Topics in Machine Learning
Machine Learning for Healthcare

Rahul G. Krishnan

Assistant Professor

Computer science & Laboratory Medicine and Pathobiology

UNIVERSITY OF TORONTO

VECTOR INSTITUTE

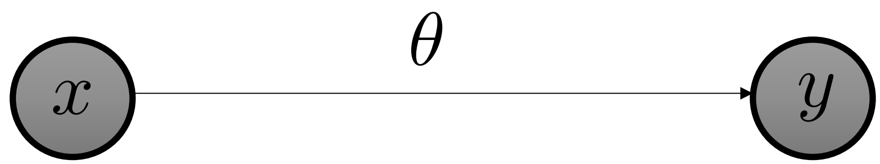
Announcements

- Thank you all for handing in your project assignments on time!
  - Discussion on how best to help overcome hurdles
- Next assignment:
  - October 29 11:59 ET
  - Paper summary assignment [15%]

Outline

- Machine learning for imaging
- Case study 1: Cardiology
- Case study 2: Histopathology
- Biases in medical images

Supervised learning – (1)



$$p(y|x)$$

Patient features — Time to readmission

Chest X-ray image — pneumonia/pneumothorax

- Step 1: Collect a dataset or curate a subset of data with labels from an existing dataset
- Step 2: Learn the model using the dataset
- Step 3: Use the output of the model to build software to help clinicians reach better decisions, faster.
- **Examples**: Logistic regression, random forests, XGBoost, Deep neural networks

Supervised learning – (2)

x_1 y_1 x_2 y_2 x_3 y_3

Dataset (N=3)

- Given a dataset, the model parameters are learned via **maximum likelihood estimation**

$$\mathcal{L}(y, x) = \log p(y|x;\theta)$$

Score function (high is good, low is bad)

$$\theta = \arg\max_{\theta} \sum_{i=1}^{N} \mathcal{L}(y_i, x_i)$$

Solve this optimization problem to **learn** the model. Often formulated as a minimization of the negative of the log-likelihood function

Deep neural networks typically learned using tools that leverage automatic differentiation

PyTorch

Computer vision

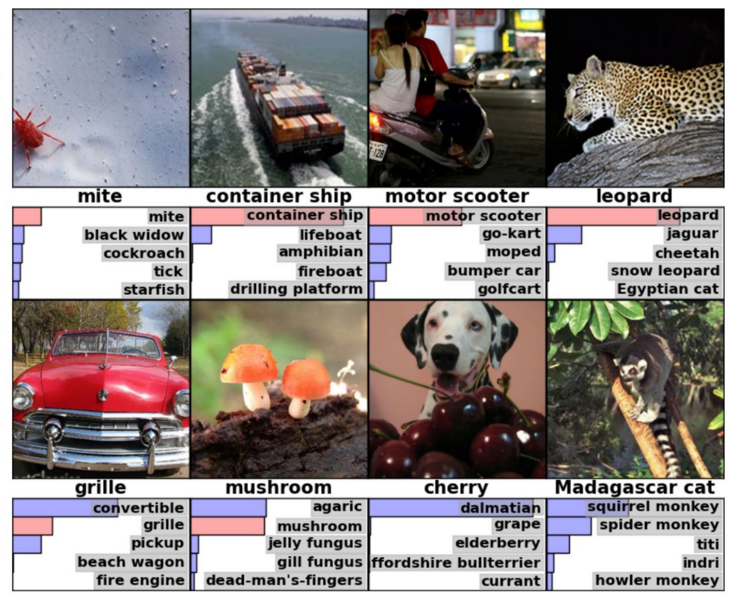
- Computer vision has had a front row seat to the advances in deep learning

ImageNet Challenge

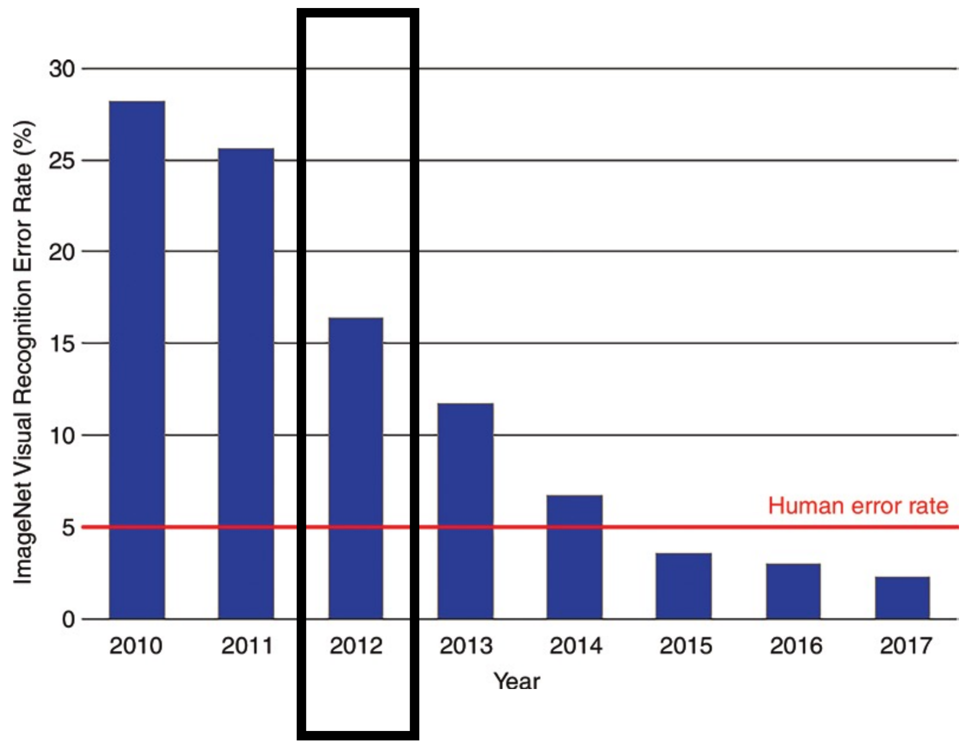
Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009.

IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.

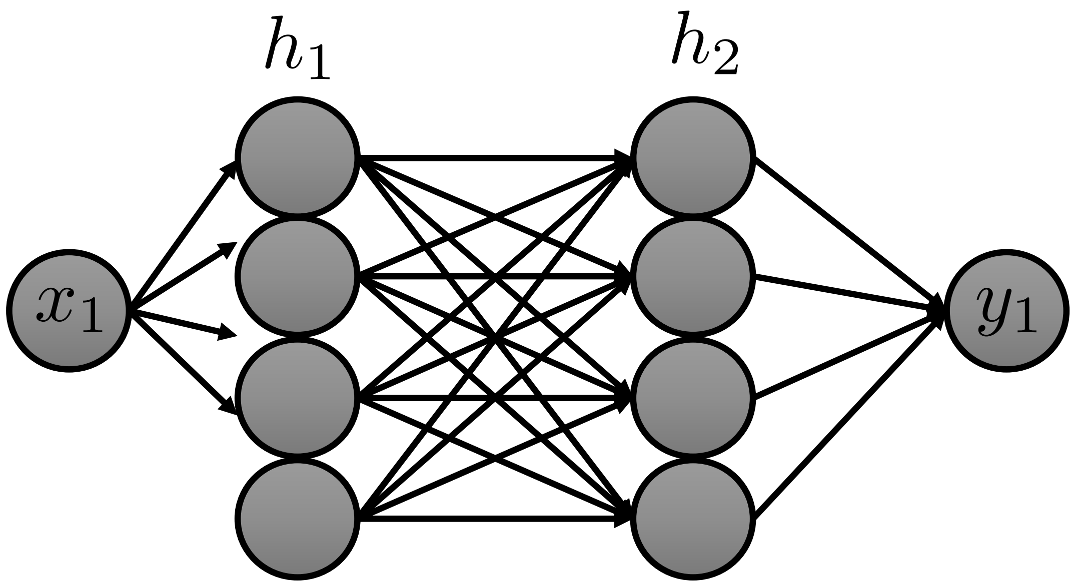


Error rates on Imagenet over time



Neural networks in a slide

- Simplest neural network is a multi-layer perceptron
- Neural networks are known to be universal function approximators

