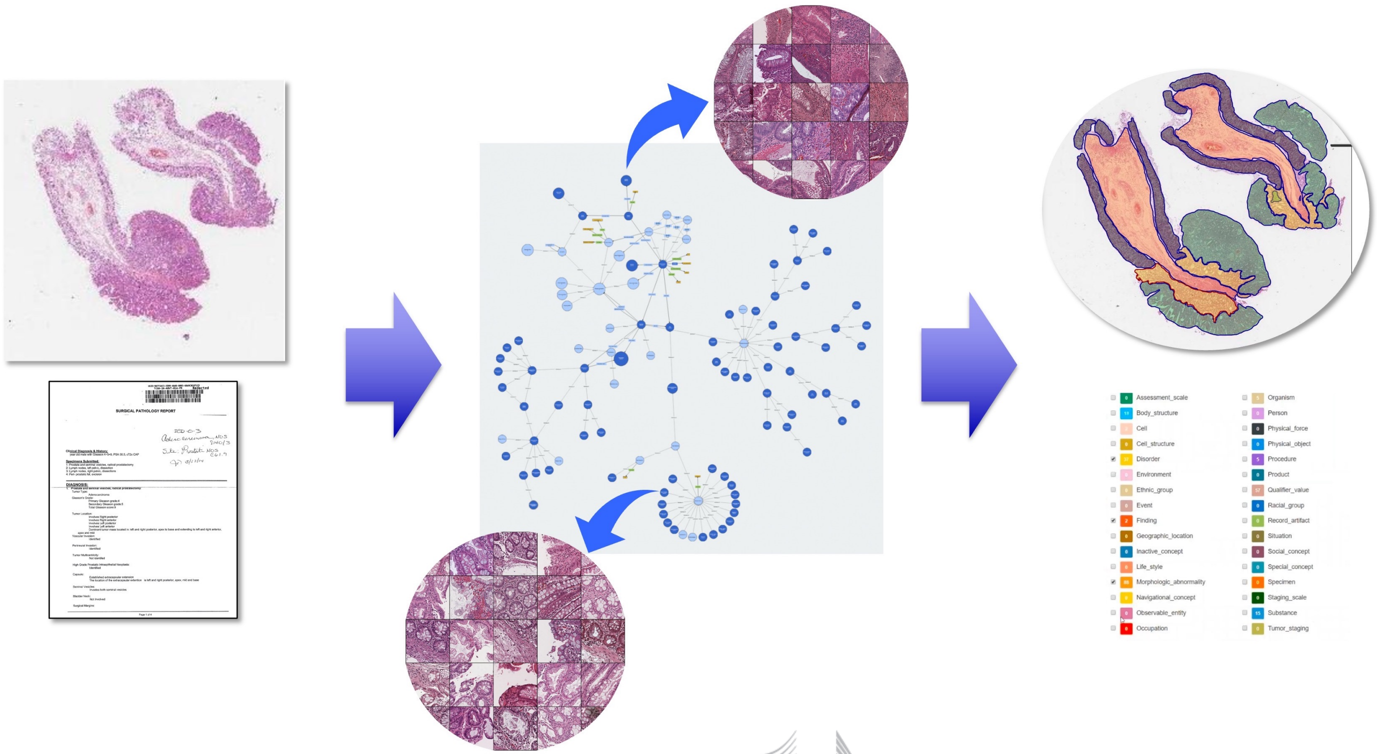
Multimodal machine learning for histopathology images: results from the ExaMode project



Henning Müller

HES-SO & UNIGE

Basel 5.9.2024

ExaMode

the sense
innovation and research center

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

Henning Müller

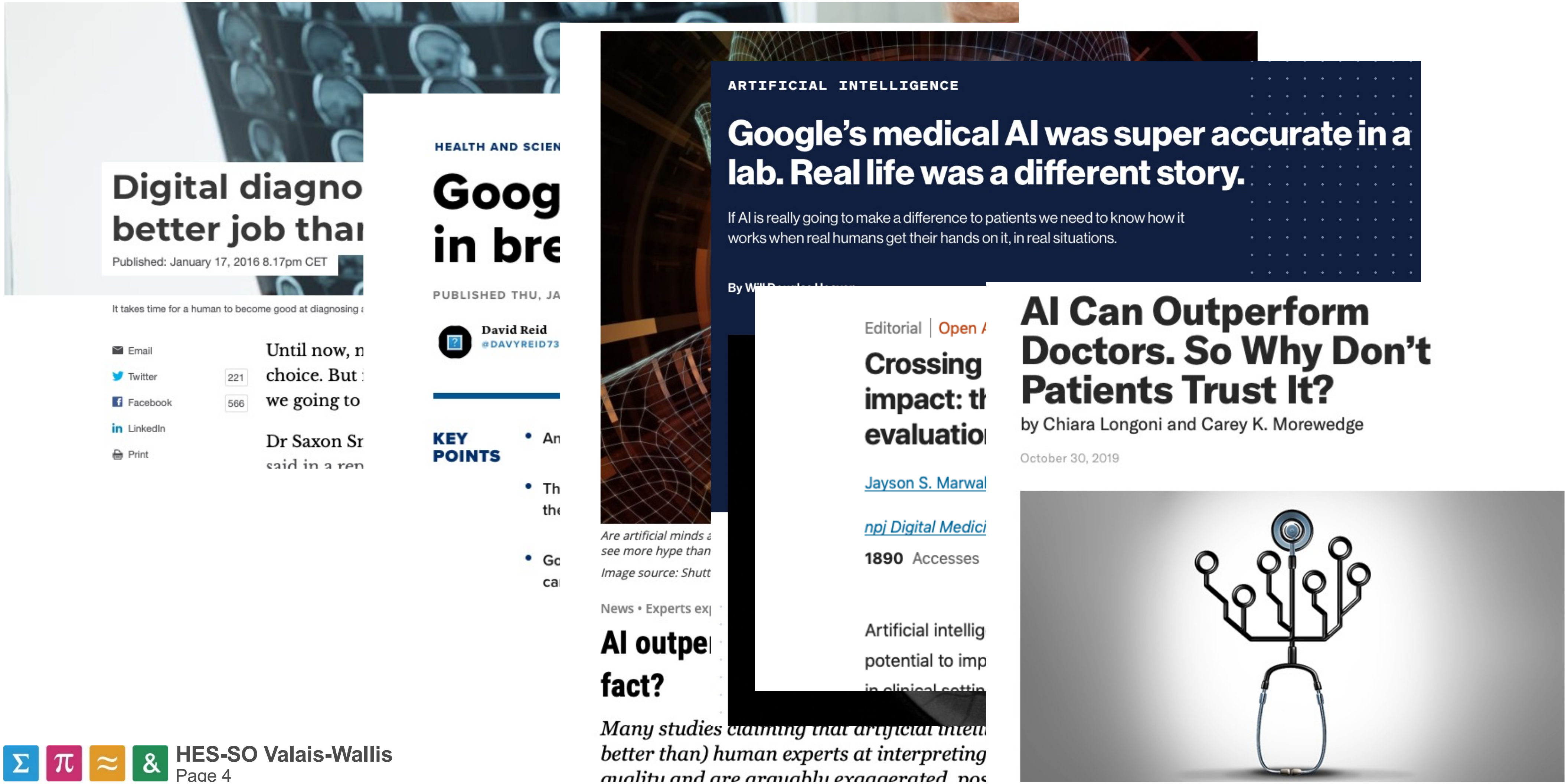
- Medical informatics studies in Heidelberg, Germany (1992-1997)
  - Exchange with Daimler Benz research, USA
- PhD in image processing, image retrieval, Geneva, Switzerland (1998-2002)
  - Exchange with Monash University, Melbourne, AUS
- Professor in radiology and medical informatics at the University of Geneva (2014-)
- Professor in Computer Science, HES-SO, Switzerland (2007-)
- Visiting faculty at Martinos Center (2015-2016)
- Member of the Swiss National Research Council (2020-)

