# Vision Transformers for Analyzing High-Resolution Pathology Images

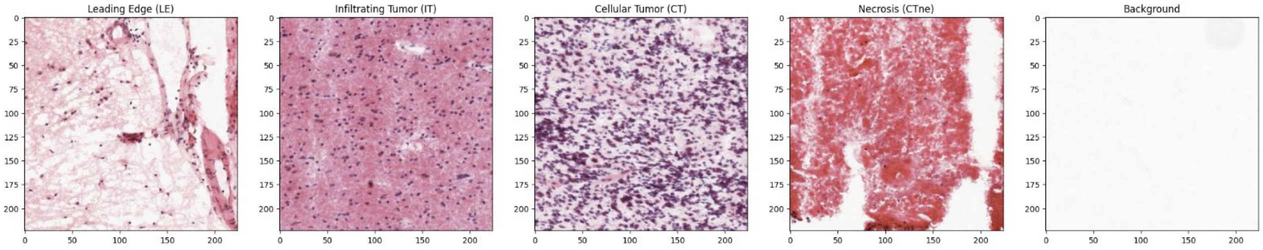
Raphaël Attias

- **Harvard**
  Yu Lab
- **EPFL**
  MIPLAB

03/24/2023, Boston MA

# Glioblastoma

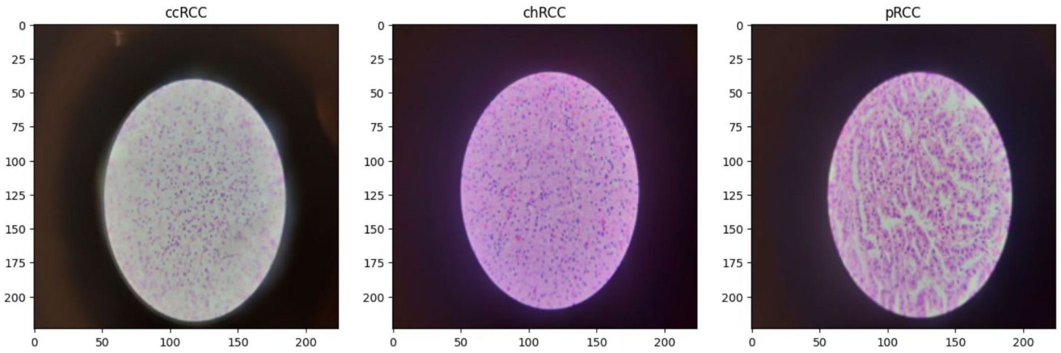
- Highly aggressive form of brain cancer with a low survival rate.
- Accounts for **15% of all brain tumors** and results in the death of over **200,000 individuals** annually.
- The median survival rate for patients receiving aggressive treatment is **12-15 months**, and the five-year **survival rate is only 6.8%**.
- In 2019, the Ivy Glioblastoma Atlas (**IvyGAP**) was introduced.
- IvyGAP cohort consisted of 41 patients, and their 42 tumors were used to generate various data types, including MRI scans, machine learning-annotated images, and clinical information.



(a) Leading Edge (b) Infil. Tumor (c) Cellular Tumor (d) Necrosis (e) Background

# Renal Cancer

- Renal cell carcinoma is the most prevalent type of kidney cancer, **accounting for 2-3% of cancers in humans.**
- Three main subtypes based on cell appearance: clear cell RCC (ccRCC), papillary RCC (pRCC), and chromophobe RCC (chRCC).
- Treatment options for RCC include surgery, radiation therapy, chemotherapy, immunotherapy, and targeted therapy.
- In 2018, approximately **403,300 new cases of kidney cancer** were reported worldwide, resulting in **175,000 deaths**.
- Dataset by the Yu Lab, which took pictures of 4694 slides with a cellphone in a microscope.



(a) Clear cell RCC (b) Chromophobe RCC (c) Papillary RCC

# Table of Content

1. **Self-Supervised Learning for robust pretraining**
   *Pretraining without labels improve model performance.*

2. **Finding Regions of Interests in Whole Slide Images**
   *Finding ROIs with Supervised and Unsupervised methods.*

3. **Uncertainty Estimation with Posterior Network**
   *Estimating model and labels confidence.*

# 1. Self-Supervised Learning for robust pretraining

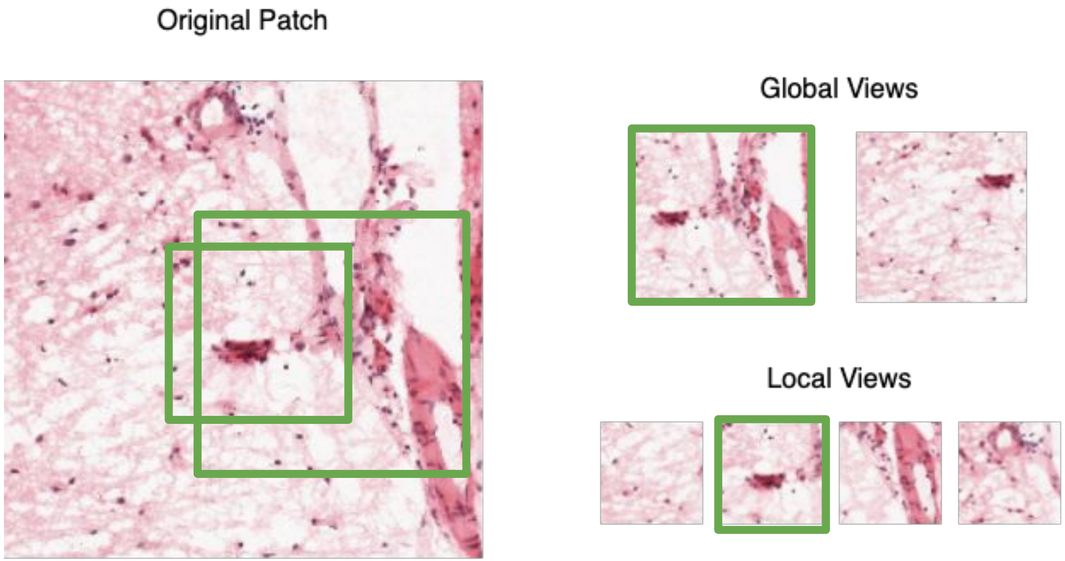
**Problem:** Most foundations models are pre-trained on general vision datasets (ImageNet), how can we leverage task specific datasets?