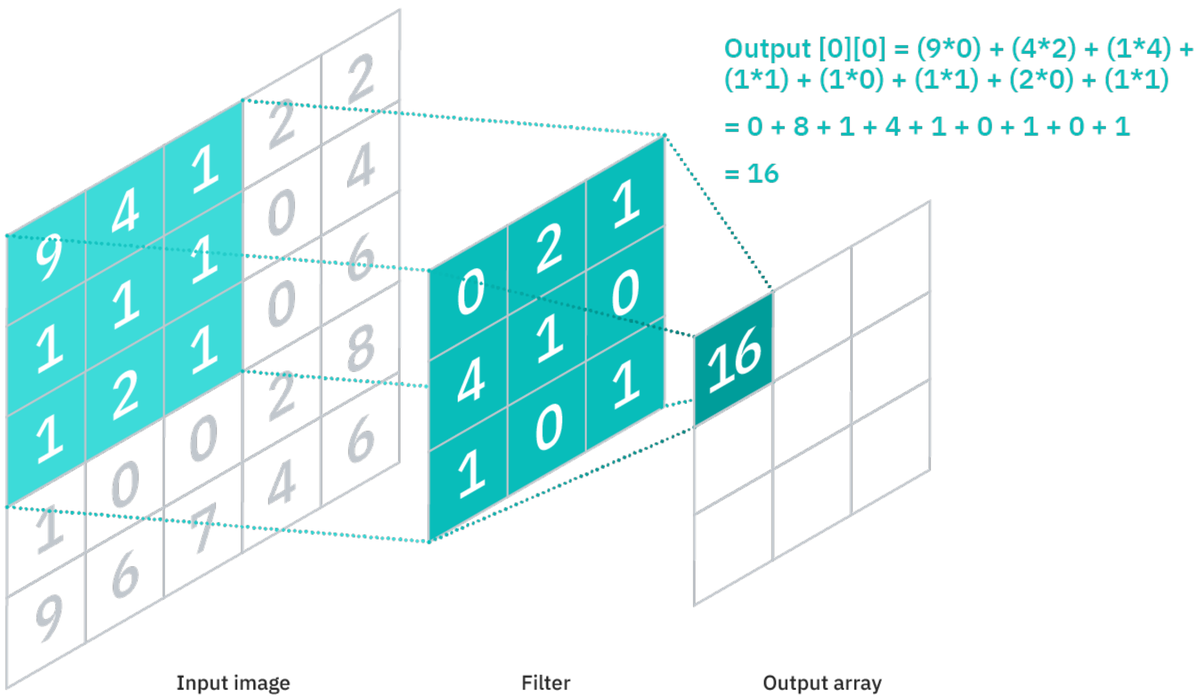


$$p(y|x)$$

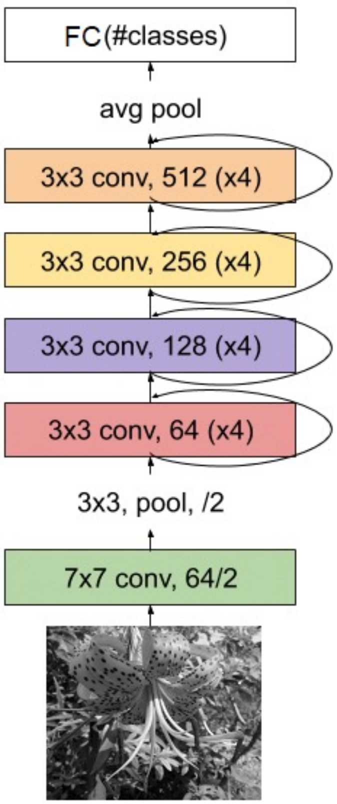
$$y = \phi(W^T h_2 + b)$$
$$h_2 = \phi(W^T h_1 + b)$$
$$h_1 = \phi(W^T x + b)$$

Convolutional neural networks

Capture the fact that we may want representations that are spatially invariant



Deep residual neural networks



Researchers found that deep networks had a hard time learning the identity function.

They added a skip-connections between layers:

$$h_k = \phi(conv(h_{k-1})) + \sum_{j<k-1} h_{k-j}$$

Deep Residual Learning for Image Recognition, He et. al, 2015

Imaging in medicine

- History of Medical Imaging, Bradley et. al, 2008
- Nuclear medicine: Using radiation to *see* inside the human body
  - X-ray discovered in 1895 (won the Nobel in 1901)
  - CT, PET discovered thereafter
- Magnetic resonance imaging: Mapping resonance in the body to images
- Ultrasound imaging: Mapping high-frequency sound waves to images
- Histopathological imaging: Images of stained tissue samples

Decision making with images

- Ultrasound:
  - Echocardiograms
    - Visualize beating of the heart to assess normal function
  - Abdominal ultrasounds
    - Assess healthy function of abdominal organs
- X-rays:
  - Breast cancer screening
  - Guiding surgery to remove blood clots, insert catheters
  - Friday: Hear from Ruizhi Liao on combining text and chest x-ray data

Technical issues in machine learning for medical imaging

- The general setup is almost always as follows:
  - Collect a large set of images [X]
  - Use notes/clinical variables/expert annotation to come up with labels [Y]
  - Use a deep learning model predict Y from X

- Fairness:
  - Reading Race: AI Recognises Patient's Racial Identity In Medical Images, Banerjee et. al, 2021
- Selection bias:
  - Causality matters in medical imaging, Castro et. al, 2019

Case study 1: Deep learning for echocardiograms

- Sound waves to image the heart
- Why:
  - Check for problems with your valves or chambers
  - Check if heart problems are causing shortness of breath
  - Assess congenital heart defects

A taxonomy of echocardiograms

- Most common: Transthoracic echocardiogram
- Transesophageal echocardiogram
  - Transducer guided down patient's throat
  - Records sound waves bouncing off the heart pumping and interprets them as images
- Doppler echocardiogram
  - Used to assess bloodflow
- Stress echocardiogram
  - Ultrasound after excercise

Case study 1: Predicting cardiac amyloidosis

- Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms, Goto et. Al, Nature Communications, 2021
- Cardiac amyloidosis
  - deposition of protein in the heart muscle, can result in heart failure
  - believed to be rare but likely underdiagnosed
  - manifests in both ECGs and echo-cardiography but features are not highly specific and difficult to spot
  - Gold standard: biopsy (costly and risky to patient)

Where machine learning can help

- How can we design a method that:
  - Fits into the clinical workflow for cardiac patients
  - If used, improve underdiagnosis of disease?
- Key-idea: Two-stage approach
  - Step 1: Build ML models from ECG data (readily available at most care providers)
    - Finding: Models have decent accuracy but not enough for conclusive diagnosis
  - Step 2: Build ML models from echocardiogram data
    - Finding: Models outperform human experts
  - Use step 1 to decide which patients should undergo an echocardiogram and apply model from step 2

A multi-center study

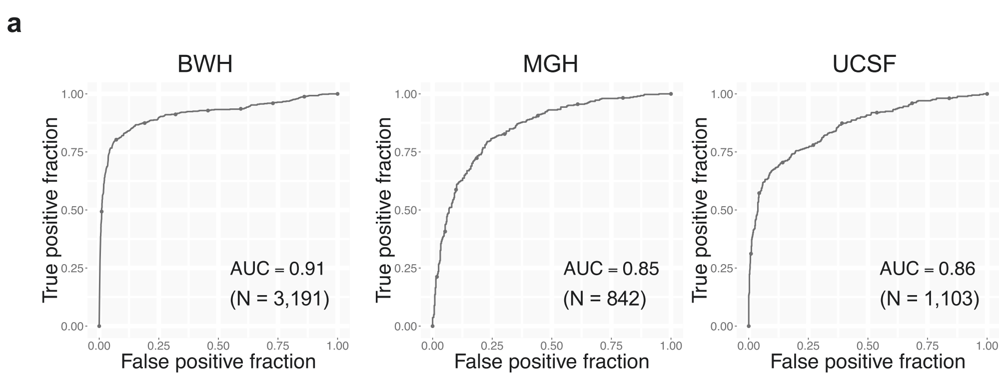
Table 1 Study-level demographic information (ECG cohort).