The promise of medical AI

Geoff Hinton: On Radiology

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

Medical AI in the media



Challenges in medical AI

Haute Ecole de Gestion & Tourisme
Hochschule für Wirtschaft & Tourismus

- Very large amounts of data, not always accessible
  - Histopathology images are 100’000x100’000 pixels
- Annotations are hard to obtain and expensive
  - Never really full images, subjectivity remains
- All available data sources should be combined
  - Multimodal data analysis: text, images, structured data, genetics
- Generalization is important, data changes over time
  - Data heterogeneity is high (acquisition parameters, staining)
- Interpretability and explainability

ExaMode

- Extreme-scale Analytics via Multimodal Ontology Discovery & Enhancement
  - Very large-scale data analysis
  - Histopathology is becoming digital
- EU Horizon 2020: ICT-12-2018-2020
  - Acceptance rate: 6 of 78 submissions
- Budget: ~5 Mio Euros
  - 7 partners: academic, commercial and hospital partners, plus a national supercomputing center

European Commission
Horizon 2020
European Union funding
for Research & Innovation

Introduction: consortium

High Performance Computing (HPC) resources: SURFSARA

ACADEMIC PARTNERS
(UNIPD, HESSO)

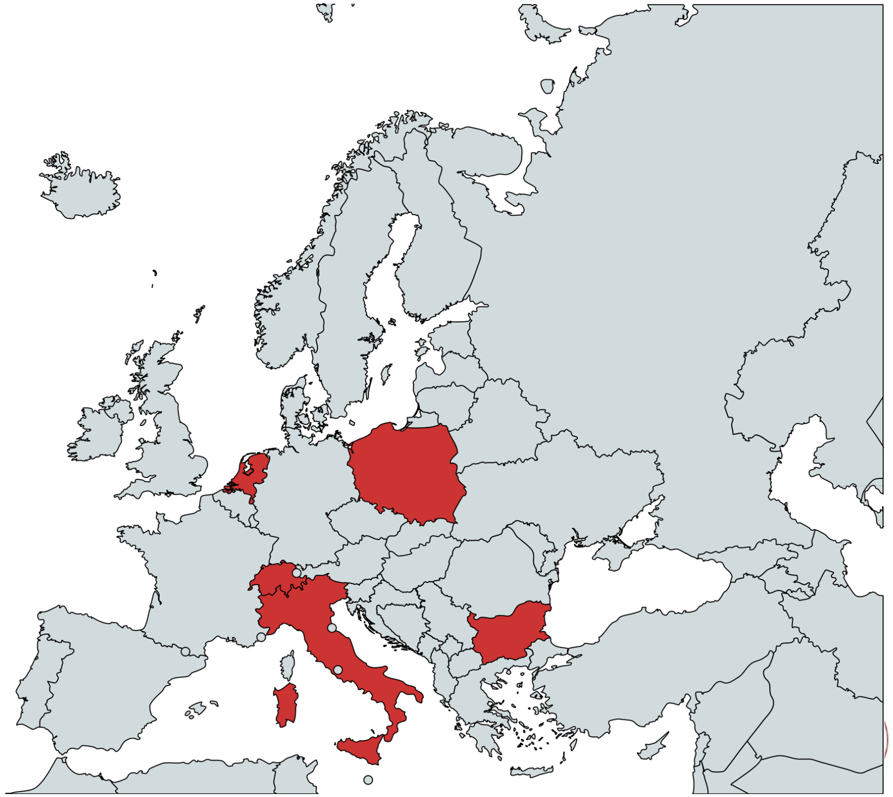
Scientific experience in extracting knowledge from:
- heterogeneous images
- text

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



Objectives of the project

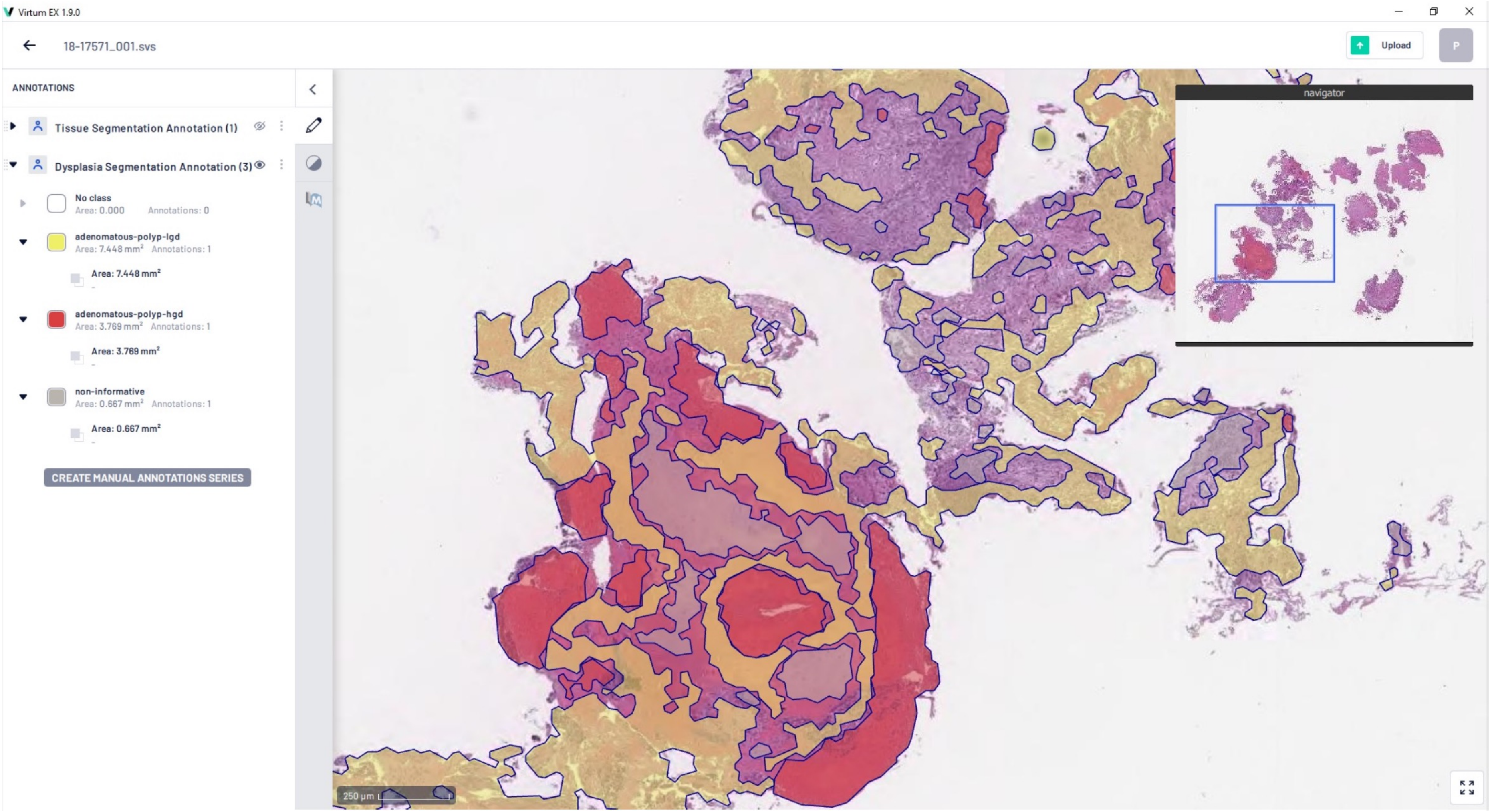
Hes·so VALAIS WALLIS
Haute Ecole de Gestion & Tourisme
Hochschule für Wirtschaft & Tourismus