**Idea:** Learn data representation by using Self-Supervised methods that leverage large unlabeled datasets.

**Application**: Provide generalist models for downstream tasks (segmentation, classification…) or finding ROIs without labels.

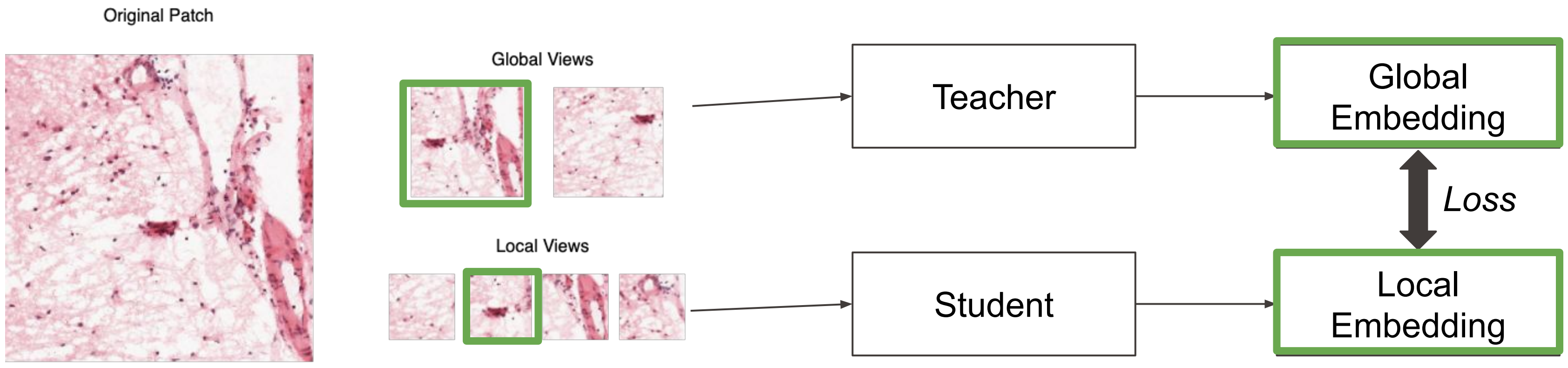
# 1. Self-Supervised Learning for robust pretraining

## DINO: Self-Supervised Learning with distilled knowledge



# 1. Self-Supervised Learning for robust pretraining

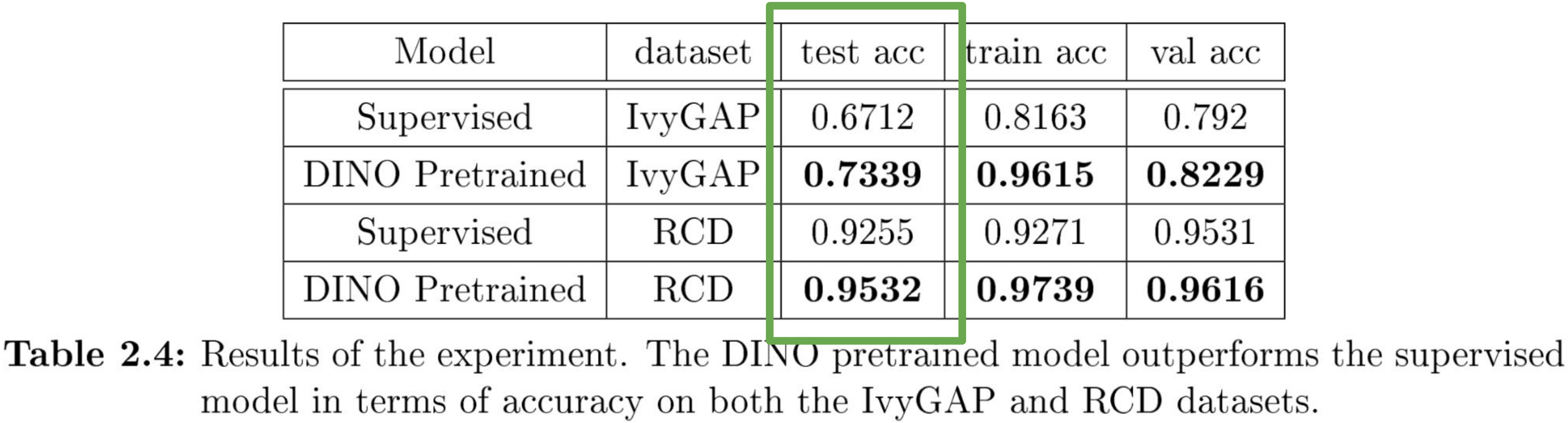
## DINO: Self-Supervised Learning with distilled knowledge



**Training is over when teacher and student provide same embedding.**
**Student/Teacher becomes a backbone for downstream tasks.**