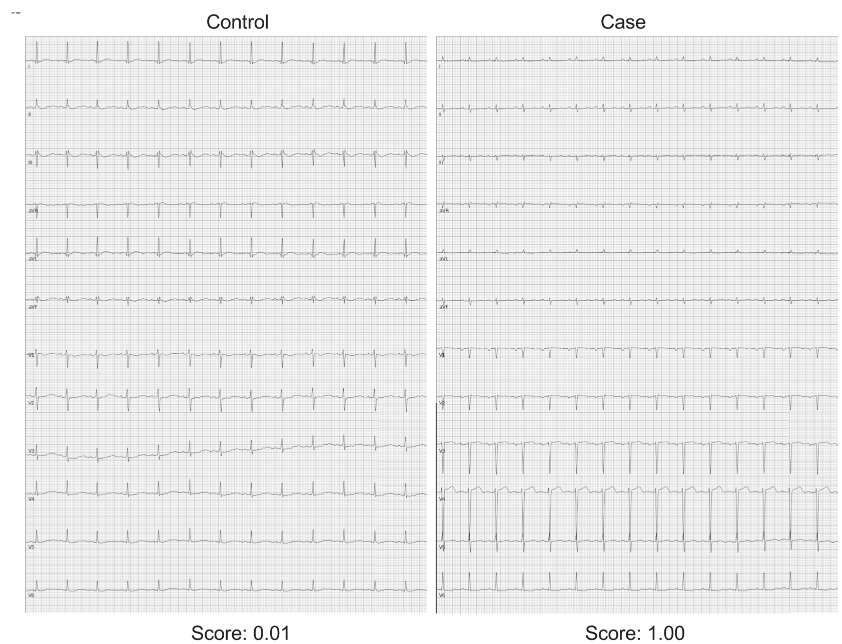
	BWH		MGH		UCSF	
	Case	**Control**	**Case**	**Control**	**Case**	**Control**
Number of studies	2249	8684	405	437	372	731
Age, years ± SD	69.9 ± 10.4	62.3 ± 13.2	72.9 ± 9.0	73.8 ± 8.8	67.7 ± 12.9	67.5 ± 11.7
Age Groups						
≤30, *n* (%)	2 (0.1)	97 (1.1)	1 (0.2)	1 (0.2)	2 (0.5)	0 (0.0)
30–50, *n* (%)	78 (3.5)	1,370 (15.8)	7 (1.7)	6 (1.4)	36 (9.7)	69 (9,4))
50–70, *n* (%)	901 (40.1)	4548 (52.4)	143 (35.3)	135 (30.9)	136 (36.6)	278 (38.0)
70–90, *n* (%)	1242 (55.2)	2606 (30.0)	254 (62.7)	295 (67.5)	198 (53.2)	384 (52.5)
>90, *n* (%)	26 (1.2)	63 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
HR, bpm ± SD	76.4 ± 16.7	75.9 ± 18.5	78.6 ± 16.6	75.1 ± 19.8	79.6 ± 18.7	72.2 ± 16.3
Sinus rhythm, *n* (%)	1,736 (77.2)	8,072 (93.0)	283 (69.9)	371 (84.9)	365 (98.1)	729 (99.7)

HR heart rate, *BWH* Brigham and Women's Hospital, *MGH* Massachusetts General Hospital, *UCSF* University of California San Francisco. *N* represents the number of studies.