- Develop training of deep learning models with weak labels, so not pixel level annotations
  - And a combination of weak with strong labels
- Combine semantic knowledge from pathology reports with image data (multimodality)
  - Development of domain ontologies
  - Use this to generate weak labels for training
- Make all this scalable to extremely large amounts of data
  - With national computing centers (Surfsara)
- Use images from the literature for training
  - Combined with clinical images

Use cases chosen in ExaMode

- Colon
  - Large number of images exist with screening, labor intensive, high economic value
- Lung
  - Large number with screening, labor intensive, economic value
- Cervix/Uterus
  - Large number with screening, labor intensive, economic value
- Celiac disease
  - For a non-oncologic application, large and increasing number

Virtum – an image viewer



Data used

	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)	TCGAPubme	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	100	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

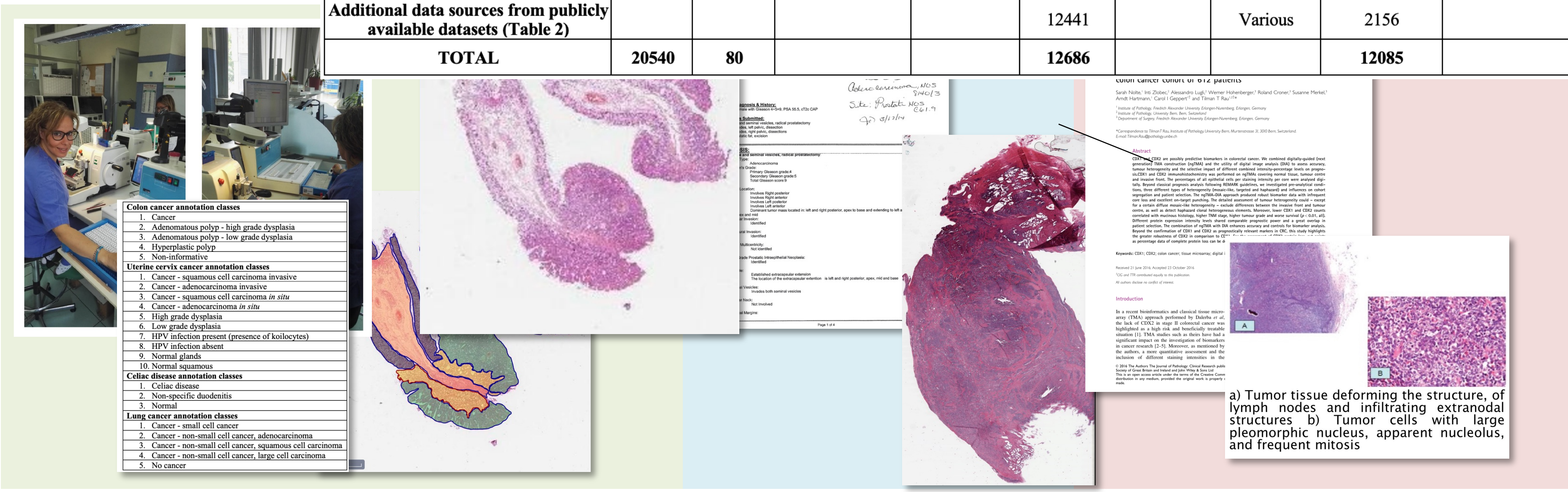
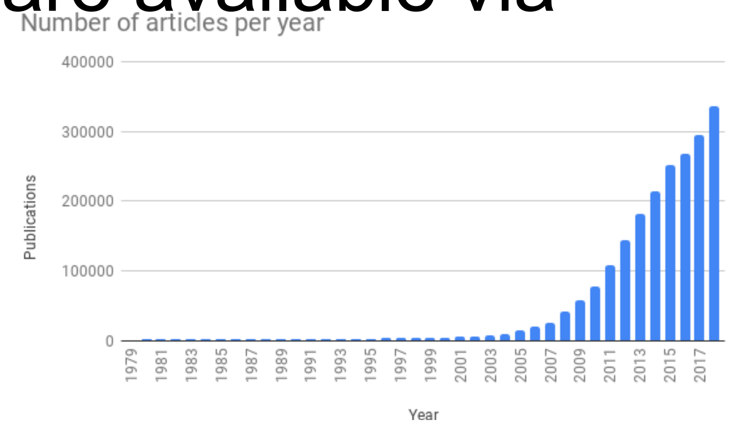


Image accessibility

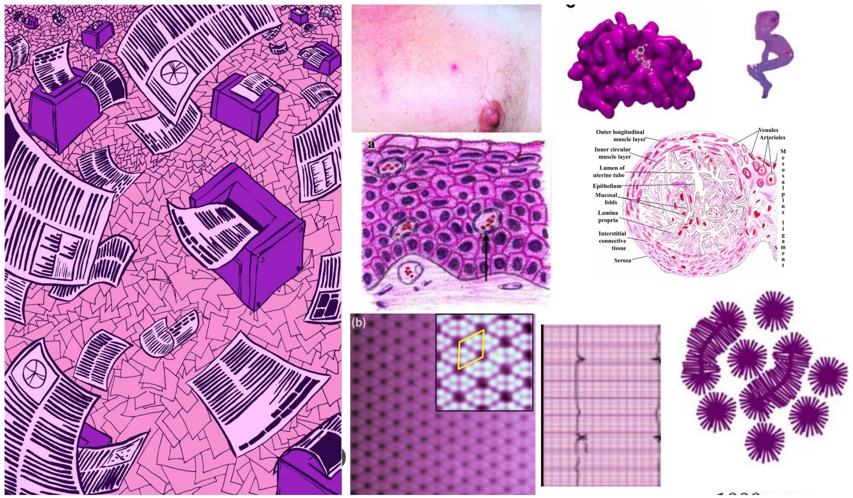
- Open data policies of funding agencies make large medical data sets available
  - Particularly NIH is pushing towards this
- TCIA and TCGA are very large repositories
  - There are many scientific challenges
- Images from the Biomedical literature are available via PubMedCentral
  - Exponentially increasing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

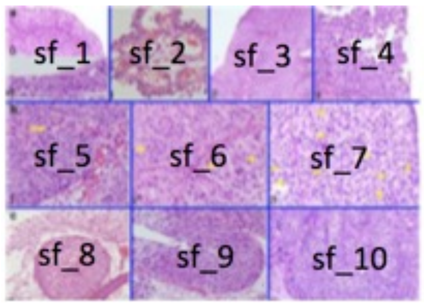


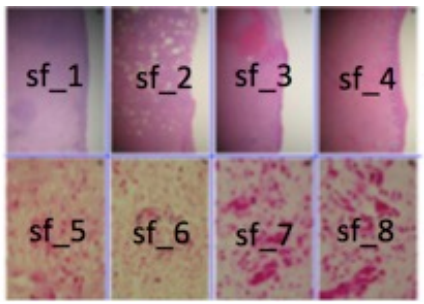
Challenges with PubMed

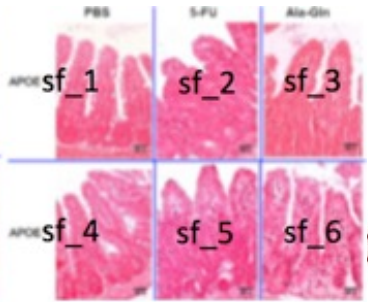
Hes·so Valais Wallis
Haute Ecole de Gestion & Tourisme
Hochschule für Wirtschaft & Tourismus

