Turing-Roche Knowledge Share Series

Transparent AI in Healthcare - Strategies and Pitfalls

Tapabrata **Rohan** Chakraborty, 30/01/2024
[email: tchakraborty at turing]

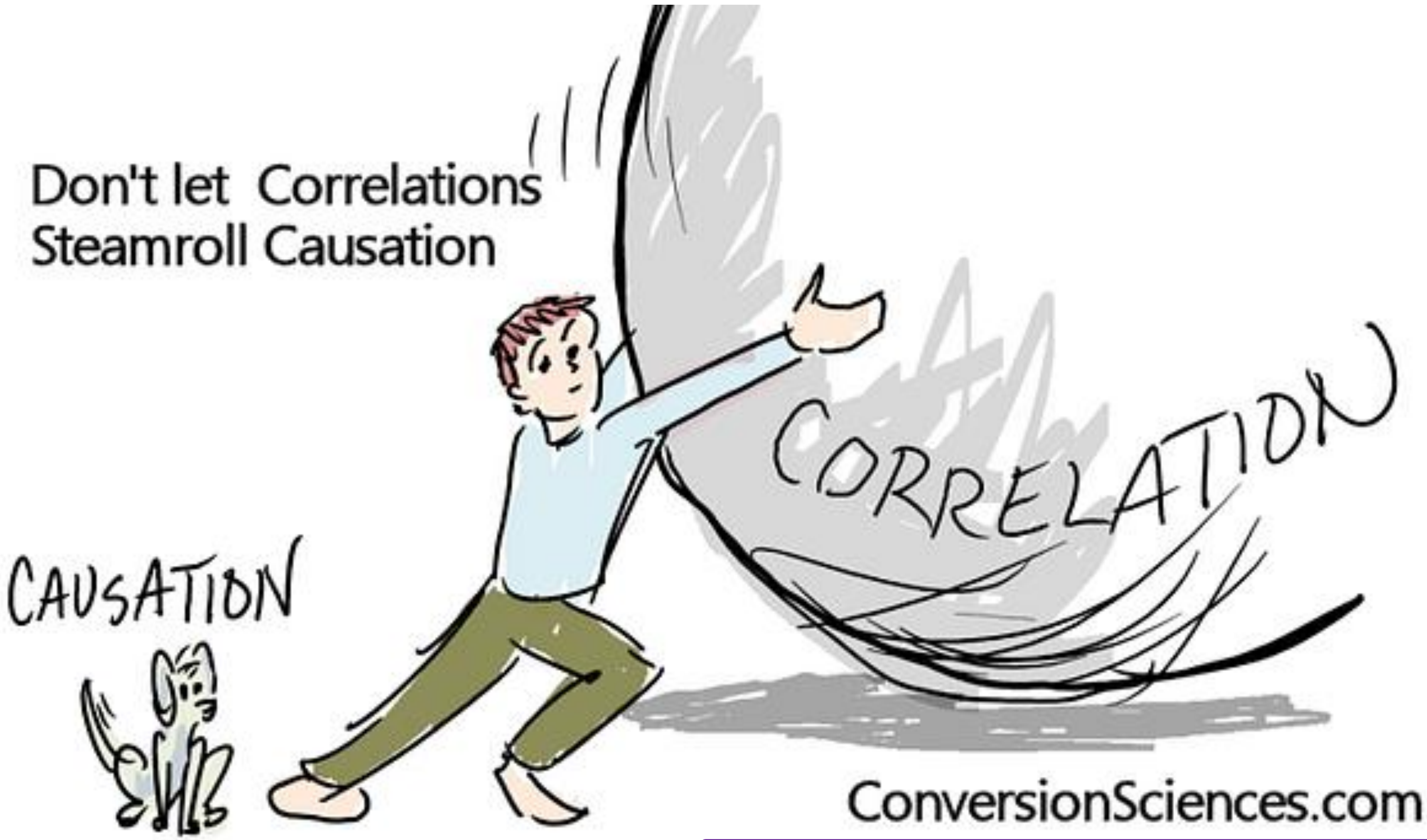
Transparent AI in Healthcare: Setting the scene

- Only a few AI models make it to the clinic, while it is important to be careful in deployment of clinical AI, there are other issues too.

- Clinical AI tools need to be transparent to be trustworthy, but do the models themselves need to be explainable for clinical translation?

- Perhaps it is enough to show that the results are interpretable and consistent with clinical decision making to be acceptable to clinicians?

Part I: Causal Machine Learning – the Holy Grail?



Source: https://sangeetm.github.io/

When not to use causal inference?

If it is possible to do actual causal experiments, causal inference should be avoided. For example, tech AI companies do A/B tests that study the effect of a change between two groups, when a new product feature is introduced, one being exposed to the old version of the product (control group), and the other group to the new version.

In the healthcare sector, the equivalent would be randomised double blind clinical control trials, for example the group getting the new treatment (vaccine/drug) vs the control group (alternative treatment or placebo), causal inference is used only when a randomised control trial (best option) is not viable.

When can we use causal inference?

The causal inference framework provides us tools to conduct causal experiments with secondary data, but in order to validate those inferences, we need enough richness in the data being used in terms of quality/quantity and confidence in the proposed model or causal diagram.

In the context of healthcare, this is starting to become possible with the advent of large multimodal health datasets from UK BioBank, Genomics England, MIMIC/eICU, TCGA/TCIA datasets.

The Three Layers of Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing, Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past two years?

Source: https://sangeetm.github.io/

SCM & RCM: 2 ways of doing causal inference

- **Structural Causal Models** (SCM) developed by Judea Pearl which enables us to infer causality via Directed Acyclic Graphs (DAG), through the "do-operator/do-calculus": **Pr(Y=y|do(X=x))**

- The **Rubin Causal Model** (RCM) developed by Donald Rubin uses "*potential outcomes*" framework by analysing the potential outcomes (Y) for different levels of X in an interventional experiment.

- Again, the gold standard is RCT (randomised control trial) which is not always possible, and we will focus on SCM only for this talk.

Contrastive Causal Fairness in Machine Learning

- *D-contrast:* Is it fair to make decision D for individual I, instead of decision D'?
- *I-contrast:* Is it fair to make decision D for individual I, while make D' for individual J?
- *T-contrast:* Is it fair to make decision D for individual I at time t, but make D' at time t'?

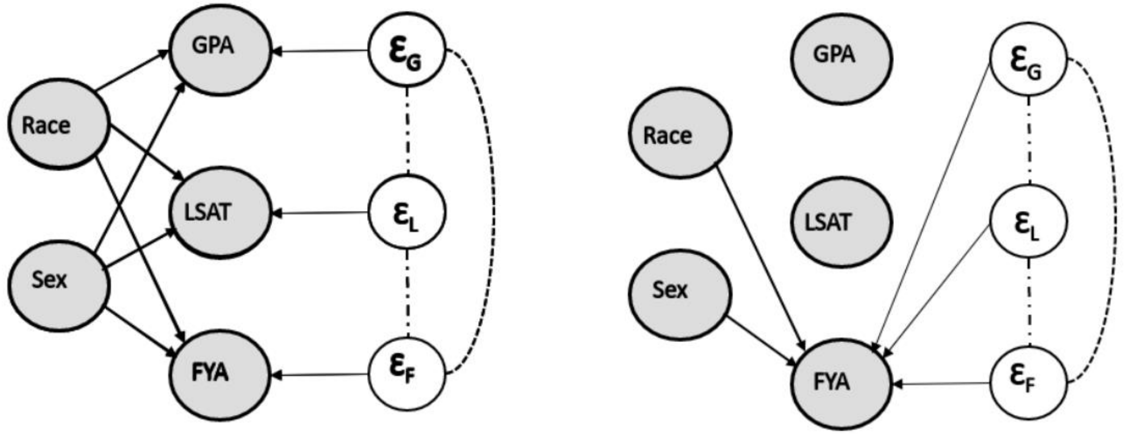


Table 1: Average accuracy (%) using logistic regression.