# 1. Self-Supervised Learning for robust pretraining

## Models pretrained with DINO outperforms supervised models

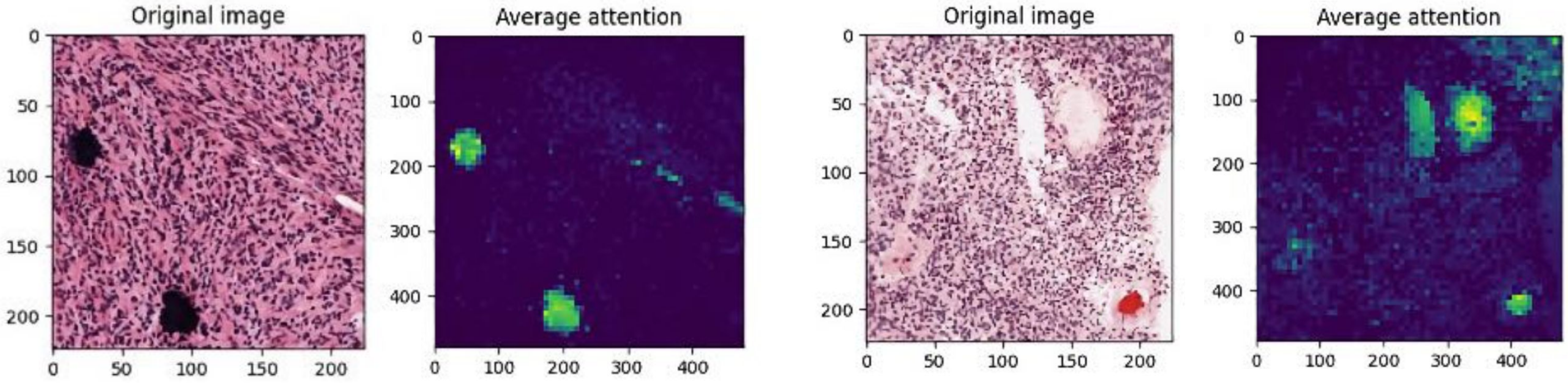


| Model | dataset | test acc | train acc | val acc |
|---|---|---|---|---|
| Supervised | IvyGAP | 0.6712 | 0.8163 | 0.792 |
| DINO Pretrained | IvyGAP | **0.7339** | **0.9615** | **0.8229** |
| Supervised | RCD | 0.9255 | 0.9271 | 0.9531 |
| DINO Pretrained | RCD | **0.9532** | **0.9739** | **0.9616** |

**Table 2.4:** Results of the experiment. The DINO pretrained model outperforms the supervised model in terms of accuracy on both the IvyGAP and RCD datasets.

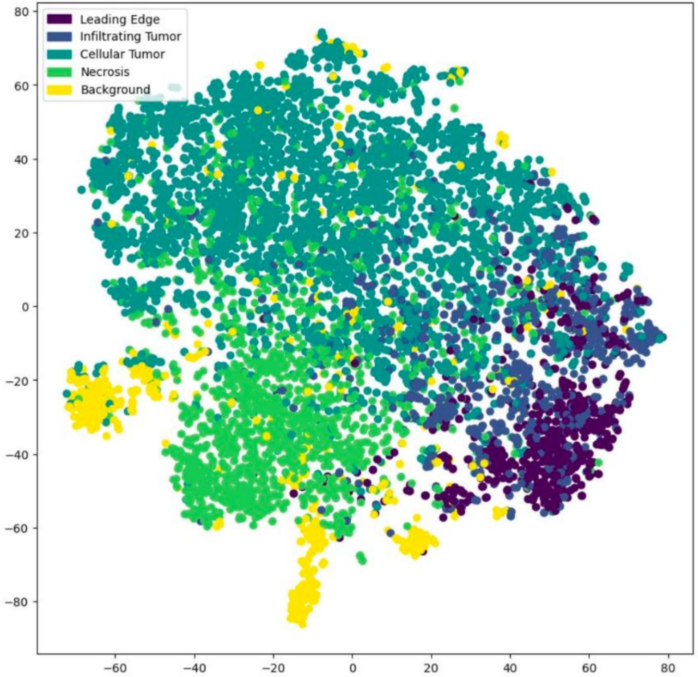
# 1. Self-Supervised Learning for robust pretraining

## Vision Transformers pre-trained with DINO provide quality attention maps



# 1. Self-Supervised Learning for robust pretraining

**Models pretrained with DINO provide embeddings with levels of semantic.**



(a) ViT embeddings with DINO

# 1. Self-Supervised Learning for robust pretraining

**Outcomes:**

- Pre-training without labels
- Better on downstream tasks
- Can provide attention maps with Vision Transformers
- Produce semantic embeddings

→ Robust foundation model for downstream tasks on IvyGap and the Renal Cancer dataset.

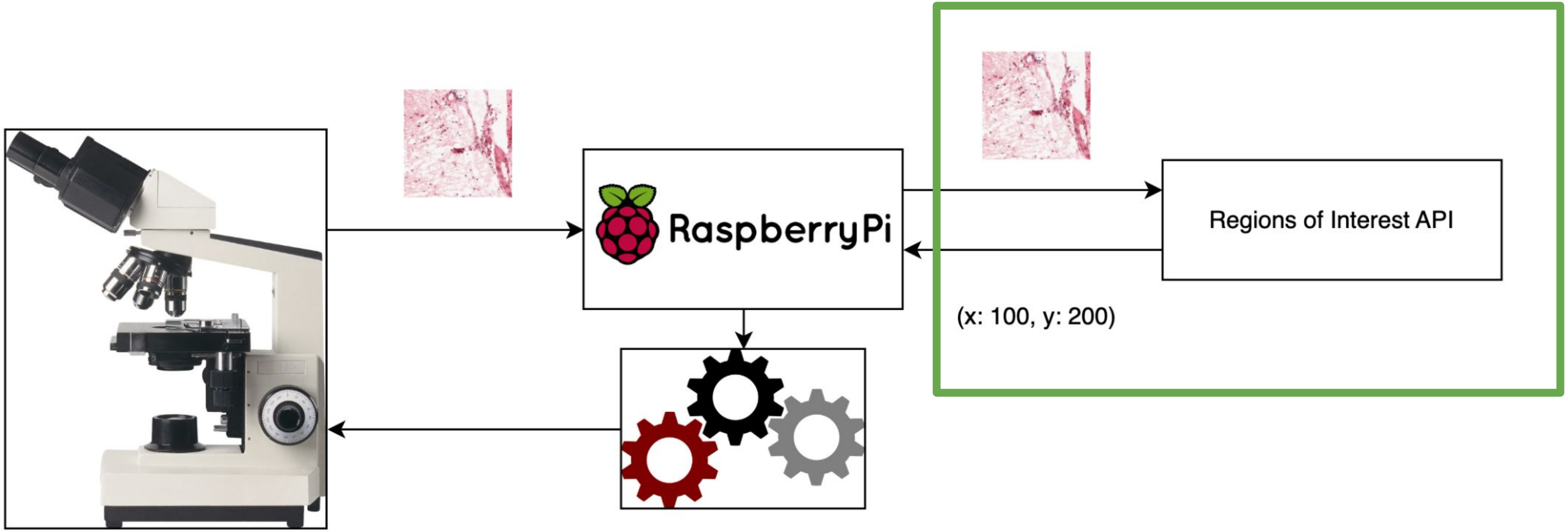
# 2. Finding Regions of Interests in WSI