Step 1: ECG model

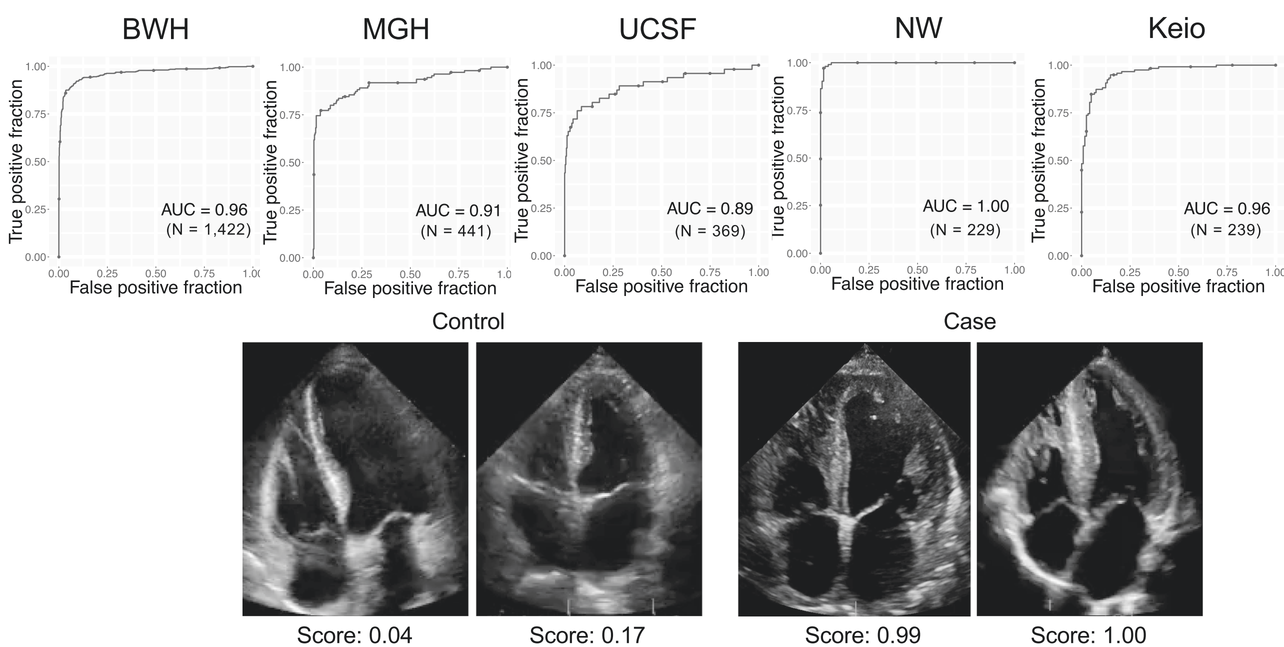
- Results Ok but not considered good enough for evaluating interventions for a rare diagnosis since it will result in a large number of false positives





Step 2: Echocardiogram model

- Performance significantly better when using a richer (but more expensive) data modality



Recall: Metrics

- Positive Predictive Value (PPV): TP/(TP+FP)
  - A high PPV will indicate that a positive result is likely correct
- Sensitivity: TP/(TP+FN)
  - A highly sensitive test will have few-false negatives

Analyzing the combined approach

- ECG model:
  - MGH: PPV 3.9% with Sensitivity 71%
  - BWH: PPV 3.4% with Sensitivity 52.4%
- Echo model:
  - MGH: PPV 32.7% with Sensitivity 66.9%
  - BWH: PPV: 33.4% with Sensitivity 67%
- Combined:
  - MGH: PPV: 76.6% with Sensitivity 47.5%
  - BWH: PPV: 73.9% with Sensitivity 34.8%

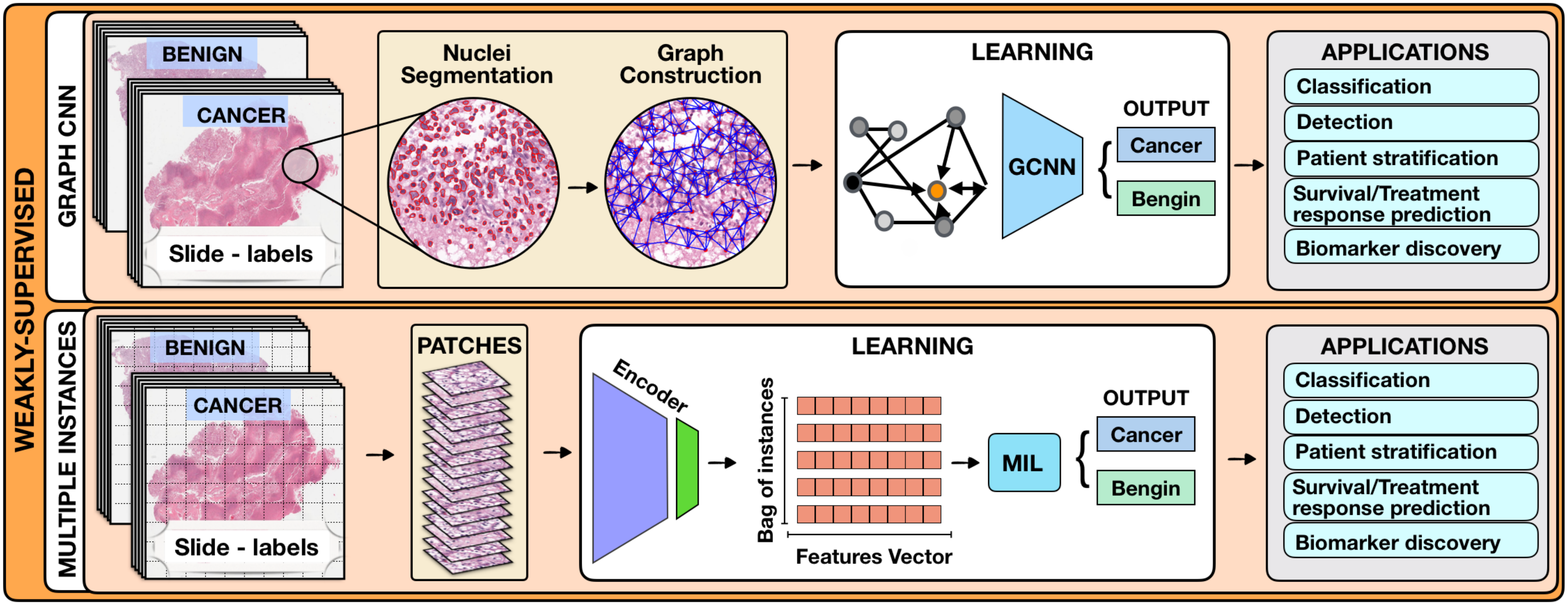
Case study 2: Deep learning for histopathological image data