Full	Unaware	Counterfactual	Contrastive
0.873	0.894	0.918	0.937

- **We start with counterfactual fairness (What IF?) and modify the equations to answer contrastive questions (Why this and not that?).**

Chakraborti et al., IEEE Letters of Computer Society

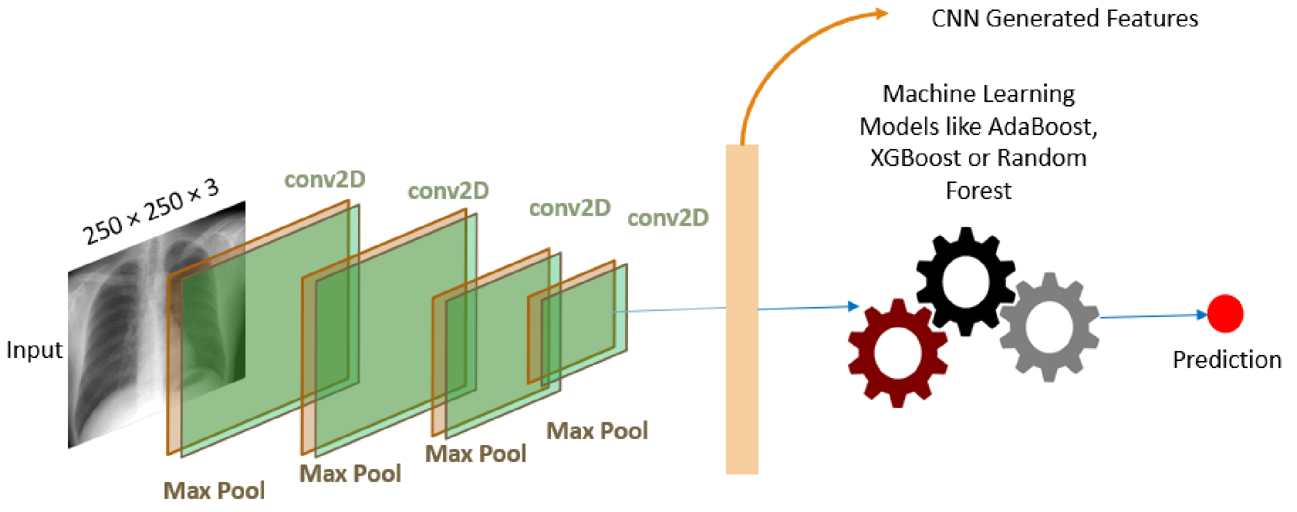
Proximal vs Distal Causality?

Part 2: Hybrid Deep Learning

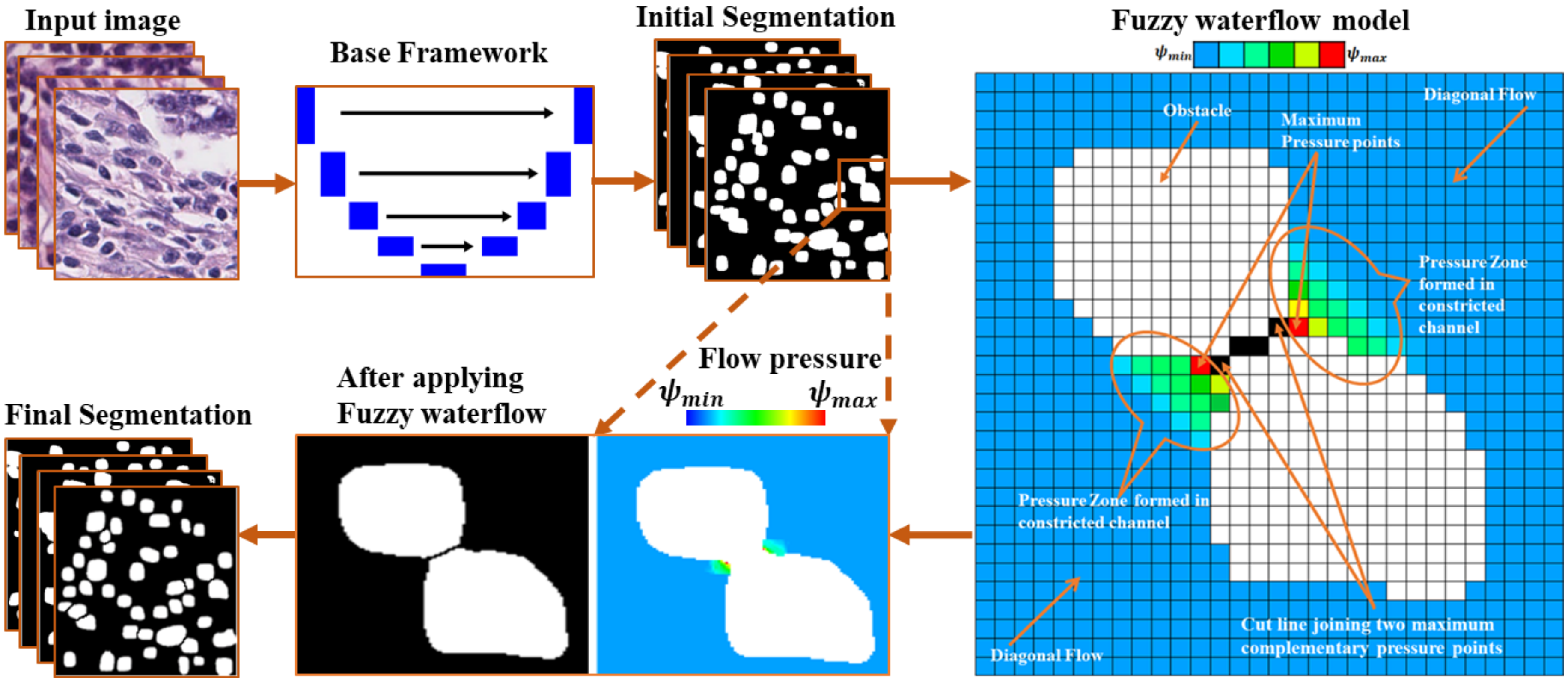
- End-to-end integration of data driven deep learning with interpretable mechanistic model
- The deep model produces coarse output that is fine-tuned by the mechanistic model
- The layers are differentiable and so the errors can be back propagated for training together

Related terms

- Physics/biology inspired deep learning
- Bayesian neural networks with domain based priors
- Declarative neural networks
- Neurosymbolic AI