- >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



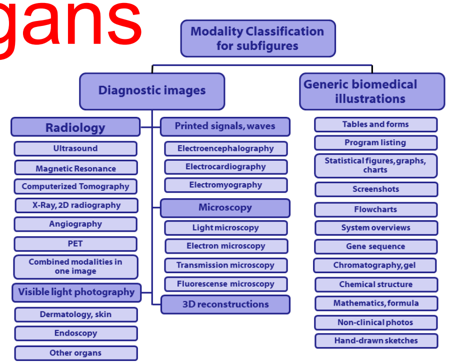






Making the images usable

- 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



Medical NLP is not trivial …

- Non-standard abbreviations
- Spelling mistakes, quickly written
- Technical language
  - Latinized terms, synonyms
- Nested, complex phrases
- Negation …
  - Several levels (“little evidence of”)
  - Not clear what terms they refer to, double negations

A/B	acid-base ratio
ab	abdomen, abdominal abortion
Ab	antibody
AB	abortion, AB Blood Type
ABC	airway, breathing, circulation aspiration biopsy cytology
ABCD	airway, breathing, circulation, disability asymmetry, borders, color, diameter (features on considering "Is it a mali ABCD rating (a staging system for prostate cancer)
ABCs ABCDs ABCDEs	airway, breathing, circulation, etc. Refers to priority of needs in emergenc recurrent.
ACA	acinic cell carcinoma Affordable Care Act
Abd	abdomen abdominal[abduction]
ABD	army battle dressing

cardiac arrest *noun* heart attack

asystole	congestive heart failure	heart arrest	heart stoppage
cardiac infarction	coronary infarction	heart attack	myocardial infarction
cardiopulmonary arrest	coronary thrombosis	heart failure	tachycardia

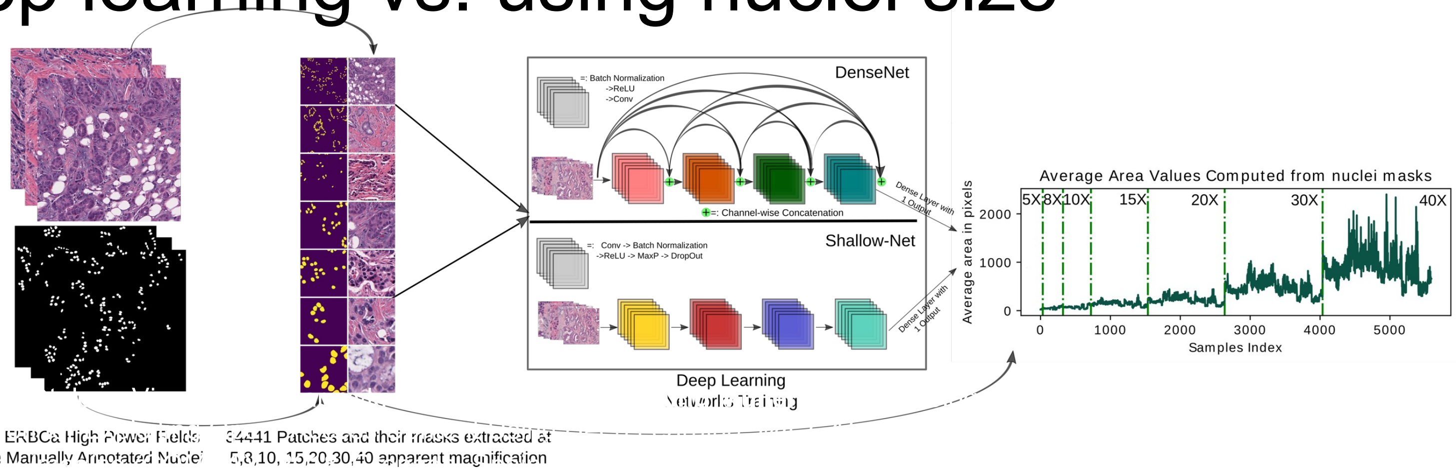
Advantages of literature images

- Rare images (unusual, untypical) are generally used for articles and case descriptions, so are oversampled
  - 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

Magnification regression

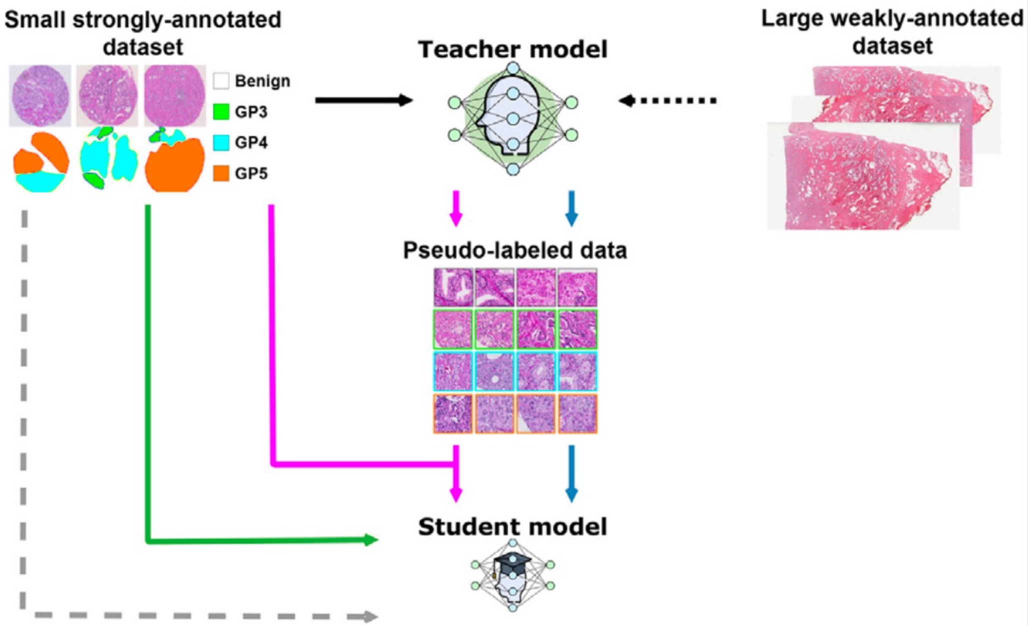
Hes·so VALAIS WALLIS
Haute Ecole de Gestion & Tourisme
Hochschule für Wirtschaft & Tourismus

- Magnification (or pixel size) in literature images is usually not known
- If we want to compare visual similarity, the scale of the structures is essential
  - Unlike object recognition where scale is irrelevant
- Brut force deep learning vs. using nuclei size



Weakly supervised learning

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



Teacher/student paradigm approaches:

Semi-supervised learning

Semi-weakly supervised

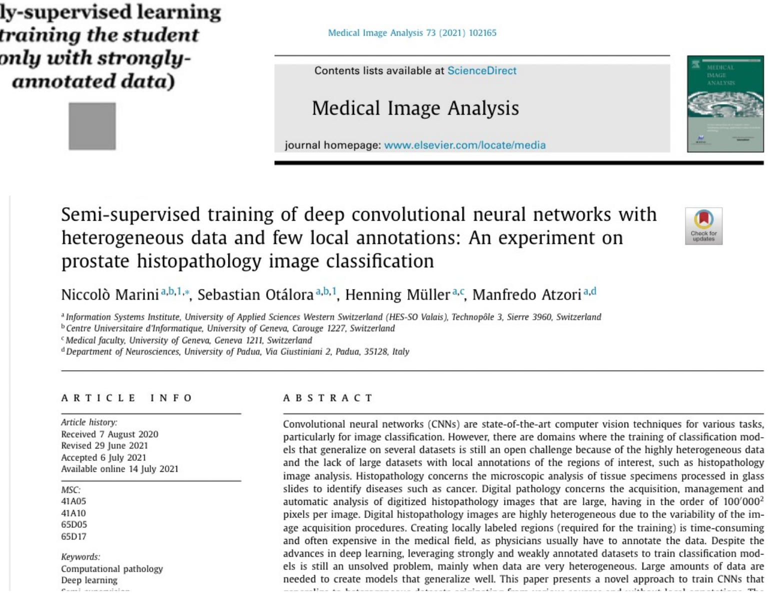
Student training variants:

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

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 6 July 2021
Available online 14 July 2021

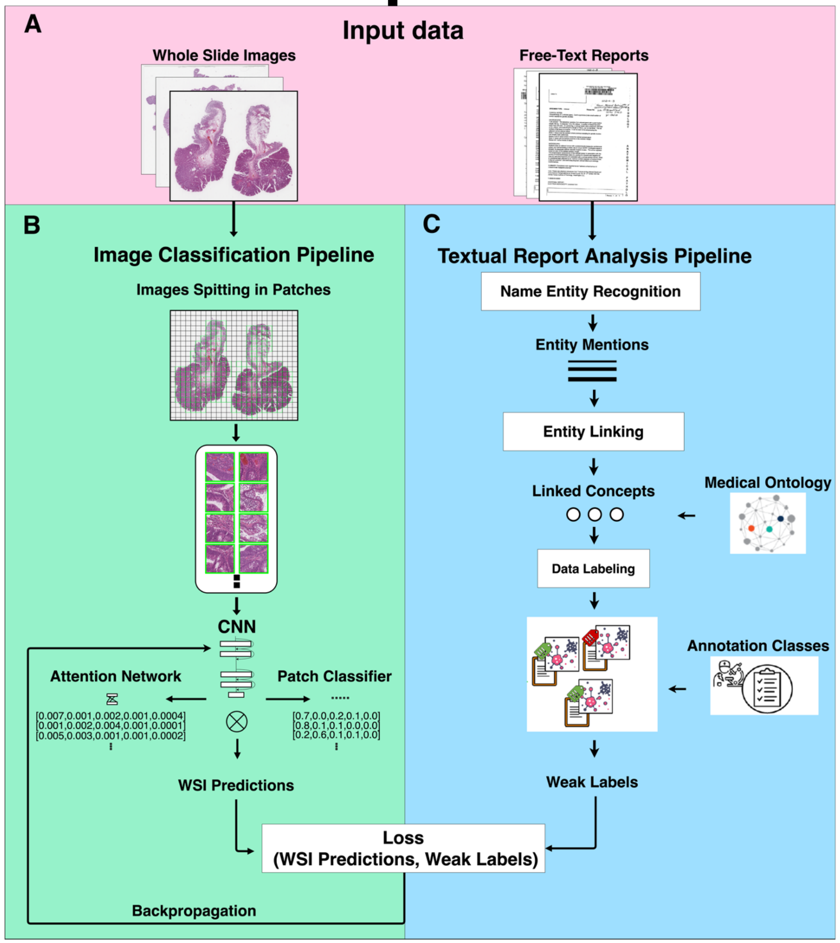
MSC:
41A05
41A10
65D05
65D17

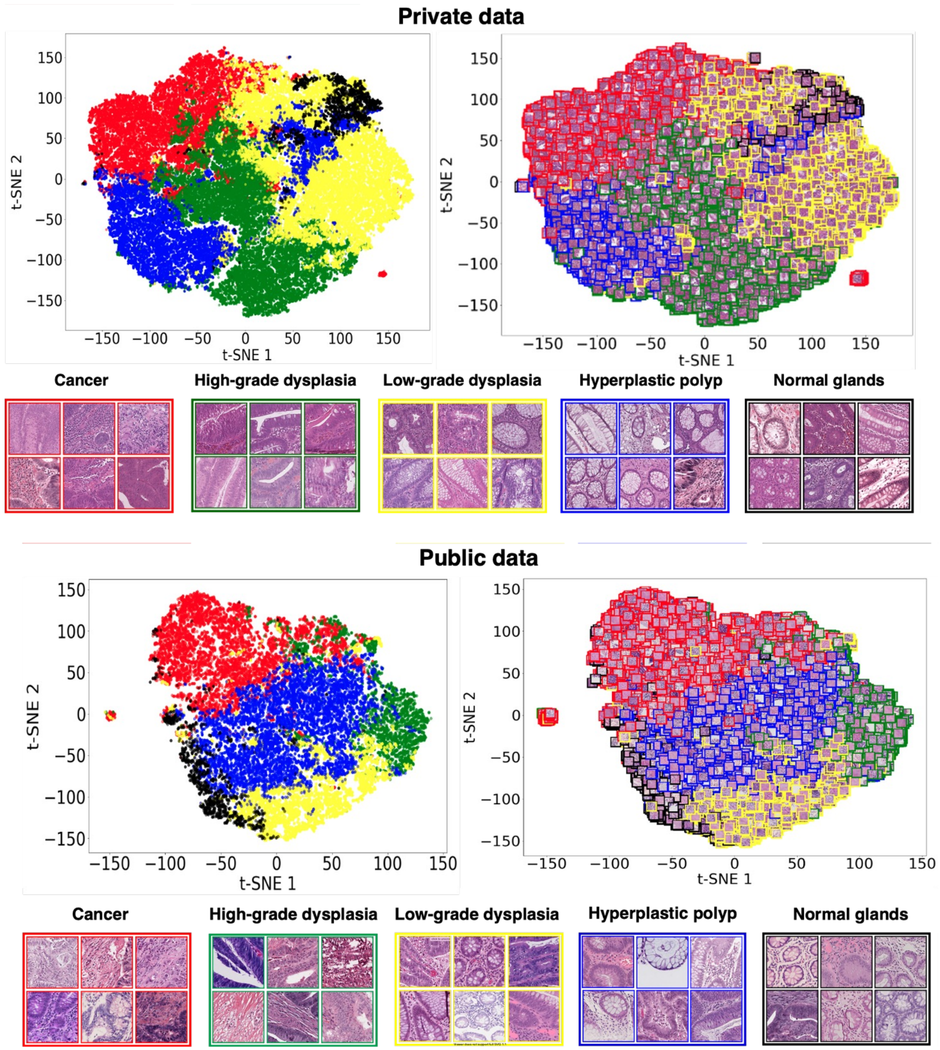
Keywords:
Computational pathology
Deep learning