- Research by Richard J. Chen
- 3rd year Ph.D. Candidate, Harvard University / BWH, Broad Institute
- Work in submission

Histopathological images in the clinical workflow

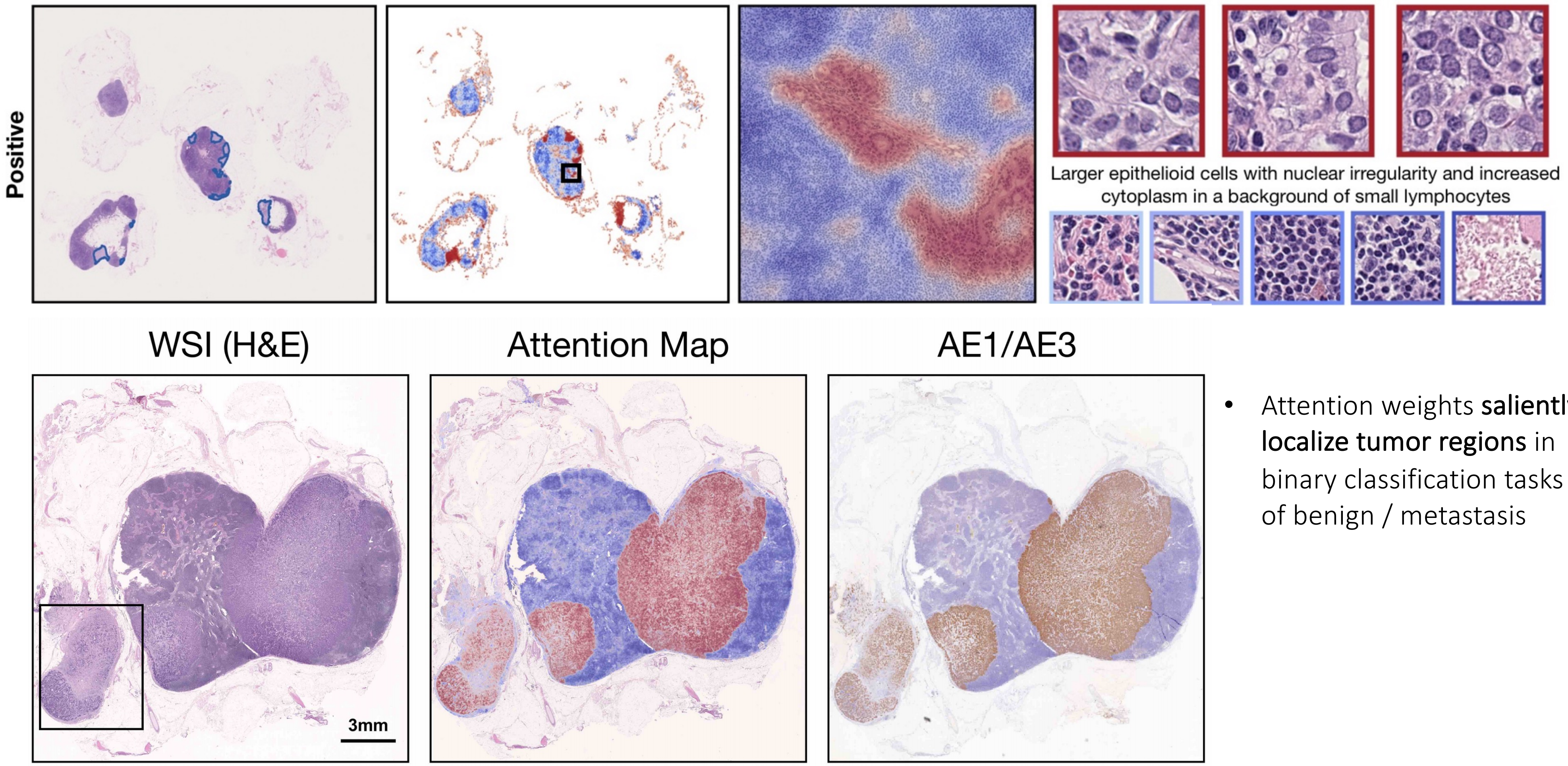
- Histopathology: Microscopic examination of tissue to study diseases and their different presentations,
- Pipeline:
  - Surgery, biopsy or autopsy for excision of tissue
  - Placed in a fixative to stabilize tissue
  - Investigated under a microscope
- Histopathological images are routinely used for clinical diagnoses of cancer
  - **Key question: How can we use machine learning to build representations of histopathological image data?**

Slide-Level Supervised Learning (Weak Supervision)