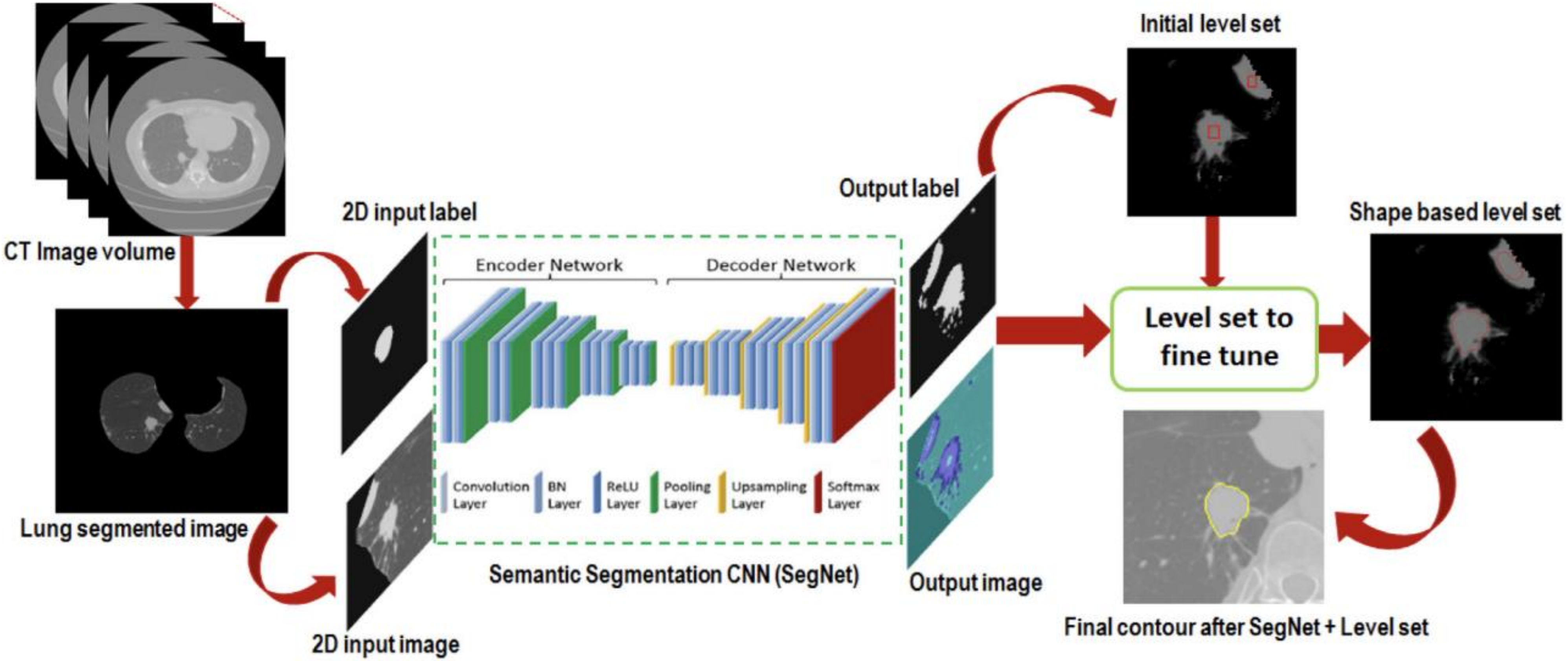


Example 1: Digital Pathology Cell Segmentation



Published in PLOS One

Example 2: Radiology Lung CT Image Segmentation



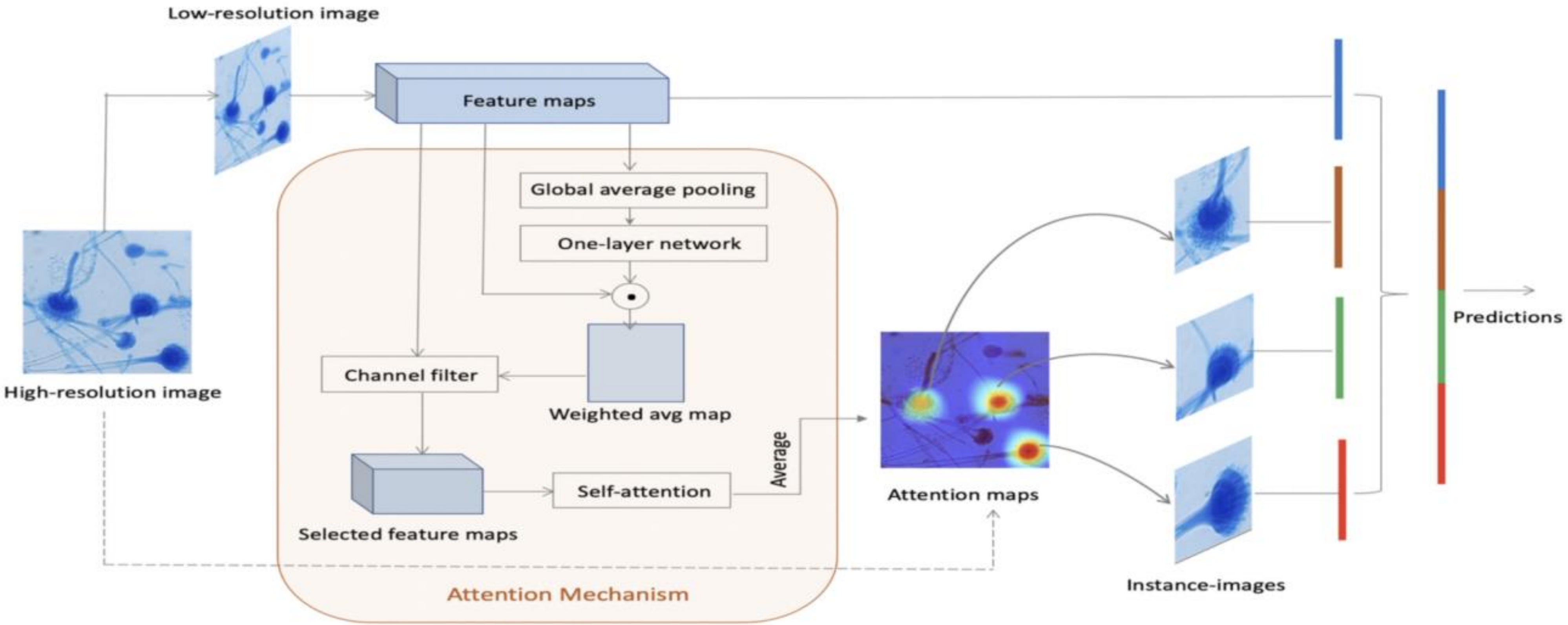
Published in Pattern Recognition Letters

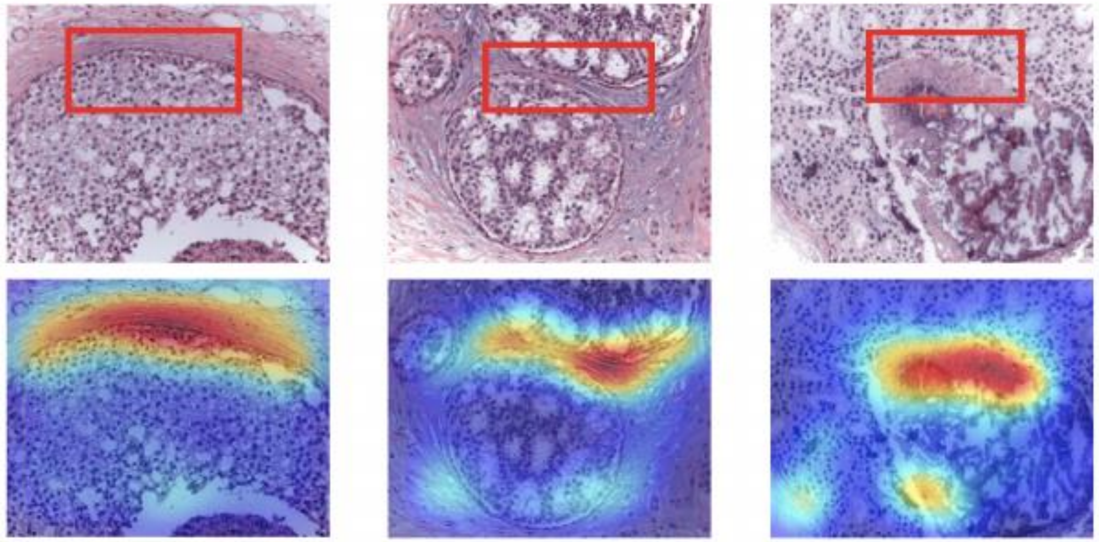
Part III: Aligning Clinical and Learned Features

- Explainable models help us to understand the inner workings of an AI decision system, however they are difficult to implement, and in some applications may not even be necessary to reach the level of trustworthiness needed for successful clinical translation.

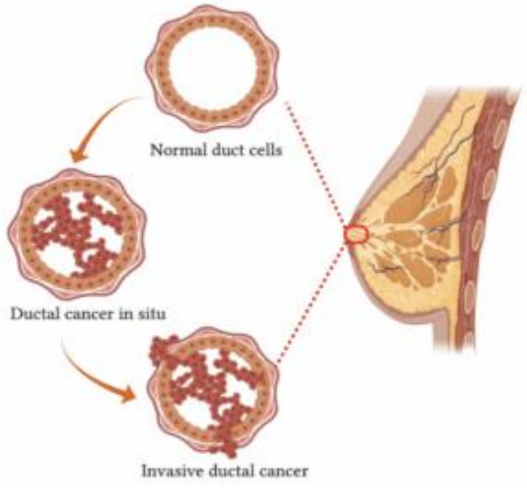
- Often it is enough to show that the decisions are based on clinically relevant attributes extracted from the data, and the results are robust to patient heterogeneity and hold true for personalised decisions.