ABSTRACT

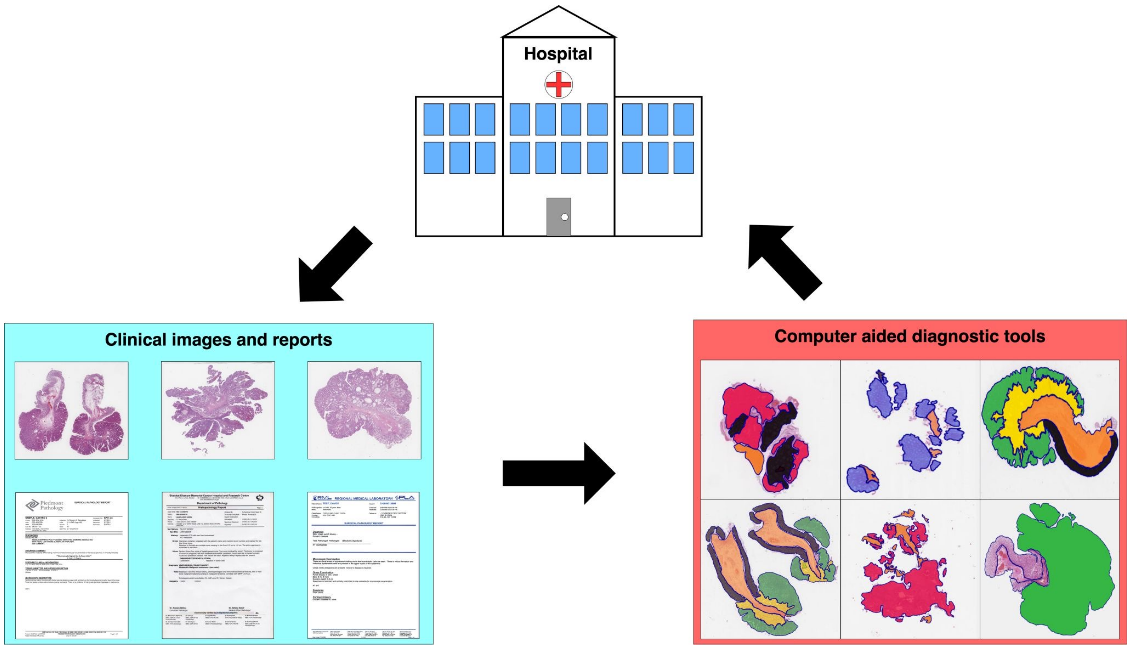
Convolutional neural networks (CNNs) are state-of-the-art computer vision techniques for various tasks, particularly for image classification. However, there are domains where the training of classification models that generalize on several datasets is still an open challenge because of the highly heterogeneous 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 $100'000^2$ 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

