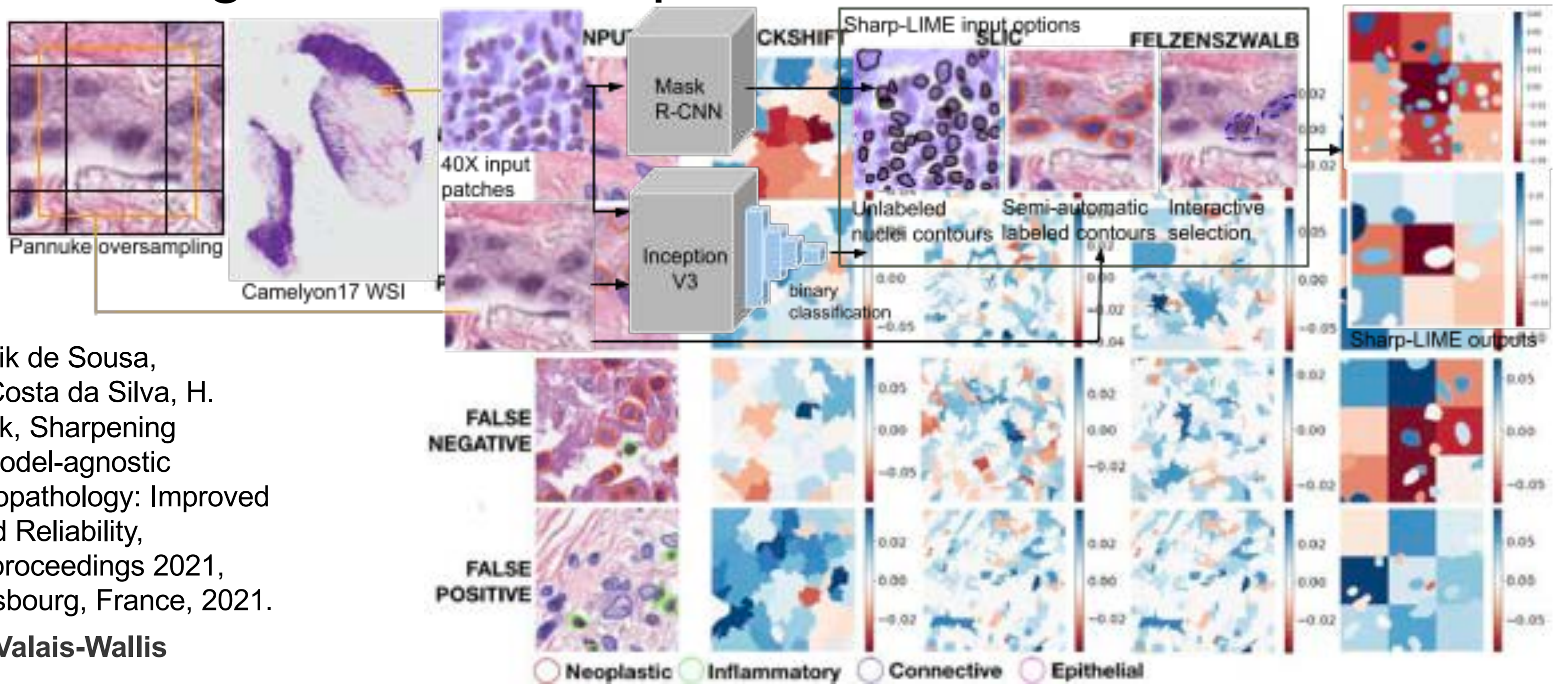
Significant improvements in performances after pruning

Regression of average nuclei size

M. Graziani, T. Lompech, H. Müller, A. Depeursinge, V. Andrearczyk, On the Scale Invariance in State of the Art CNNs Trained on ImageNet, *MDPI Make*, 2021.

Visualizations

- Improve visualizations of regions that are relevant for the decision of a DNN
  - LIME is commonly used to highlight regions, but interpretations can be difficult



M. Graziani, I. Palatnik de Sousa, M.B.R. Vellasco, E. Costa da Silva, H. Müller, V. Andrearczyk, Sharpening Local Interpretable Model-agnostic Explanations for Histopathology: Improved Understandability and Reliability, MICCAI conference proceedings 2021, Springer LNCS, Strasbourg, France, 2021.