Multi-instance Self-Attention based Networks

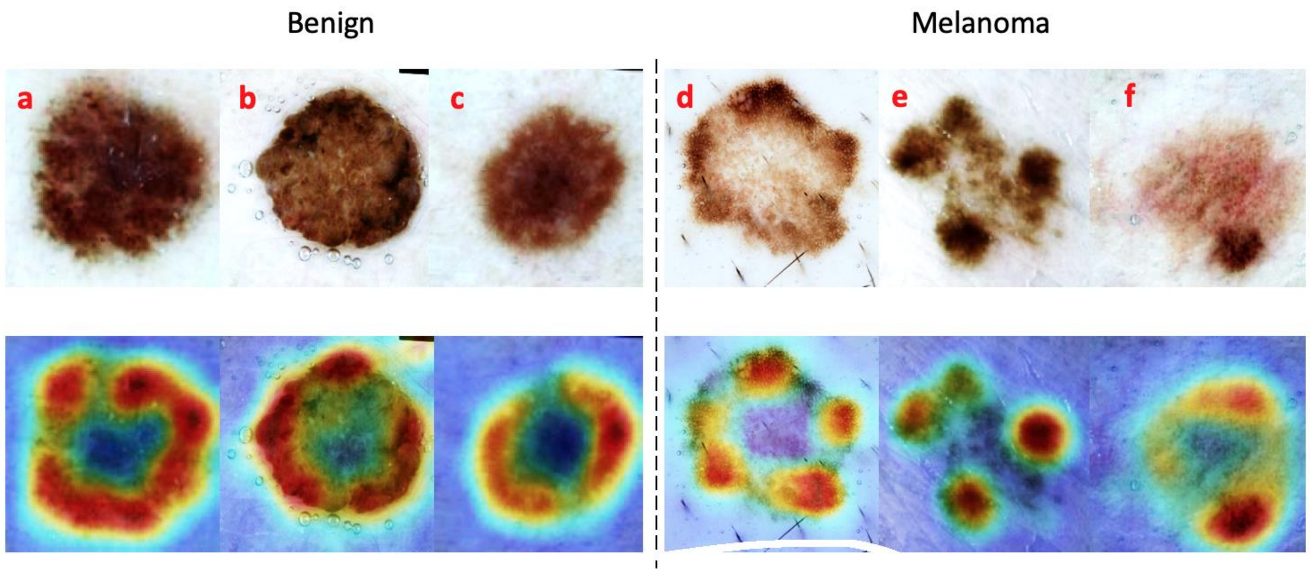


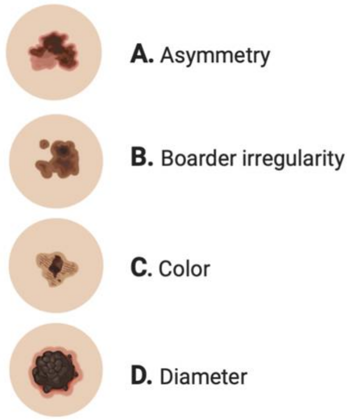


(a) Visualization of the attention maps in different classes.

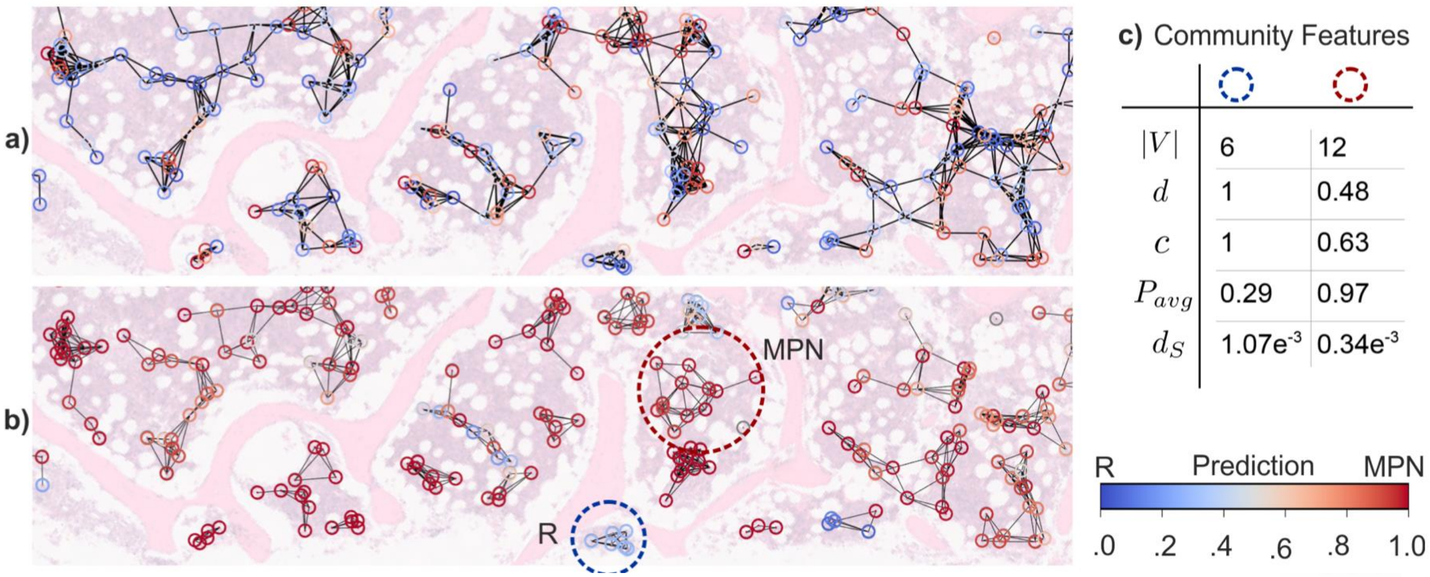


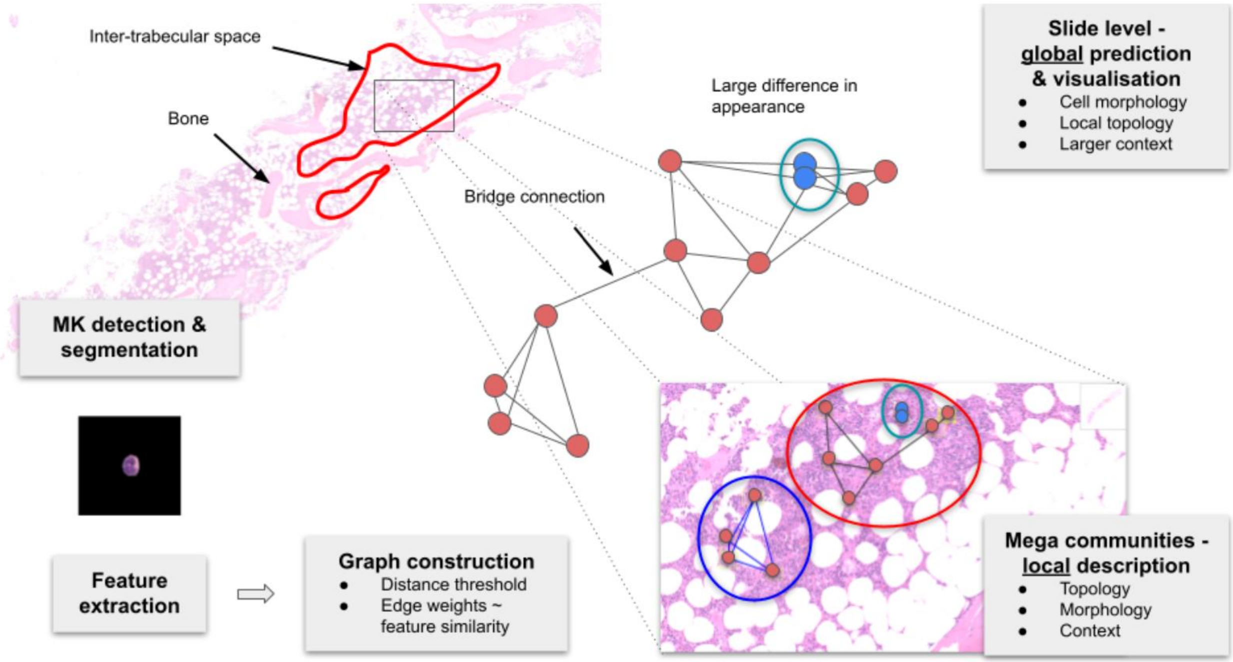
(b) Schematic representation





Published in MICCAI, ISBI





Graphical Neural Networks

- **Hierarchical graph neural network for tissue image analysis.**
- **3 stages, starting with the most zoomed in view, where graph nodes are placed on relevant cells.**
- **These graphs are then grouped into tumour micro-environments as the 2nd meso-stage.**
- **The final overview is given by clustering and having a graph at whole slide image level.**

Published in EMBC

PART IV: Quantify Predictive Uncertainty

Two Patients present with a headache

Clinical Analogue:

Most likely diagnosis (Point Estimate)