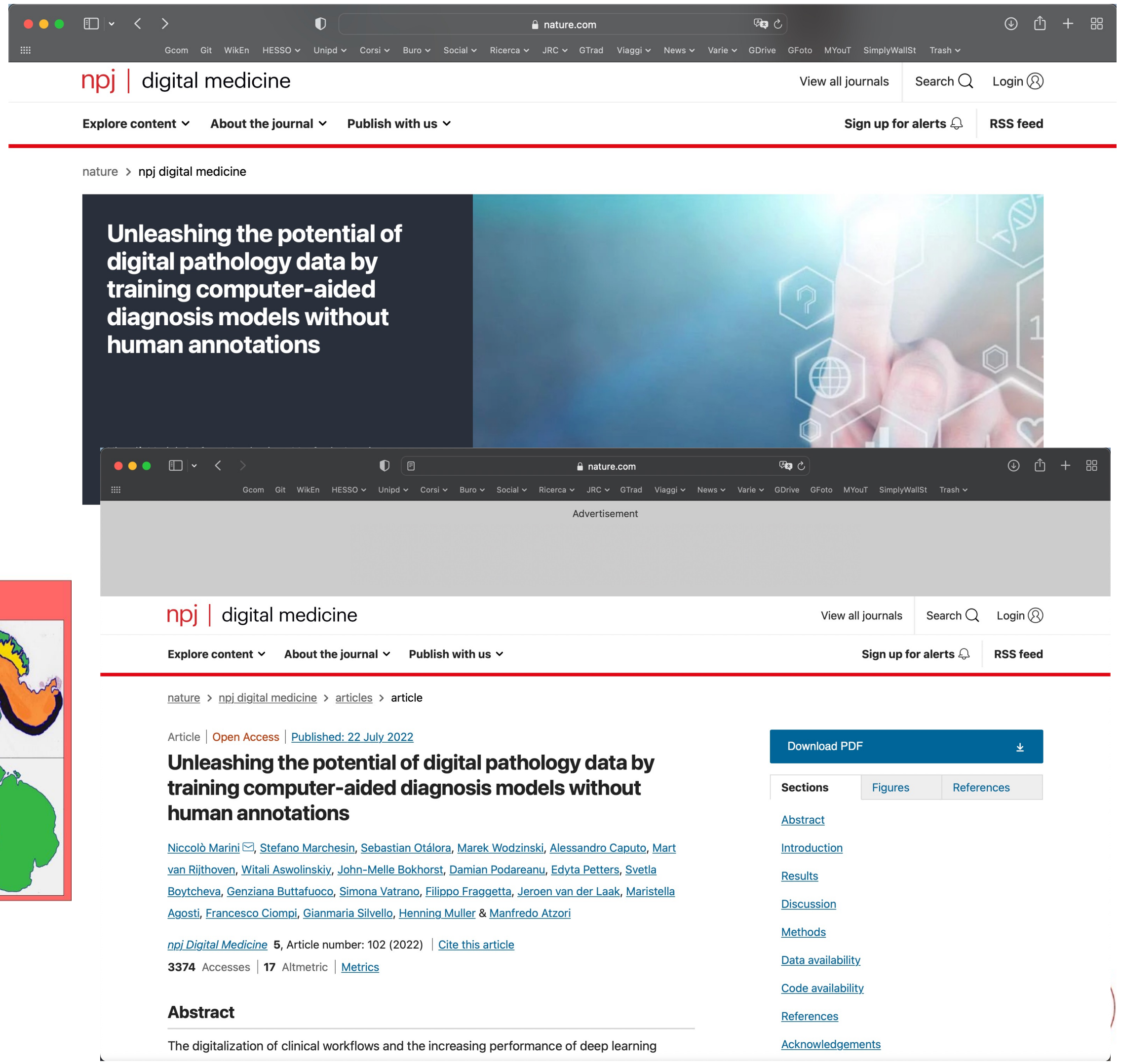
Weakly supervised learning from reports





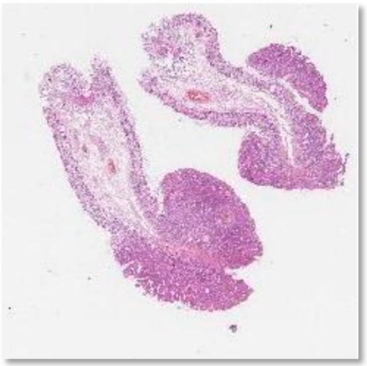
First results on multimodal data

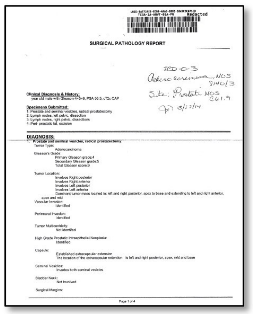


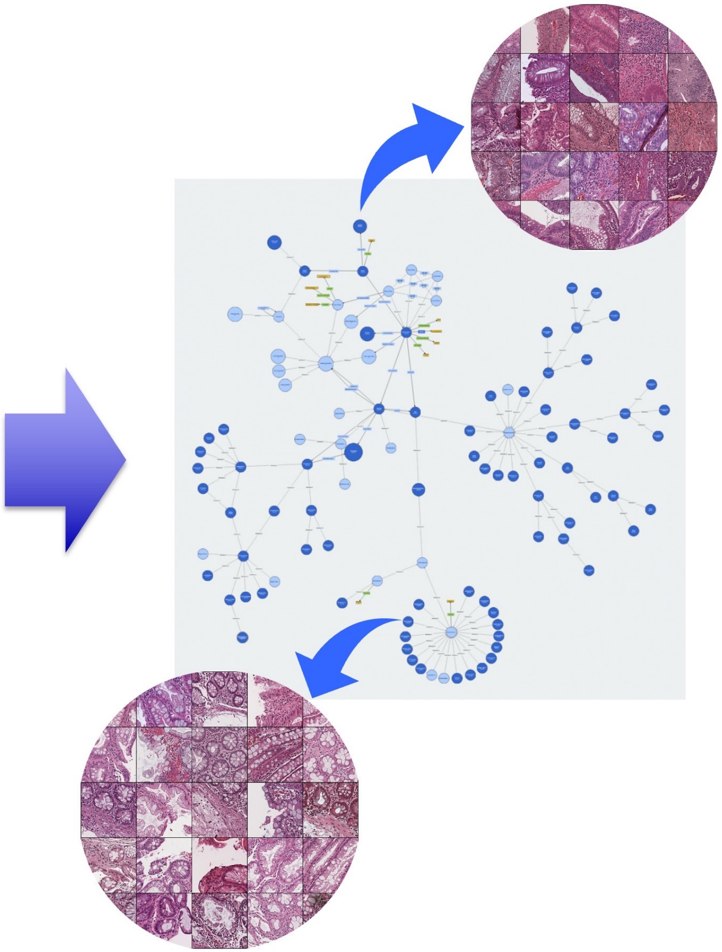


Fully automatic learning

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




SURGICAL PATHOLOGY REPORT



Multimodal embeddings

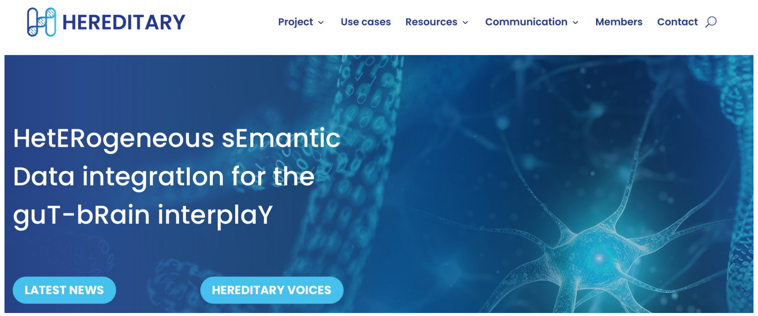
- Multimodal network trained with images and texts
  - Create links between text and visual information
  - Single representation
- Outperforms unimodal models, requires less training data
  - Works on smaller data sets for WSI classification
  - Separate input branches for images and visual information
- Creates a representation of visual semantic information

ELSEVIER
Medical Image Analysis
Volume 97, October 2024, 103303

Multimodal representations of biomedical knowledge from limited training whole slide images and reports using deep learning

Niccolò Marini[a 1], Stefano Marchesin[b 1], Marek Wodzinski[a c], Alessandro Caputo[d e], Damian Podareanu[f], Bryan Cardenas Guevara[f], Svetla Boytcheva[g h], Simona Vatrano[e], Filippo Fraggetta[e i], Francesco Ciompi[i], Gianmaria Silvello[b], Henning Müller[a j], Manfredo Atzori[a k]

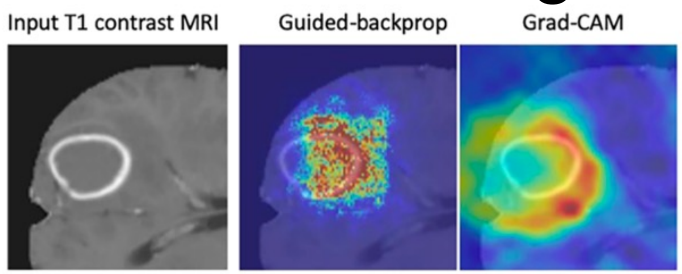
Even more multimodal



- Hereditary project that started in 2024
  - Horizon Europe, budget of 13 Mio €, 18 partners
- Explore the link between the gut (microbiome) and the brain
  - Neurodegenerative diseases: MS, ALS, Alzheimers, Parkinsons
  - Microbiome, genetics, clinical data, signals (EEG), imaging
    - Several types of imaging: MRI, OCT, eye fundus, histopathology
  - All information is mapped onto semantics
  - Requires to deal with missing data

Interpretability of Deep Learning

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



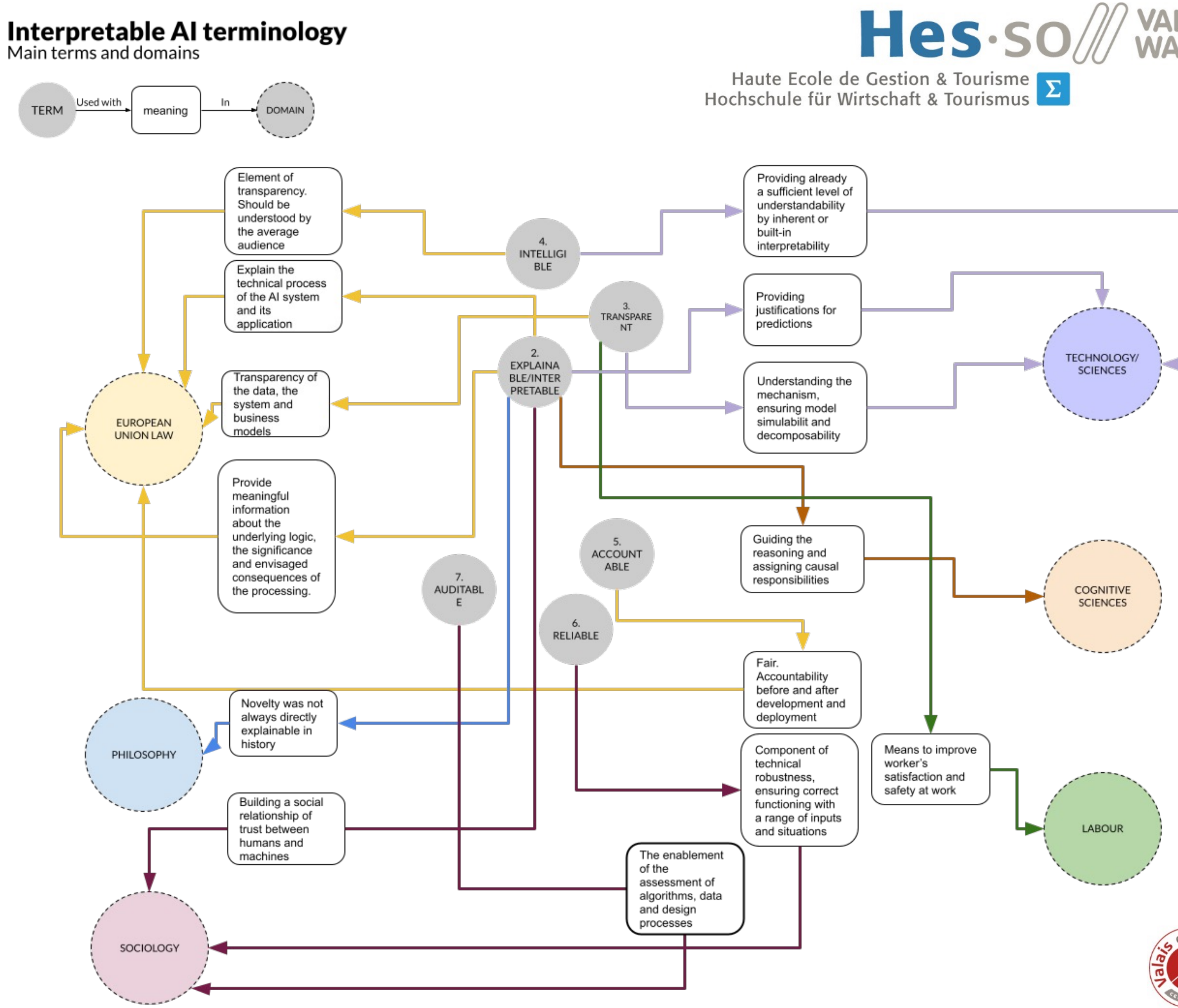
- In medicine it is particularly important to make sure that results can be explained & reproduced
  - High impact of wrong decisions
- There are many approaches for interpretability
  - 2D projections, PCA, TSNE
  - Class activation maps, saliency, …

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 early summer 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.