**Motivations:**

- Critical component of a potential surgical or pathological pipeline for treating tissue.
- Component in the Yu Lab project for automated scanning, focus and detection on histopathology slides using a Raspberry Pi.

# 2. Finding Regions of Interests in WSI

Component in the Yu Lab project for automated scanning, focus and detection on histopathology slides using a Raspberry Pi.



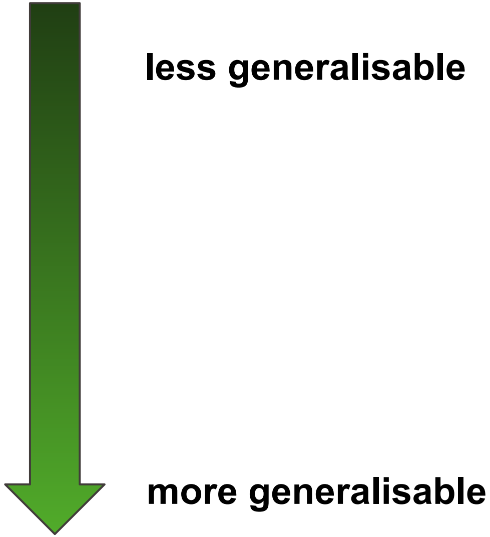
# 2. Finding Regions of Interests in WSI

**We have studied 3 approaches:**

A. **Strongly Supervised**: applying segmentation models when ground truth maps are available.

B. **Weakly Supervised**: employing interpretability methods on models trained for patch classification.

C. **Self-Supervised:** using the attention maps of Vision Transformers pretrained with DINO.



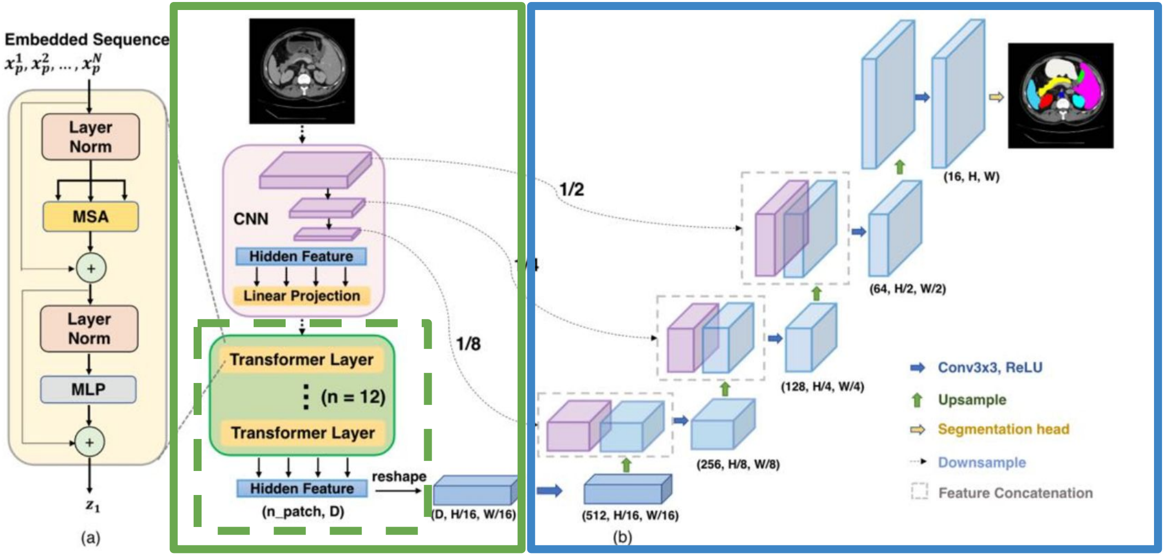
**less generalisable**

**more generalisable**

# 2. Finding Regions of Interests in WSI

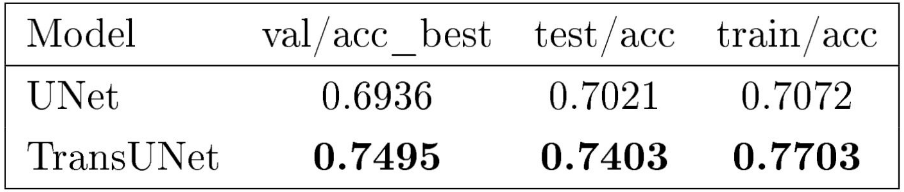
## A. Strongly Supervised: TransUNet

- Trained with ground truth segmentation maps.
- Hybrid between a Transformer and UNet architecture.