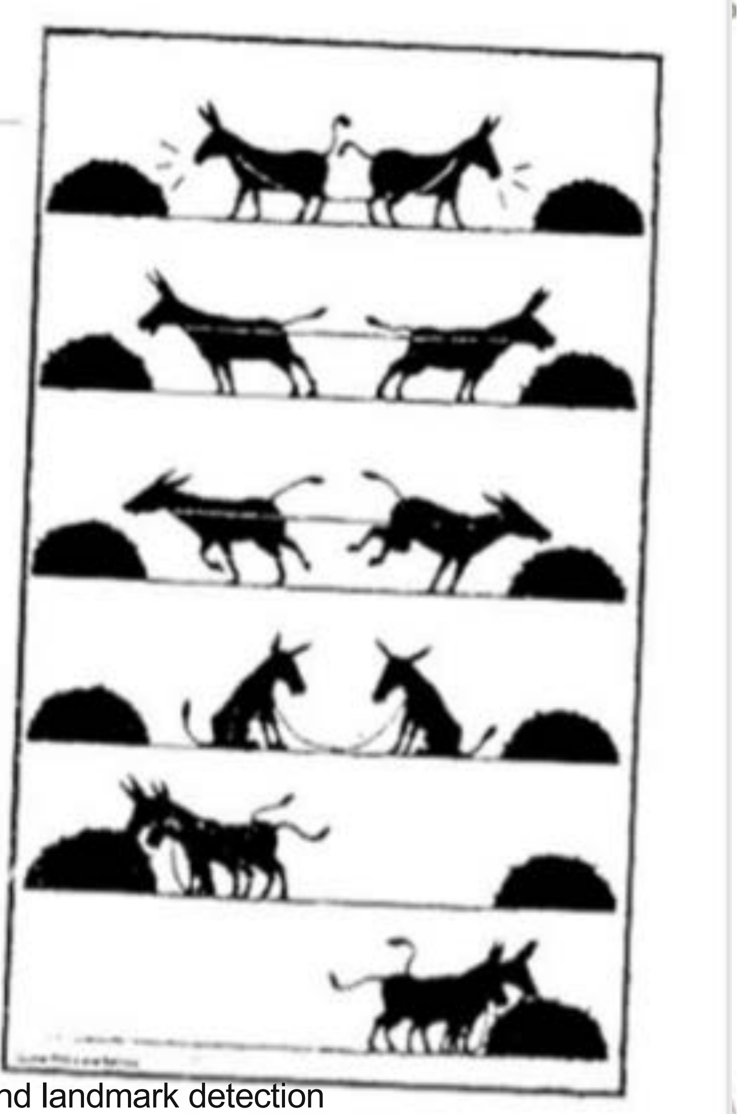
The importance of user tests!

- Most systems are scripts run under laboratory conditions
  - Does not give many indications of routine use
- Impact of the system is hard to measure
  - Better decisions, more confidence, faster, satisfaction?
- What is the influence on the patient?
  - Better treatment? Longer survival? Quality of life?
- User tests are more complex to set up but can really help
- AI and users are usually best together

https://news.harvard.edu/gazette/story/2020/11/risks-and-benefits-of-an-ai-revolution-in-medicine/