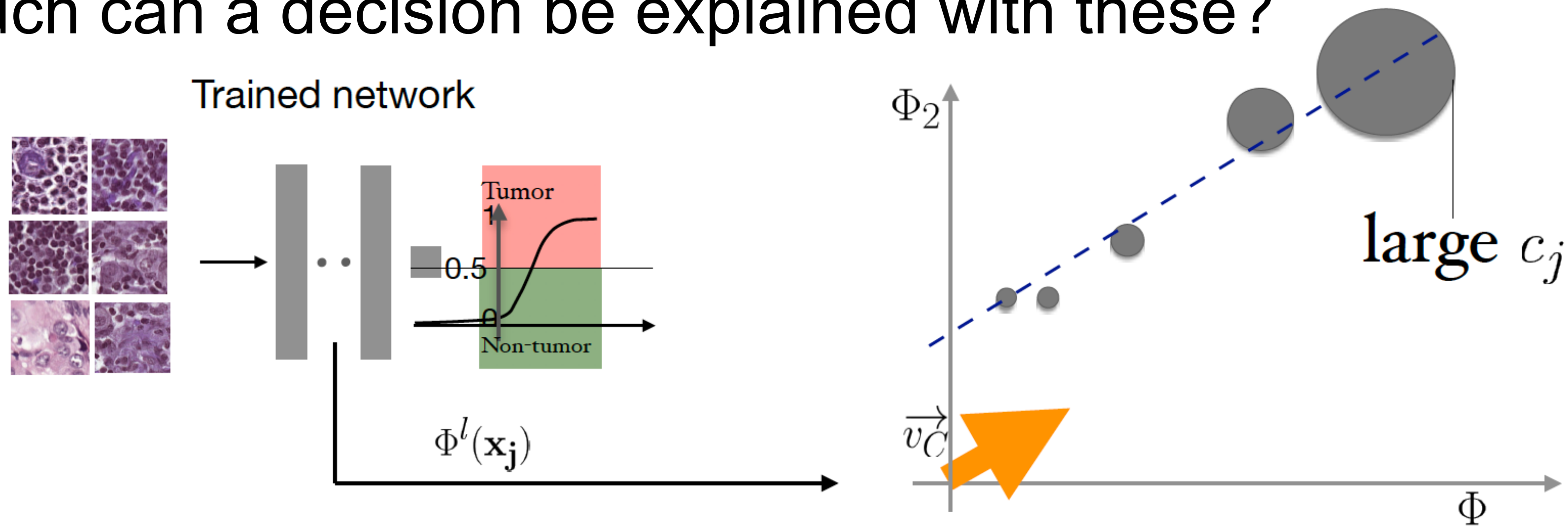
Taxonomy



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.
EUROPEAN UNION LAW
4. INTELLIGIBLE
3. 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
5. ACCOUNTABLE
7. AUDITABLE
6. RELIABLE
Guiding the reasoning and assigning causal responsibilities
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

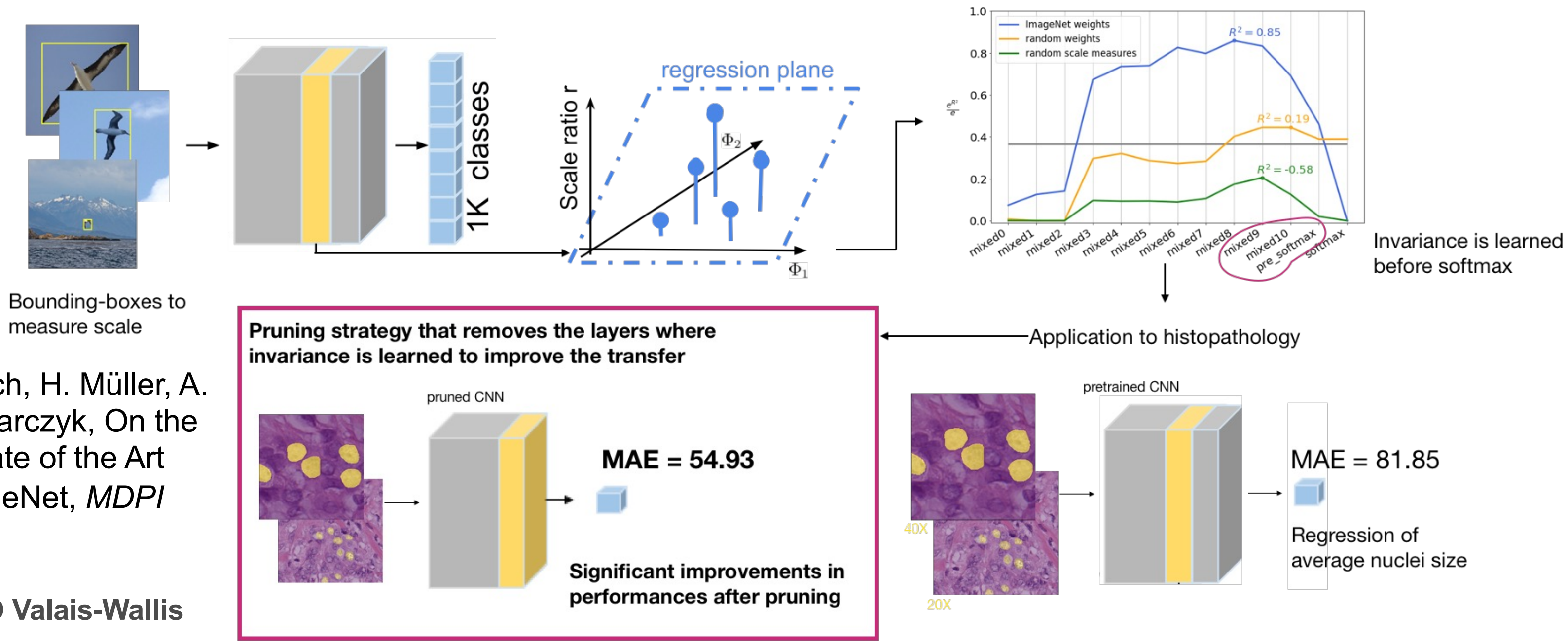
- Identify existing features and check how the decision layers correlate to these features
  - i.e.: nuclei size, internal heterogeneity, borders, …
  - How much can a decision be explained with these?



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



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.

The importance of user tests!

- Most systems are scripts run under laboratory conditions
  - Does not give essential 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 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/

Conclusions

- Interpretability of deep learning is a key for integration of tools into clinical workflows
  - Explain decisions, understand a potential bias, …
- ExaMode addressed many current challenges in ML
  - Making things scalable (internal and external resources)
    - Also to allow for a better generalization
  - Learn from weak labels
    - We should also use strong labels when they are available
  - Learn from multimodal data, create multimodal embeddings
    - Most medical images have a report attached

Contact

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

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