# 2. Finding Regions of Interests in WSI

## A. Strongly Supervised: TransUNet



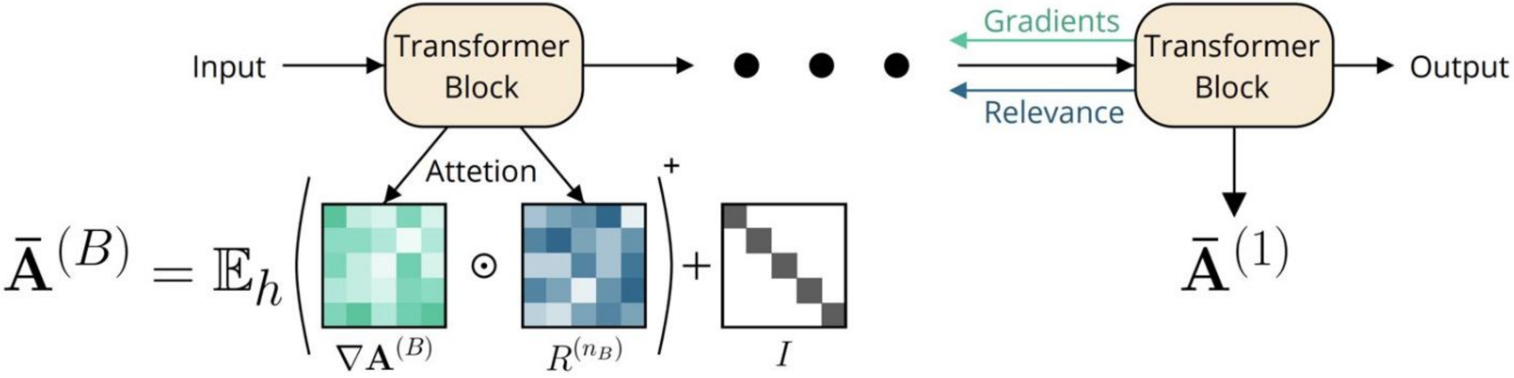
| Model | val/acc_best | test/acc | train/acc |
|---|---|---|---|
| UNet | 0.6936 | 0.7021 | 0.7072 |
| TransUNet | **0.7495** | **0.7403** | **0.7703** |

- **TransUNet outperforms UNet architecture on WSI segmentation.**
- **Make use of the skip connections of UNet and local features.**
- **Leverage global and context relations thanks to Transformers.**

# 2. Finding Regions of Interests in WSI

## B. Weakly Supervised: Transformer Interpretability

- Find regions that are relevant for patch classification
- Gradient based methods to find activations in Attention heads

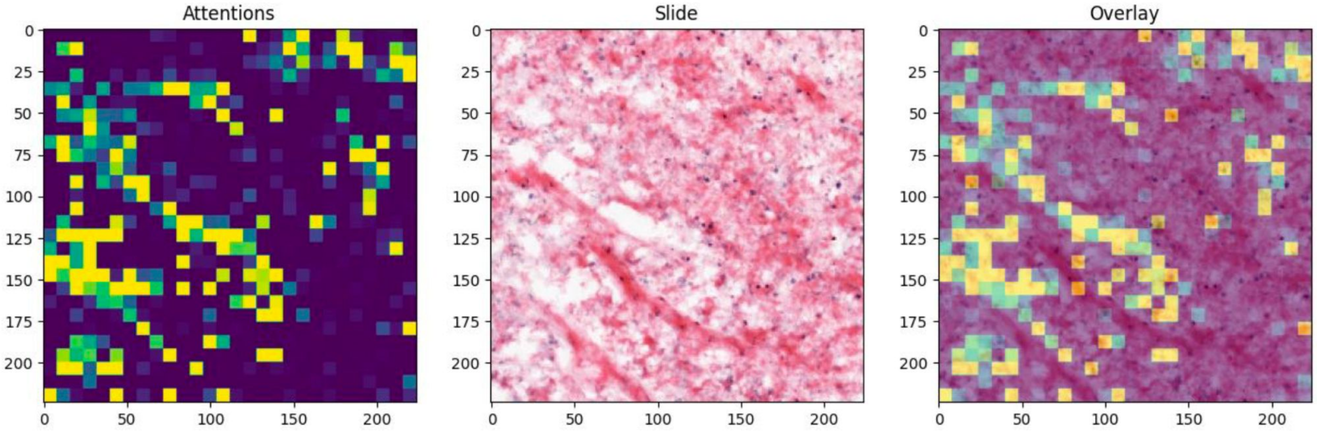


**Relevancy Map** $= \bar{\mathbf{A}}^{(1)} \cdot \bar{\mathbf{A}}^{(2)} \cdot \ldots \cdot \bar{\mathbf{A}}^{(B)}$

# 2. Finding Regions of Interests in WSI

## B. Weakly Supervised: Transformer Interpretability

**Transformer Interpretability produces relevant attention maps without the need of segmentation maps.**

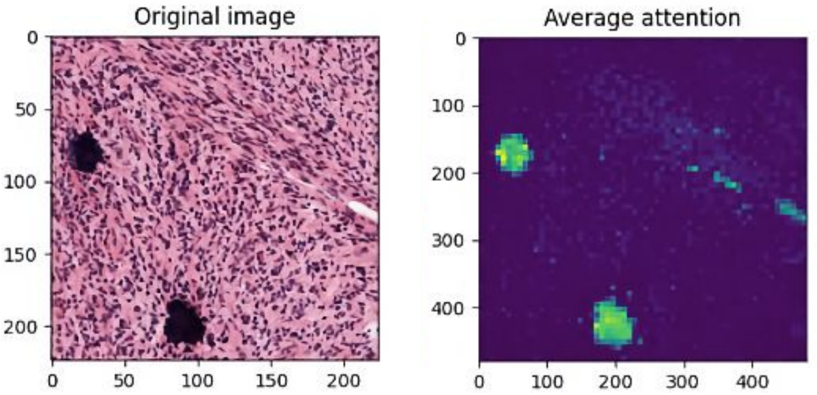


(a) Necrosis patch, and the gradient-based relevancy map. We see that the image features that correlates strongly with Necrosis are the absence of tissue in the patch.