Scientific challenges

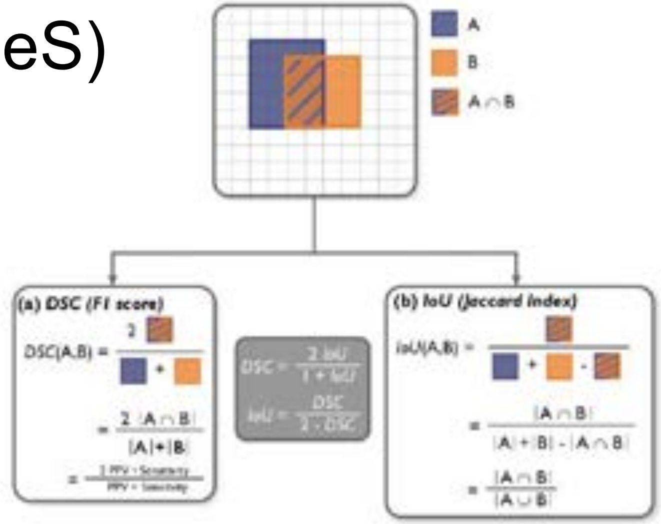
- Cooperation, Coopetition, Competition
- Many data sets are now available
  - Also many medical data sets
- Strong baselines help to judge quality
  - Not only the results count!
- Challenges can be run without sharing confidential data
  - Provide VMs or Docker containers



Jimenez-del-Toro, Oscar, et al. "Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks." *IEEE transactions on medical imaging* 35.11 (2016): 2459-2475.

Some more best practices

- Reporting for scientific challenges in medical imaging
  - BIAS (Biomedical Image Analysis challengeS)
  - Avoid bias, use the right measures
  - Use meaningful data sets and scenarios
  - How to chose the best evaluation metrics
  - …