Initial Impression

Prediction with uncertainty

Personalised differential diagnosis

Model trained on patients with known diagnoses

Diagnostic AI model

Patient data collected

Migraine
-> Treat with painkillers

Migraine	Cluster Headache	ischaemic stroke	subarachnoid haemorrhage
98%	<1%	<1%	<1%

-> Treat with painkillers

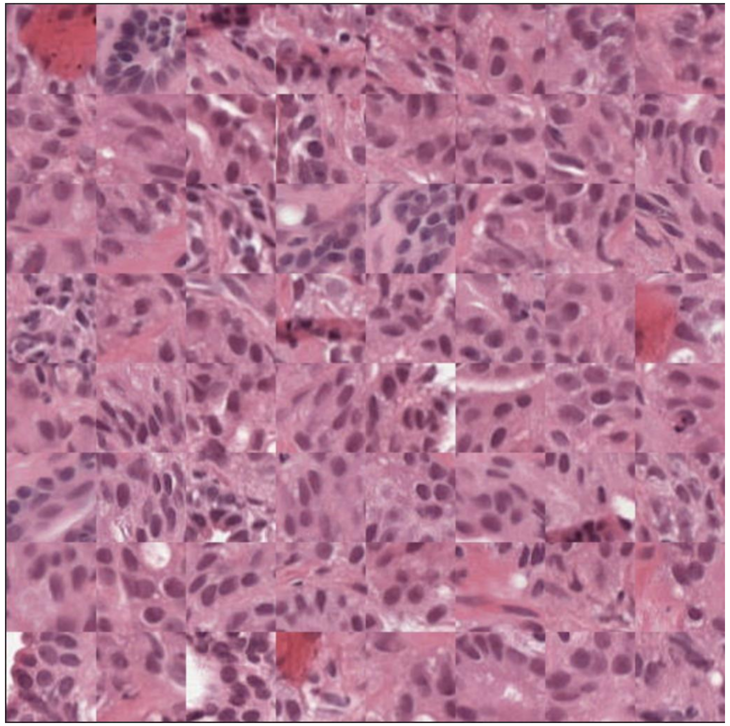
Migraine
-> Treat with painkillers

Migraine	Cluster Headache	ischaemic stroke	subarachnoid haemorrhage
80%	10%	5%	5%

-> Run further diagnostic tests

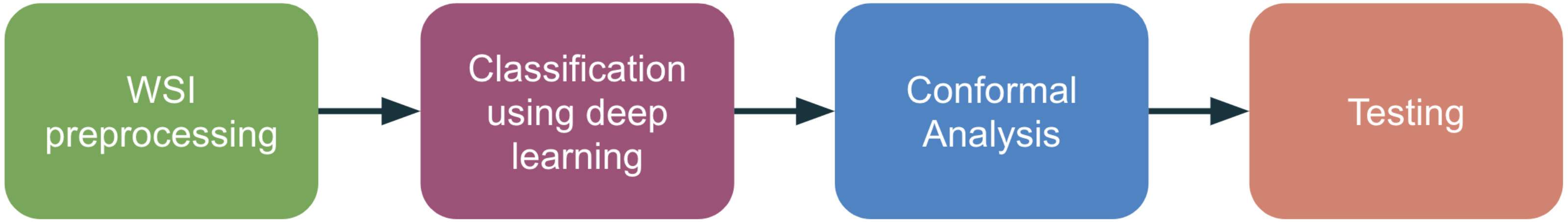
Led by Dr. Chris Banerji, Turing-Roche SRA (Nature Medicine)

Quantifying uncertainty in automated prostate cancer gradation using conformal analysis

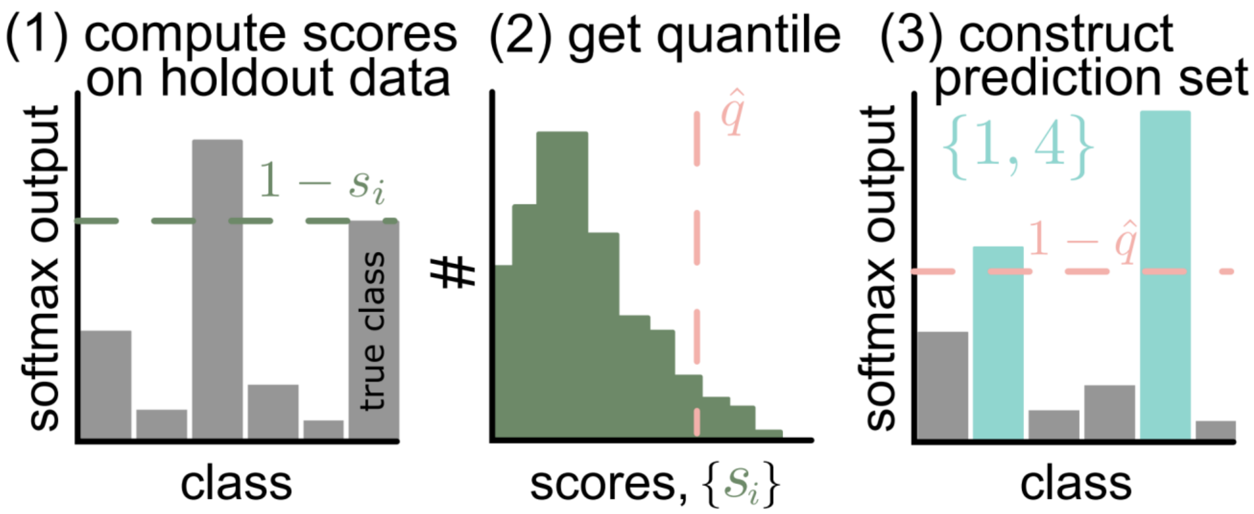


Preprocessed prostate cancer histopathology image from PANDA dataset.

There are two data sources – Karonlisnka (labelled by an expert) and Radboud (labelled by trained students)