# 2. Finding Regions of Interests in WSI

## C. Self-Supervised: DINO and Vision Transformers

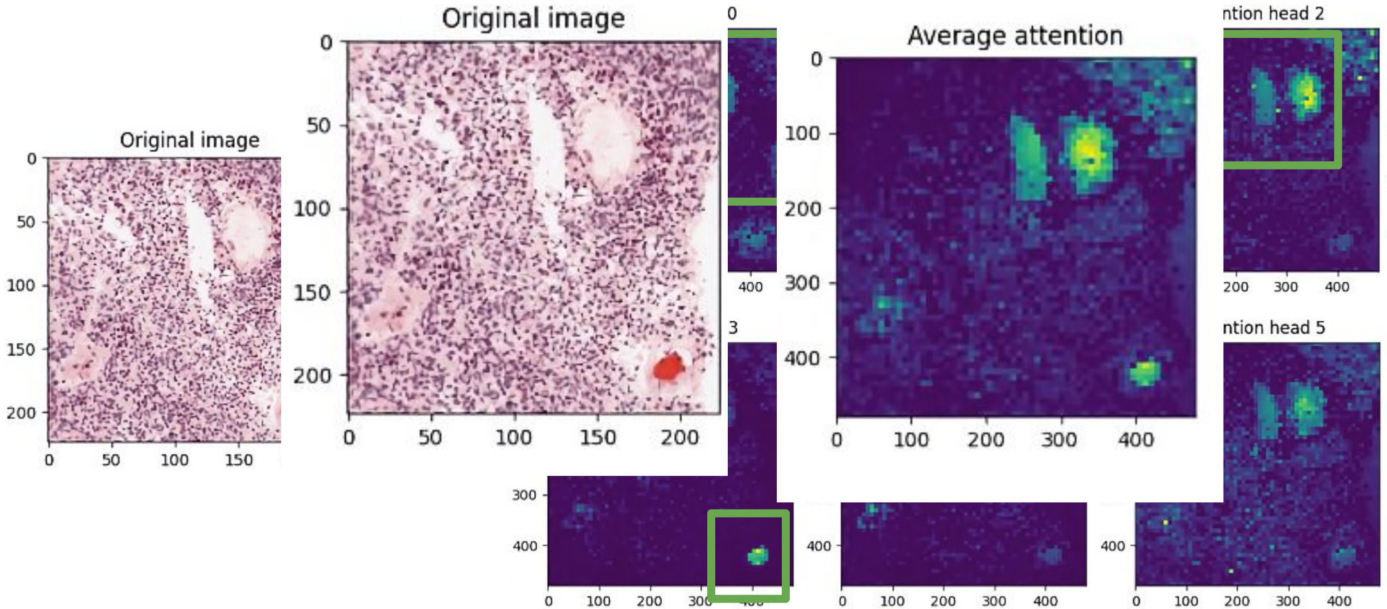
- As seen before, we can pretrain ViT using DINO to learn semantic embeddings without the need of labels.
- Look at the activation of the last layer of the each attention heads.
- 6 attention heads → 6 semantic attention maps.



# 2. Finding Regions of Interests in WSI

## C. Self-Supervised: DINO and Vision Transformers

6 attention heads → 6 semantic attention maps.



# 2. Finding Regions of Interests in WSI

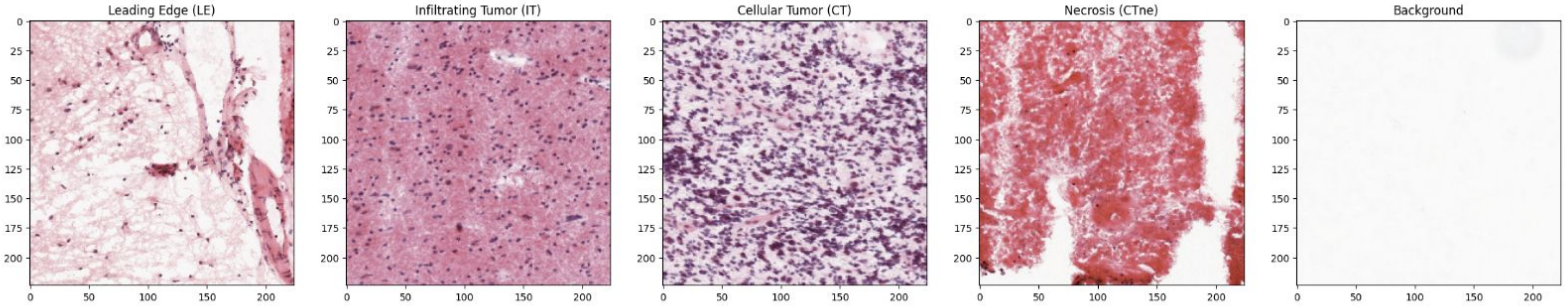
## Outcomes:

- Explored 3 approaches to finding regions of interests in whole slide images.
- Evaluated the pros and cons of each method.
- Produced and deployed a remote API that can be called by the Raspberry Pi for findings ROIs.

# 3. Uncertainty Estimation with Posterior Network

## Motivations:

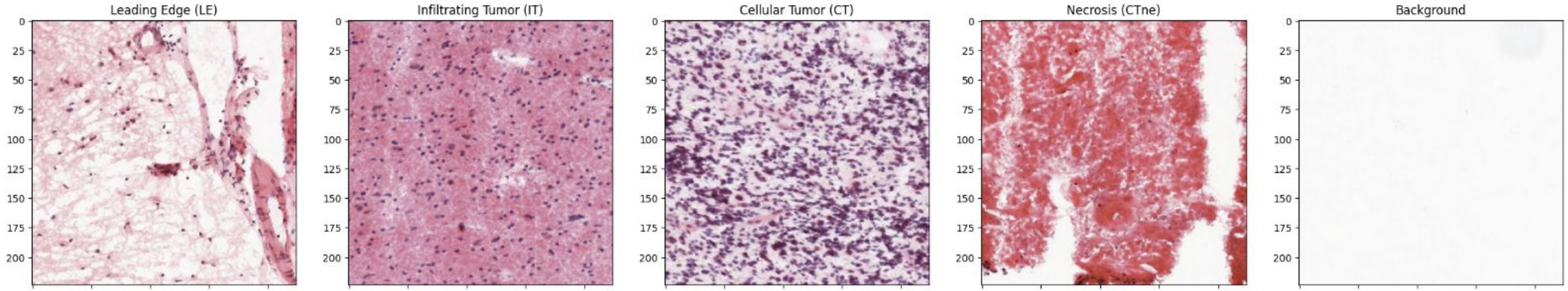
- Key to make ML reliable, gives us an idea how much we can trust ML systems.
- Segmentation map in IvyGap were obtained in a semi-automated way using decision forests, labels are prone to error.