Lipkova *et al.* 2021, In Review

Weakly-Supervised Learning: Finding Needles in Haystacks via Attention



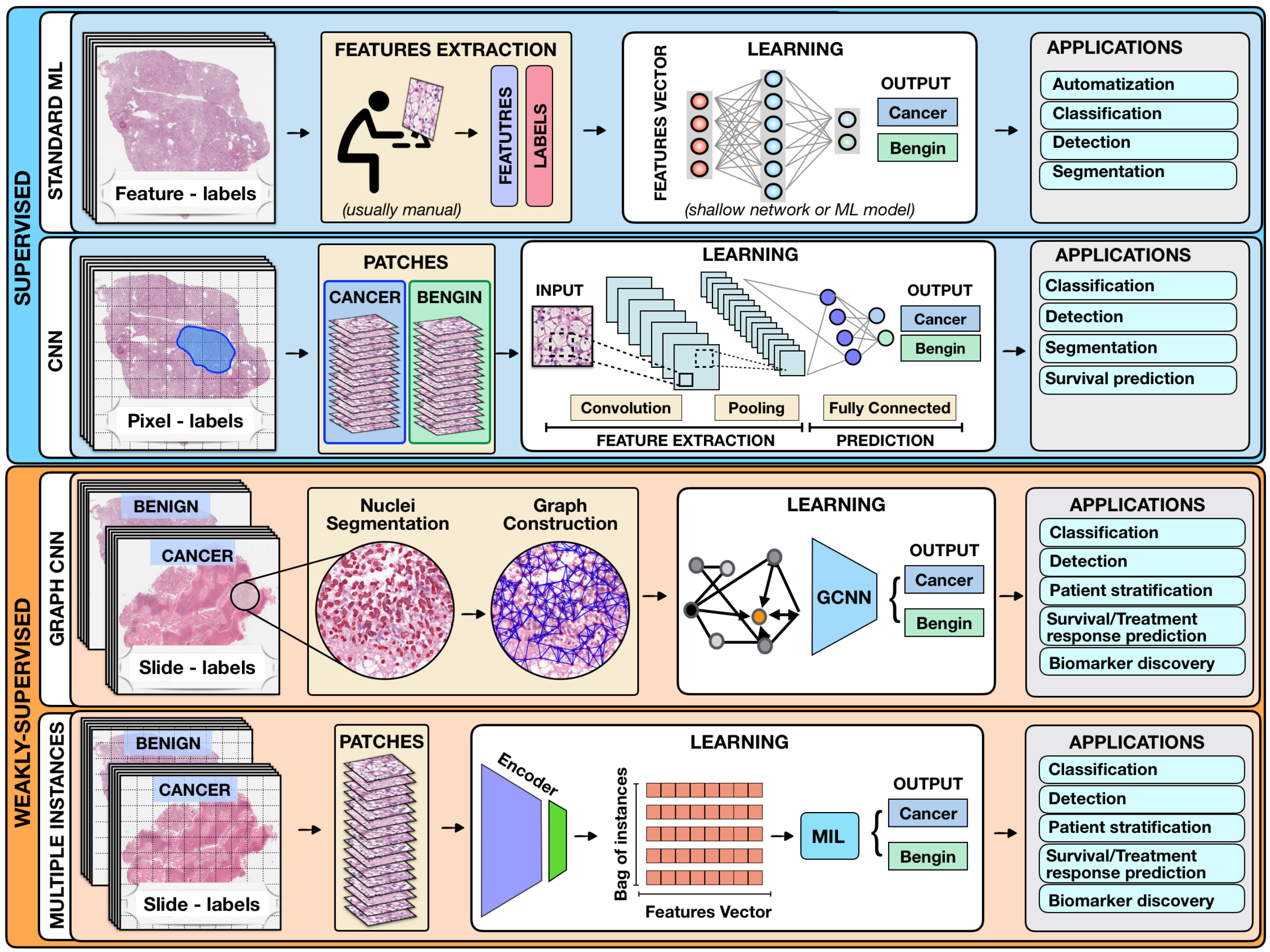
Lu *et al.* 2021, Nature BME

- Attention weights **saliently localize tumor regions** in binary classification tasks of benign / metastasis

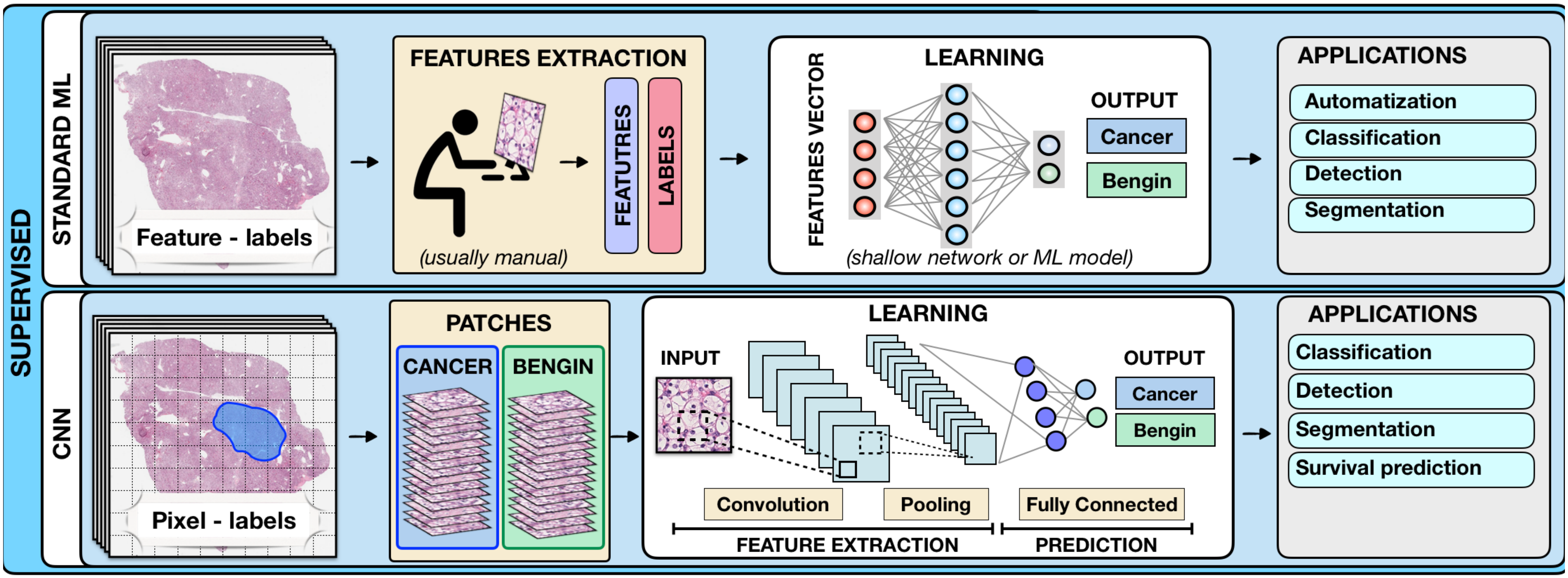
Current Paradigm is limited by: Clinical Domain Knowledge