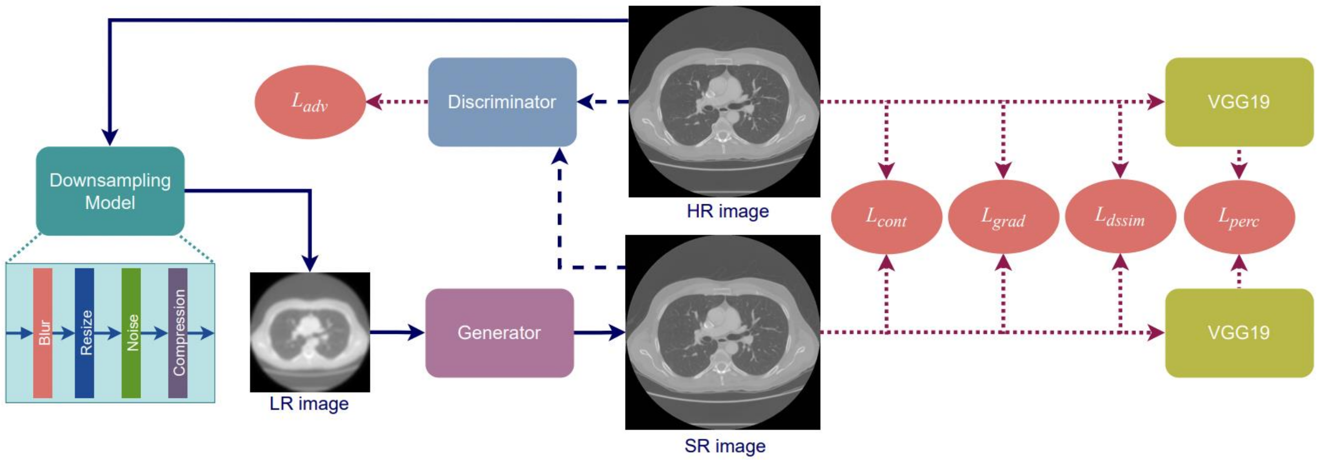


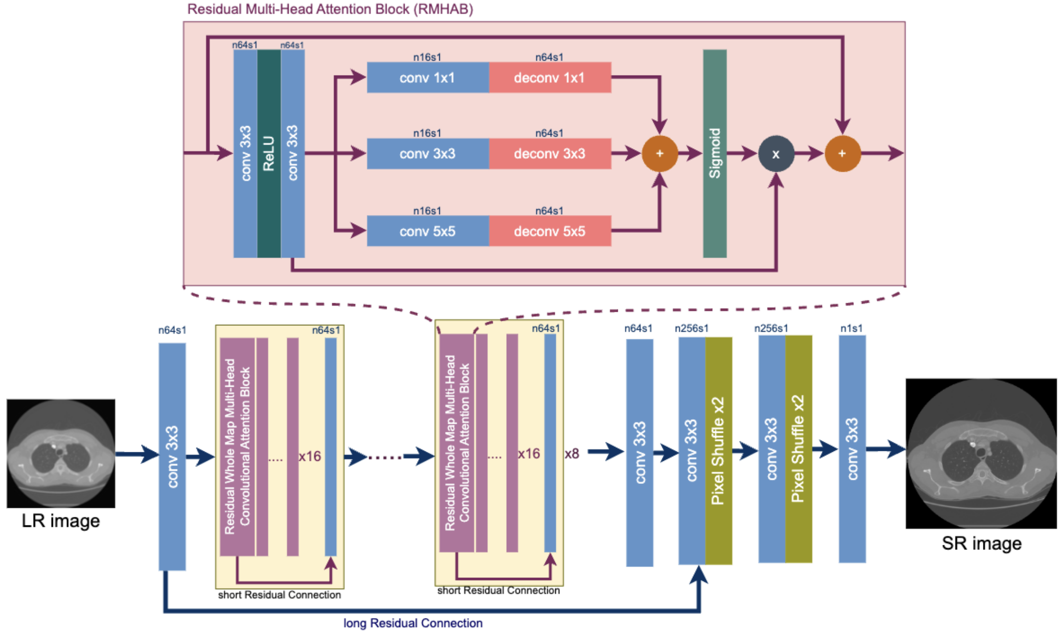
Method pipeline

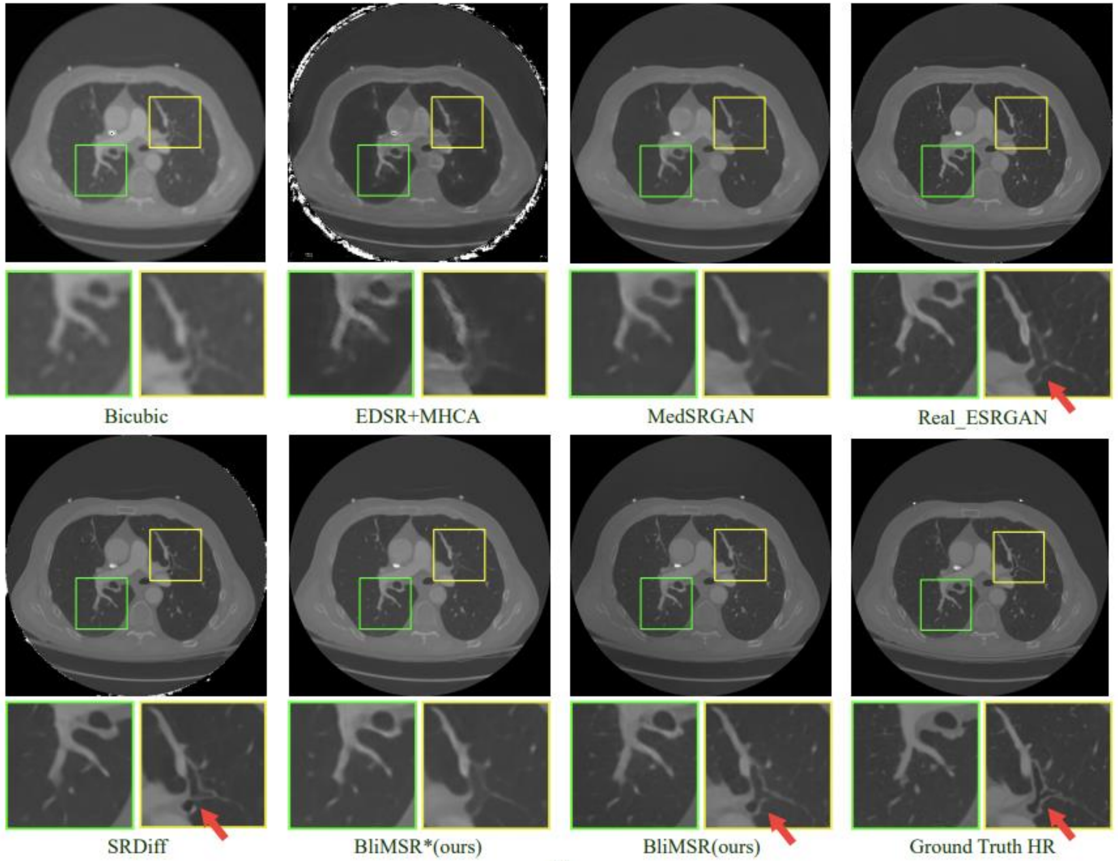


Conformal Prediction

Generating High-Resolution Medical Images







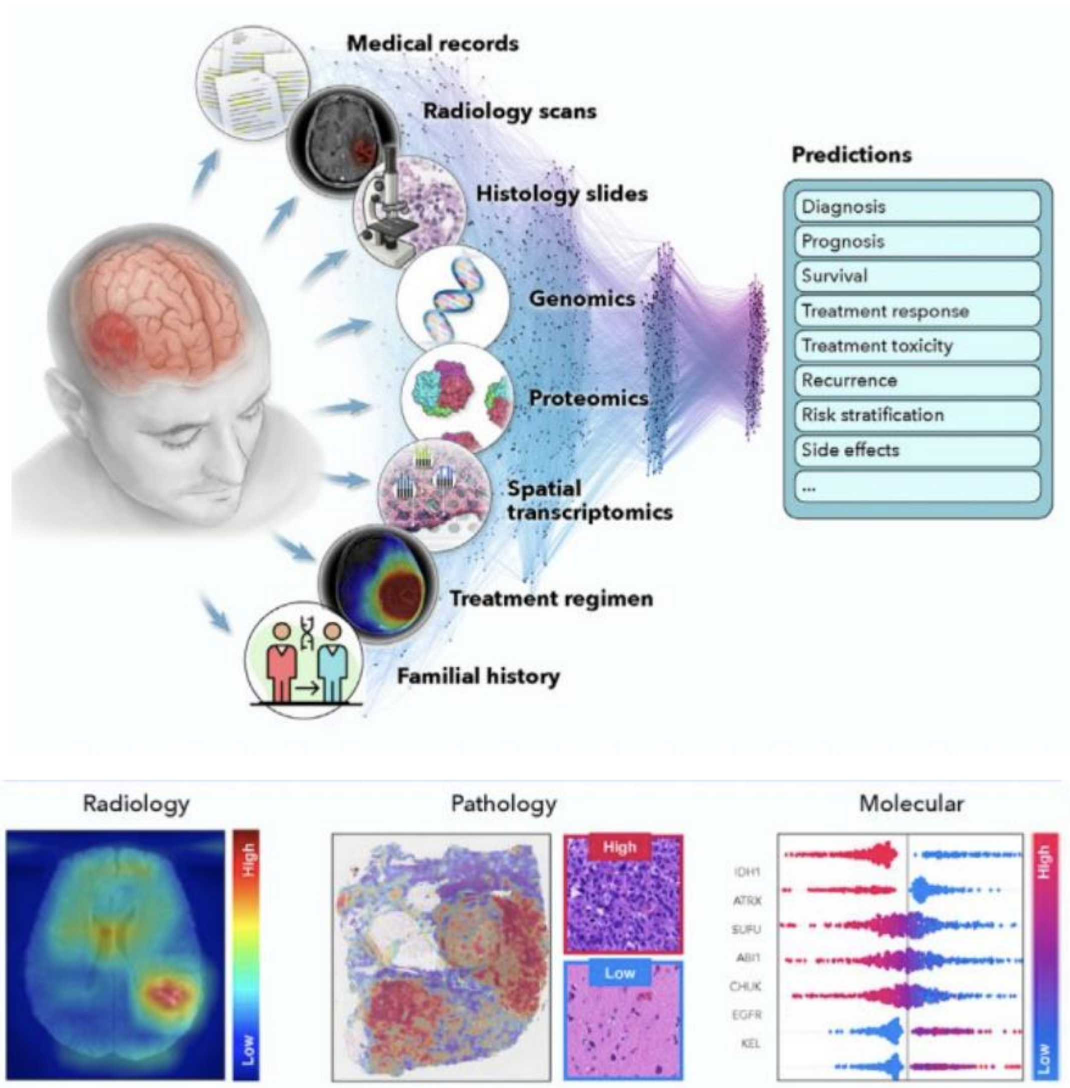
Won award at ICVGIP 2023

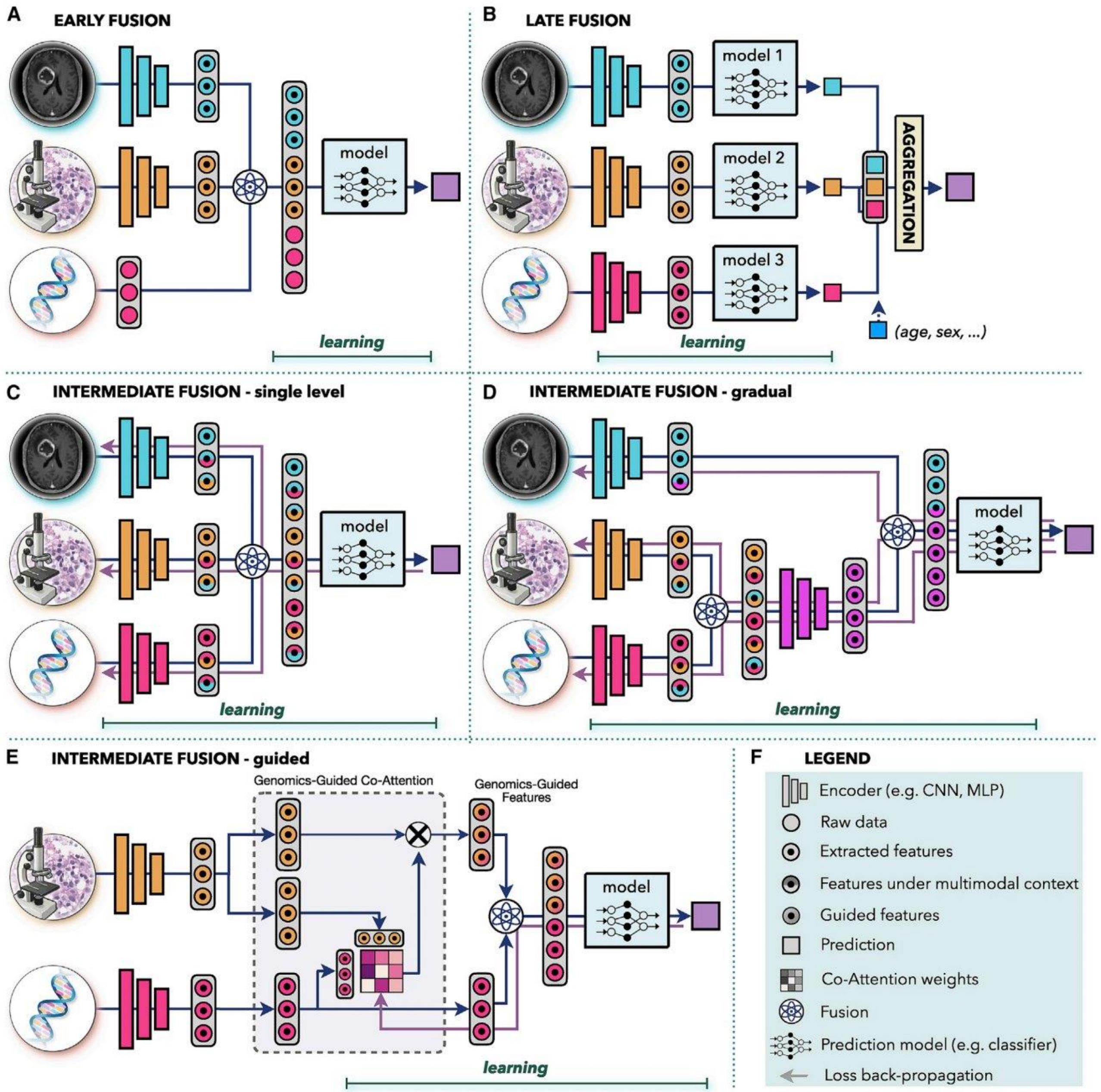
Transparent AI in Health: Grand Challenges

- We have covered several real world examples from the two broad technical approaches to transparent AI, that is explainable models and interpretable results in this talk.

- However as we enter into the era of "Big Multimodal Data" in healthcare, we should be aware of the grand challenges that face us, so as to avoid any pitfalls while expanding the field of transparent AI.