(a) Leading Edge (b) Infil. Tumor (c) Cellular Tumor (d) Necrosis (e) Background

# 3. Uncertainty Estimation with Posterior Network

1. Can we use OoD detection methods to find problematic samples?
2. To what extent can we improve the accuracy of our model?



(a) Leading Edge (b) Infil. Tumor (c) Cellular Tumor (d) Necrosis (e) Background

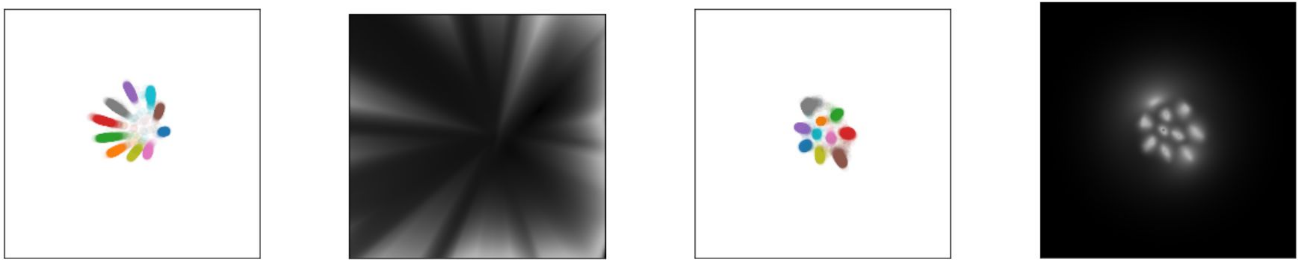
# 3. Uncertainty Estimation with Posterior Network

- **Key to make ML reliable**, gives us an idea how much we can trust ML systems.
- Current models can be overconfident.
- **Epistemic Uncertainty**: knowledge/model uncertainty.
- **Aleatoric Uncertainty**: data/label uncertainty.

# 3. Uncertainty Estimation with Posterior Network

## Posterior Networks:

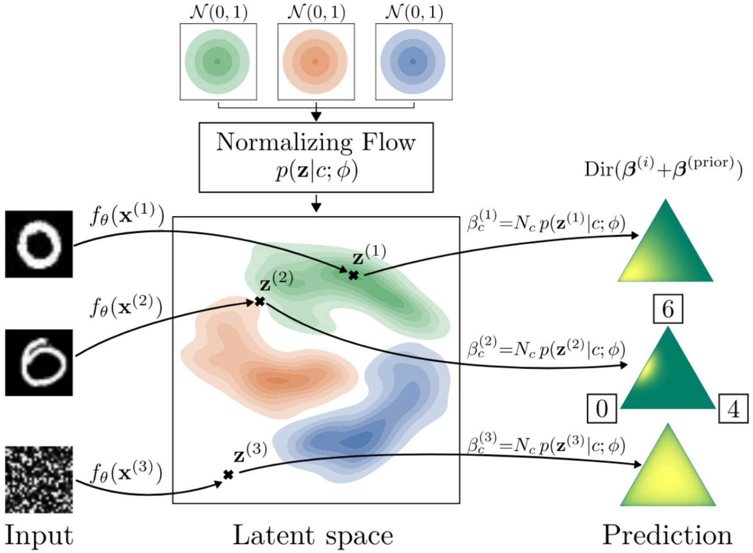
1. Use a **Normalizing Flow** to predict the posterior distribution over **predicted probabilities**.
2. Does not require Out of Distribution data.
3. Reaches state of the art in OoD detection and uncertainty calibration



(a) Data labels - PriorNet(b) Uncertainty - PriorNet (c) Data labels - PostNet (d) Uncertainty - PostNet

Posterior Network: Uncertainty Estimation without OOD Samples via Density-Based Pseudo-Counts, Charpentier et Al (2020)

# 3. Uncertainty Estimation with Posterior Network


$\mathcal{N}(0,1)$
$\mathcal{N}(0,1)$
$\mathcal{N}(0,1)$
Normalizing Flow
$p(\mathbf{z}|c;\phi)$
$\mathrm{Dir}(\boldsymbol{\beta}^{(i)}+\boldsymbol{\beta}^{(\mathrm{prior})})$
$f_\theta(\mathbf{x}^{(1)})$
$f_\theta(\mathbf{x}^{(2)})$
$f_\theta(\mathbf{x}^{(3)})$
$\mathbf{z}^{(1)}$
$\mathbf{z}^{(2)}$
$\mathbf{z}^{(3)}$
$\beta_c^{(1)}=N_c\, p(\mathbf{z}^{(1)}|c;\phi)$
$\beta_c^{(2)}=N_c\, p(\mathbf{z}^{(2)}|c;\phi)$
$\beta_c^{(3)}=N_c\, p(\mathbf{z}^{(3)}|c;\phi)$
6
0
4
Input
Latent space
Prediction