- Requires clinical domain knowledge to:
  1. label image regions in WSIs with known morphological phenotypes **(patch-level tasks)**
  2. Make prognostic decisions from subjective interpretation of the entire WSI **(slide-level tasks)**

- How can we identify new phenotypic biomarkers?
- What are we missing in current decision-making that can guide prognosis?

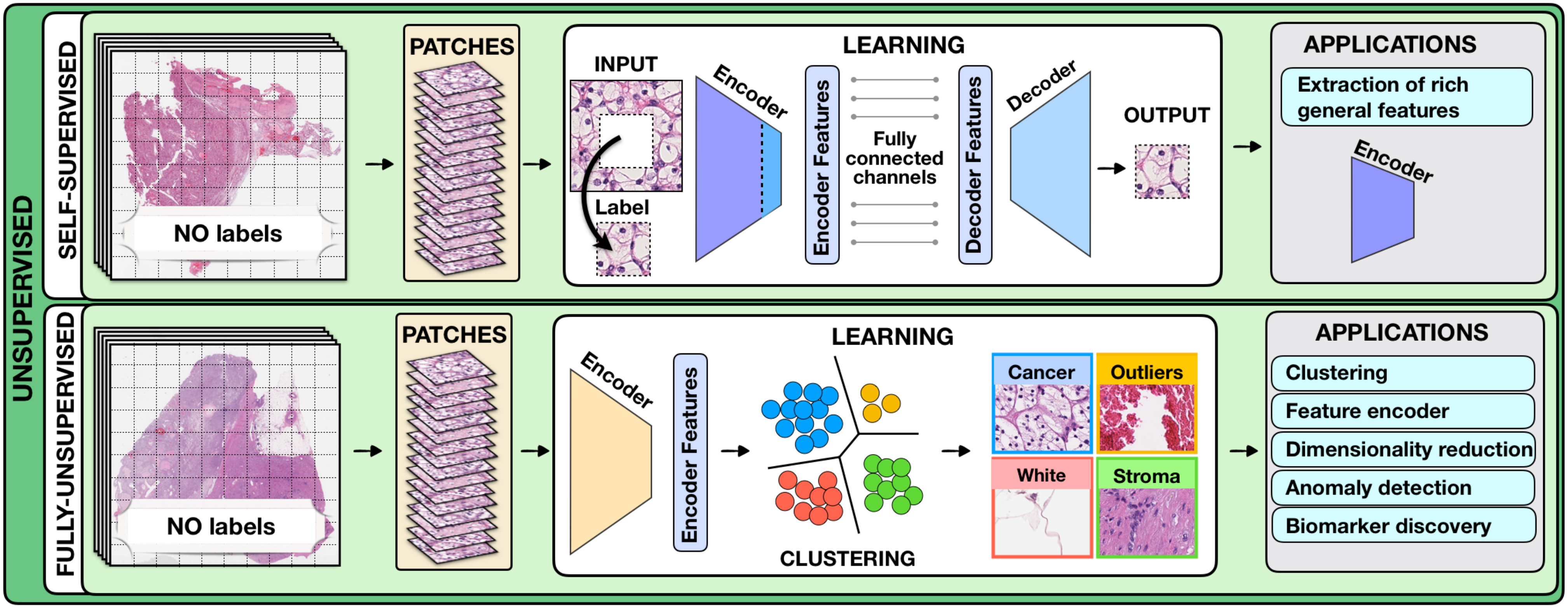


Current Paradigm is limited by: Clinical Domain Knowledge



Current pipelines for creating representations of whole slide images make use of ResNet50 architectures pretrained on imagenet.

Self-Supervised Learning: Pixel-Level Annotations are Not Needed!



Lipkova *et al.* 2021, In Review, Ciga et. al

We build upon recent work [Resource and data efficient self supervised learning, Ciga et. al, 2021] who show that self-supervision yields general purpose representations of histopathological images