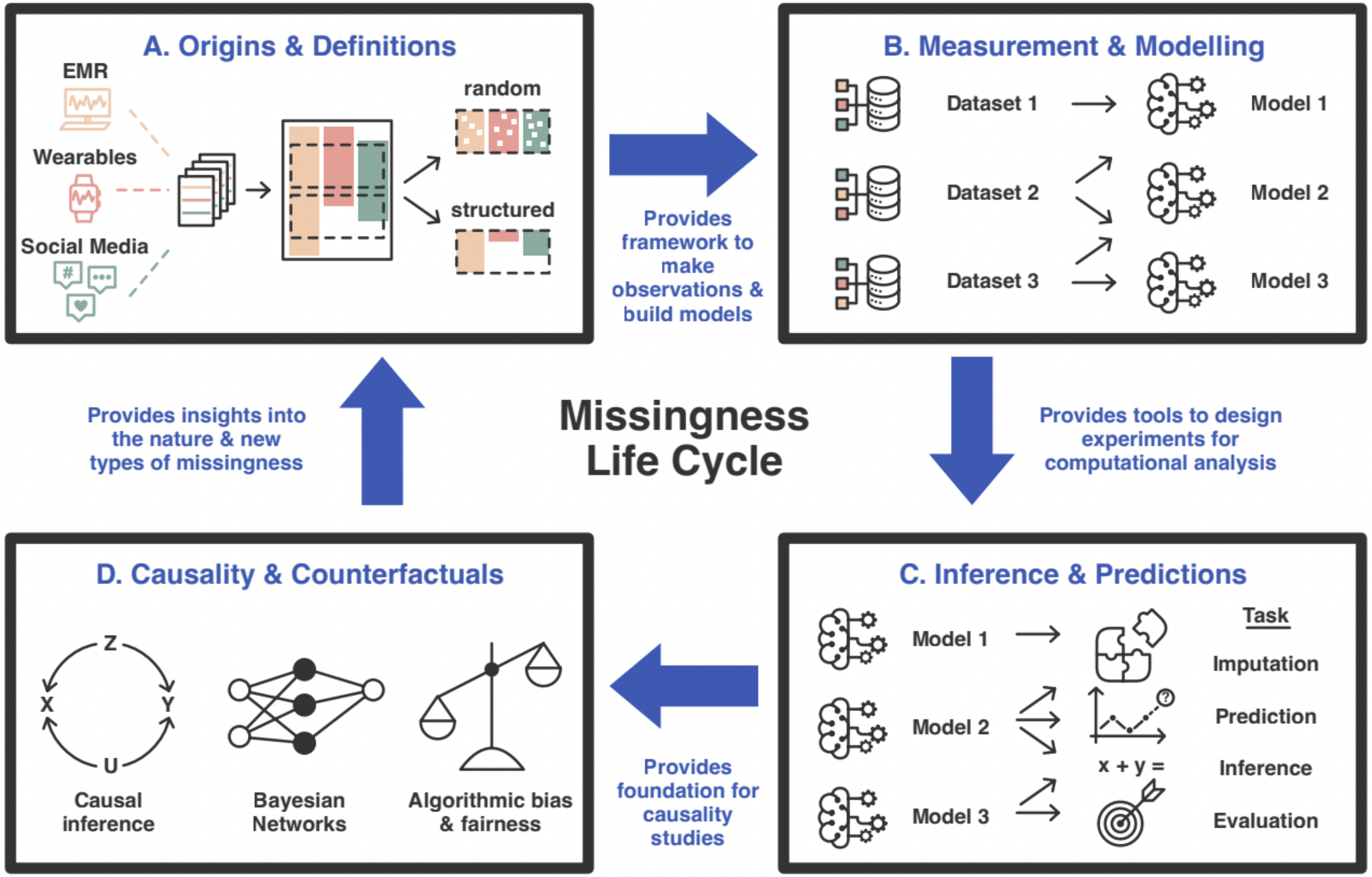
Multi-modal AI





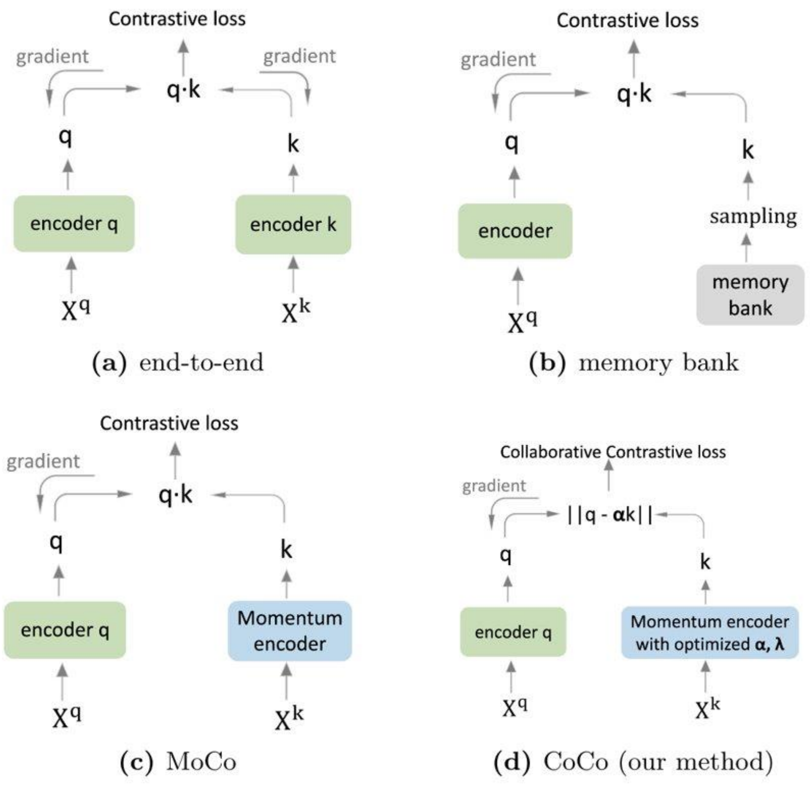
Source: Cancer Cell, Mahmood Lab, 2022

Learning from Data with Structured Missingness

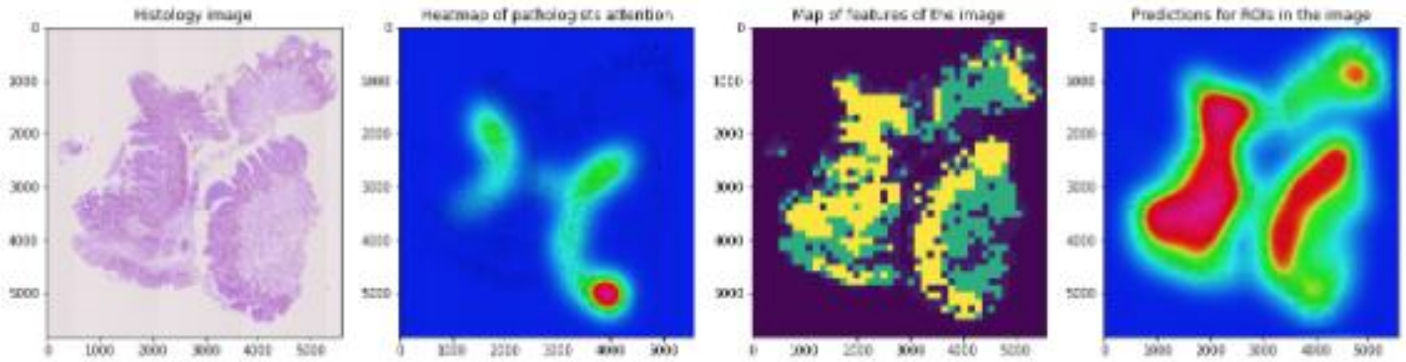


Published in Nature MI, 2023 through the Turing-Roche partnership

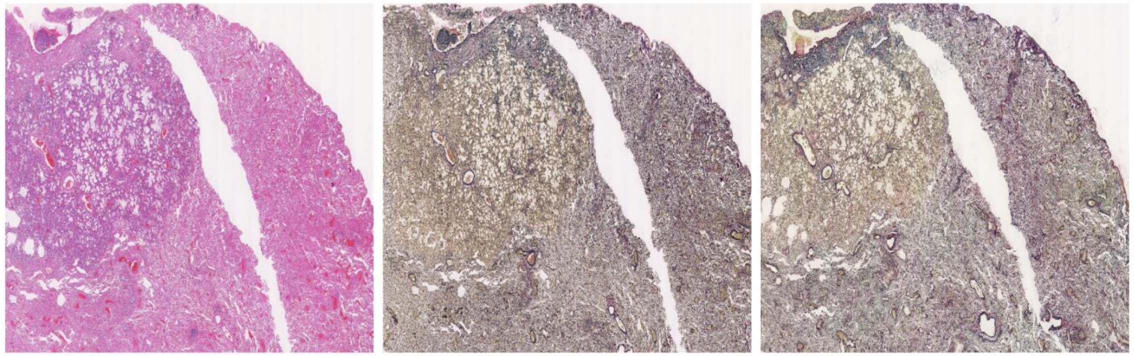
Robust Semi-supervised Continual Learning



Semi-supervised Continual Learning



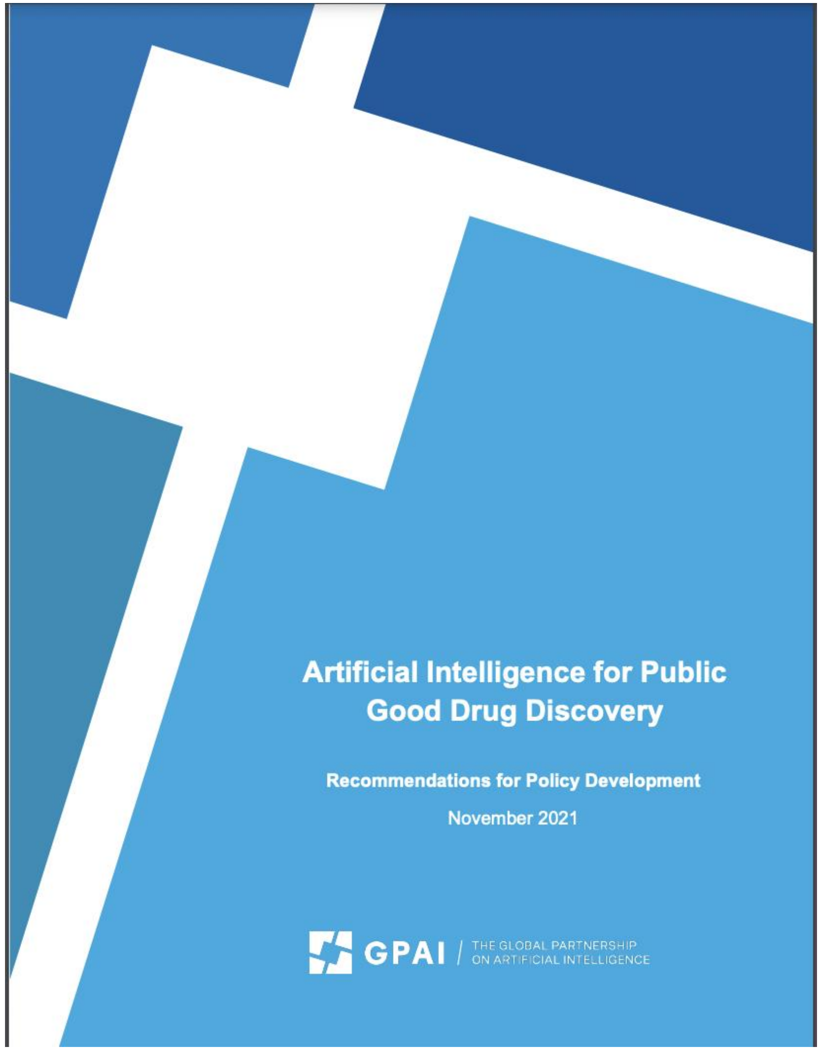
Automated Annotator



Staining Generator

Published in MICCAI, EMBC

Transparent AI for Public Good through Law/Regulations with the Global Partnership on AI (GPAI: https://gpai.ai/)



ceimia

Mila

Work with Yoshua Bengio reflected in EU AI Act