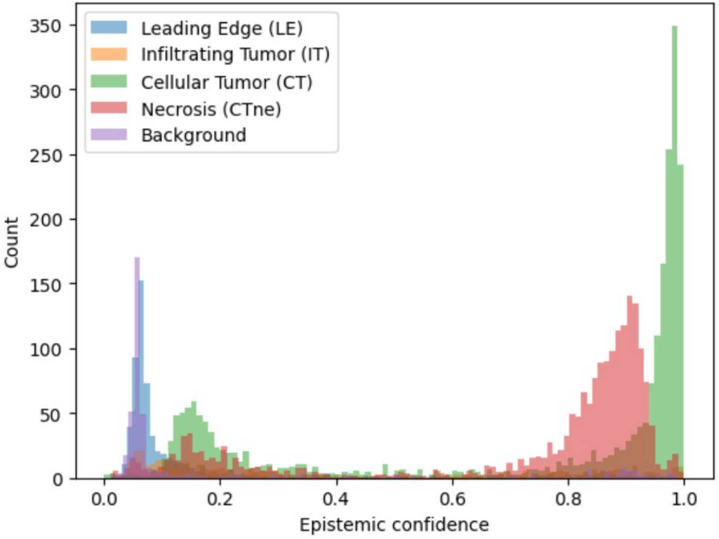
1. An input x is mapped to z into the latent space by f
2. The normalizing flow learns density functions p for each class.
3. We evaluate for z the density for its class.
4. This density is used to parameterize a Dirichlet distribution.
5. Higher Density => Higher confidence

Posterior Network: Uncertainty Estimation without OOD Samples via Density-Based Pseudo-Counts, Charpentier et Al (2020)

# 3. Uncertainty Estimation with Posterior Network

## Results on IvyGap:

- Model makes prediction **and** evaluates its confidence
- **Leading Edge** and **Infiltrating Tumor** have low epistemic (prediction) confidence
- **Necrosis** and **Cellular Tumor** have high epistemic confidence



(a) Epistemic uncertainty distribution

# 3. Uncertainty Estimation with Posterior Network

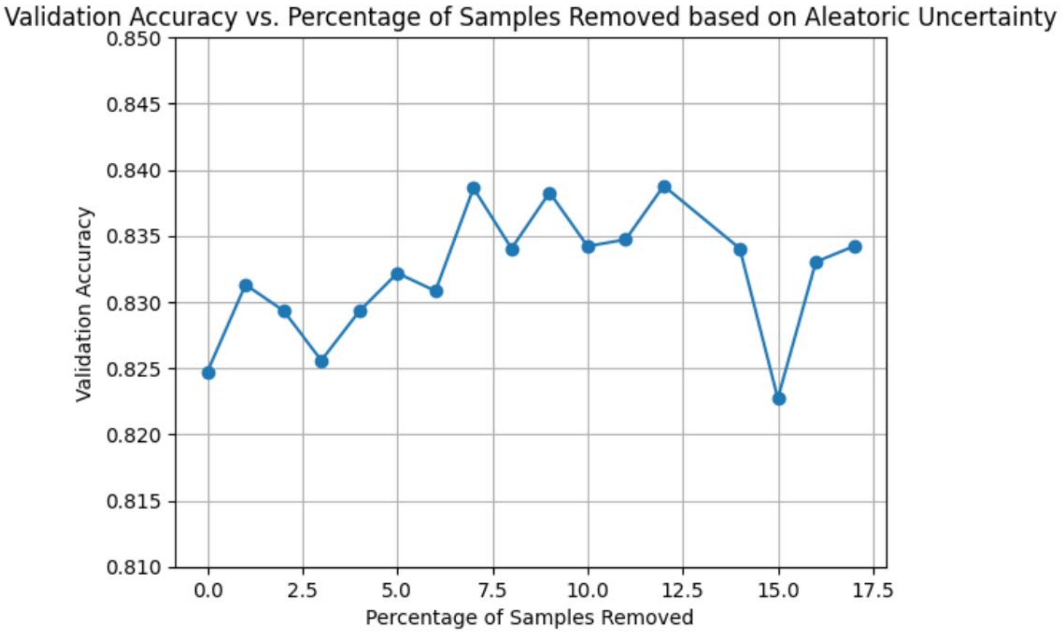
**Idea:** Does removing the samples with the worst label confidences makes for a better training set?

**Method**:

1. Remove in the training set for each class the same proportion of samples with worst label confidence.
2. Train a new model on this pruned training set.
3. Evaluate on unchanged validation set.

# 3. Uncertainty Estimation with Posterior Network

**We see an improvement in validation by removing the samples in the training set that are most likely to be mislabeled.**



# 3. Uncertainty Estimation with Posterior Network

**Outcomes:**

1. Calibrated a uncertainty estimation model for IvyGap (PostNet)
2. Posterior Net can estimate the prediction and labeling confidence.
3. Posterior Net can detect samples that are out of distribution (OoD) compared to the training distribution, without the need of OoD dataset.
4. We proposed a method to improve model accuracy by calibrating first a model on detecting mislabeled samples and pruning the training set.

# Conclusion

1. **Self-Supervised Learning for robust pretraining**
   *SSL and DINO algorithm makes for great pretrained model with semantic embeddings.*
2. **Finding Regions of Interests in Whole Slide Images**
   *DINO, Transformer Interpretability and TransUNet are 3 approaches to finding ROIs with various degree of flexibility.*
3. **Uncertainty Estimation with Posterior Network**
   *PostNet provide prediction and label confidence, which we leveraged in a new method to prune training sets.*

# Thank you!

# Relevant Papers

- Krishnan, R., 2022. **Self-supervised learning in medicine and healthcare**. Nat. Biomed. Eng 1–7. https://doi.org/10.1038/s41551-022-00914-1
- Chen, R.J., **Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning** 12.
- Jiang, Y.Q., 2020. **Recognizing basal cell carcinoma on smartphone‐captured digital histopathology images with a deep neural network**. Br J Dermatol 182, 754–762. https://doi.org/10.1111/bjd.18026
- Chen, J., 2021. **TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation.**
- Caron, M., 2021. **Emerging Properties in Self-Supervised Vision Transformers.**
- Zheng, Y., 2021. **A deep learning based graph-transformer for whole slide image classification** (preprint). Oncology. https://doi.org/10.1101/2021.10.15.21265060
- Marostica, E., 2021. **Development of a Histopathology Informatics Pipeline for Classification and Prediction of Clinical Outcomes in Subtypes of Renal Cell Carcinoma**. https://doi.org/10.1158/1078-0432.CCR-20-4119