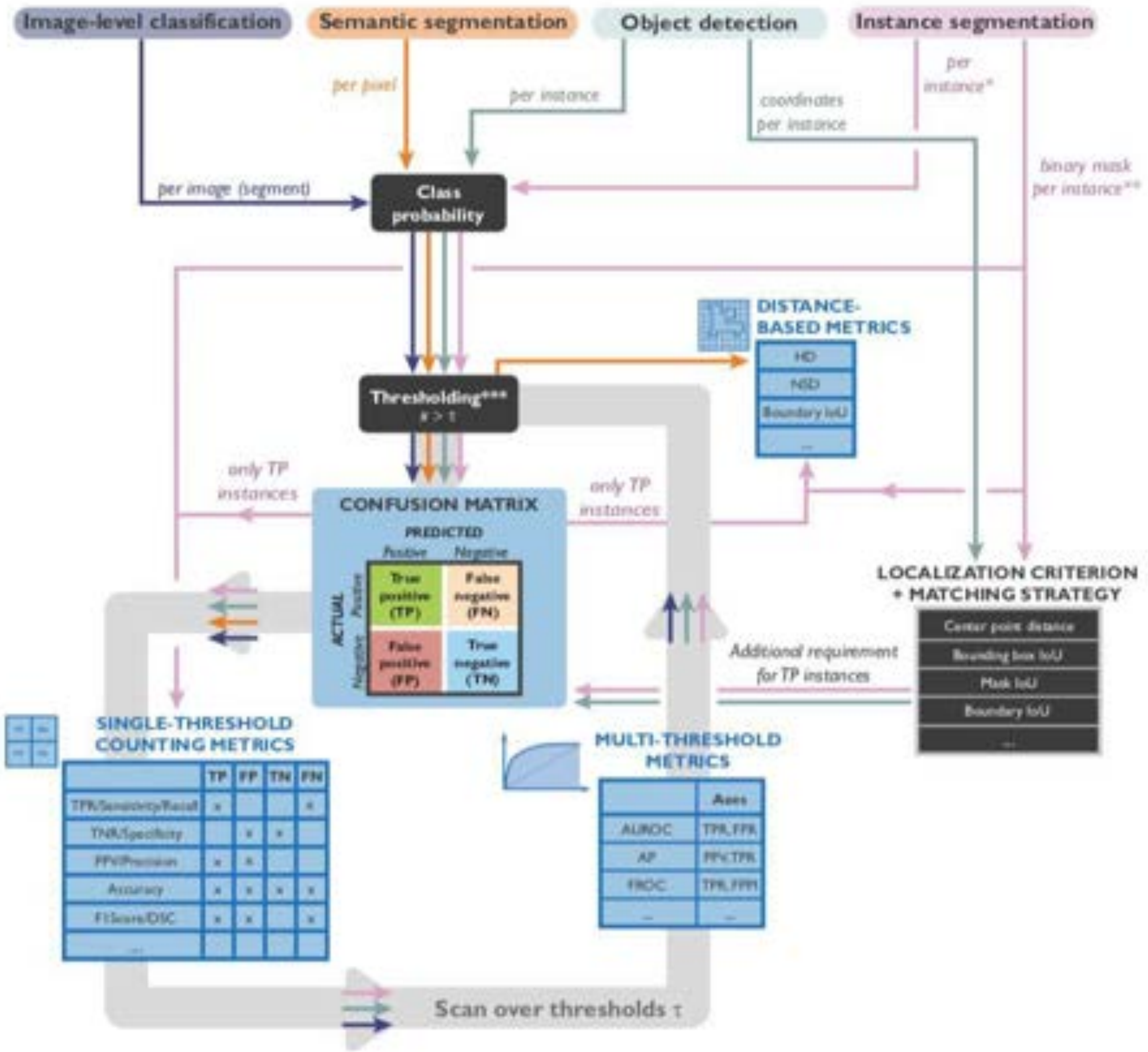

A
B
A ∩ B
(a) DSC (F1 score)
$DSC(A,B) = \frac{2\,|A \cap B|}{|A| + |B|}$
$= \frac{2\,PPV \cdot Sensitivity}{PPV + Sensitivity}$
$DSC = \frac{2\,IoU}{1 + IoU}$
$IoU = \frac{DSC}{2 - DSC}$
(b) IoU (Jaccard index)
$IoU(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$
$= \frac{|A \cap B|}{|A \cup B|}$

Maier-Hein, L., Eisenmann, M., Reinke, A., Onogur, S., Stankovic, M., Scholz, P., Arbel, T., Bogunovic, H., Bradley, A.P., Carass, A. and Feldmann, C., 2018. Why rankings of biomedical image analysis competitions should be interpreted with care. Nature communications, 9(1), pp.1-13.

Maier-Hein, L., Reinke, A., Kozubek, M., Martel, A.L., Arbel, T., Eisenmann, M., Hanbury, A., Jannin, P., Müller, H., Onogur, S. and Saez-Rodriguez, J., 2020. BIAS: Transparent reporting of biomedical image analysis challenges. *Medical image analysis*, *66*, p.101796.

Reinke, Annika, et al. "Common limitations of image processing metrics: A picture story." *arXiv preprint arXiv:2104.05642* (2021).

Tasks and measures



DISTANCE-BASED METRICS

HD
NSD
Boundary IoU
...

LOCALIZATION CRITERION + MATCHING STRATEGY

Center point distance
Bounding box IoU
Mask IoU
Boundary IoU
...

SINGLE-THRESHOLD COUNTING METRICS

	TP	FP	TN	FN
TPR/Sensitivity/Recall	x			x
TNR/Specificity		x	x	
PPV/Precision	x	x		
Accuracy	x	x	x	x
F1Score/DSC	x	x		x
...				

MULTI-THRESHOLD METRICS

	Axes
AUROC	TPR, FPR
AP	PPV, TPR
FROC	TPR, FPPI
...	...

Certification of medical SW

- Any use of AI in medicine needs to be certified (CE, FDA)
  - Software is a “medical device”
  - Unless only for a research study
  - Avoid risks for the patient, tedious process
- In-vitro diagnostics is more complex since 2022
  - Transition period for already certified tools
- Expensive to do, so not usable for research tools

Conclusions

- Medical AI is an extremely interesting domain
  - With high impact on people's lives!
- AI in medical imaging has many challenges remaining!
  - Some can be addressed relatively easily
  - Many will require much more research
- Consequences of (wrong) decisions are important
- Run user tests (also on prospective data)

Contact

- More information can be found at
  - http://medgift.hevs.ch/
  - http://publications.hevs.ch/
- Contact: Henning.mueller@hevs.ch

Hes·so VALAIS WALLIS
Haute Ecole de Gestion & Tourisme
Hochschule für Wirtschaft & Tourismus

European Commission | Horizon 2020
European Union funding
for Research & Innovation