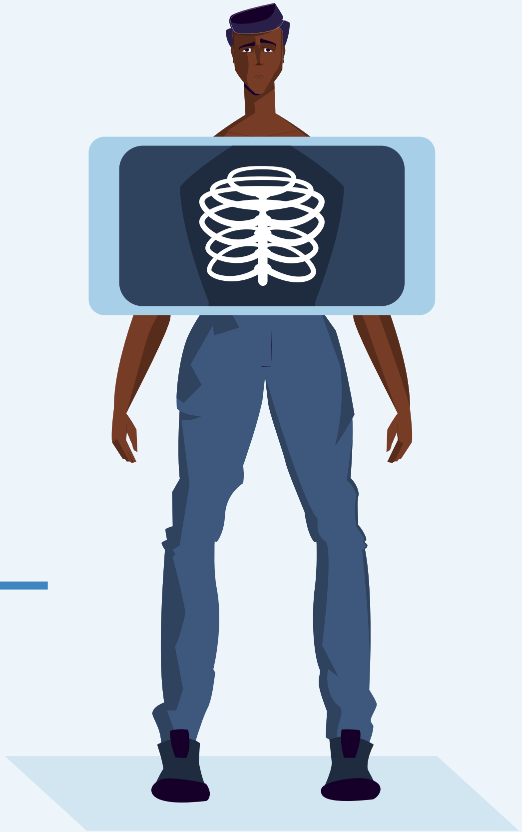
PatchCamelyon Challenge

Rodrigo Chavez, Ioanna Gogou, Christelle Klein, Alexia Muresan and Sina Taslimi

01

Introduction



Breast Cancer & Metastasis

- One of the leading causes of female **mortality** worldwide
- **Abnormal cells** in the breast grow in an **uncontrolled** manner
- Metastasis is a **severe case** of cancer where the cancerous cells spread to the rest of the body, among others in the **lymph nodes**
- **Timely detection** plays a crucial role in effective treatment

Why computer-aided diagnosis?

- ➔ Diagnosing cancer metastasis in lymph nodes is a **challenging task**
- ➔ Potentially **more accurate** than trained pathologists
- ➔ **Faster** diagnosing

Dataset: PatchCamelyon Benchmark

- ➔ Image **classification** dataset
- ➔ 300k **96x96 patches** from 400 scans of lymph node sections of Camelyon16

Dataset: PatchCamelyon Benchmark

327,680

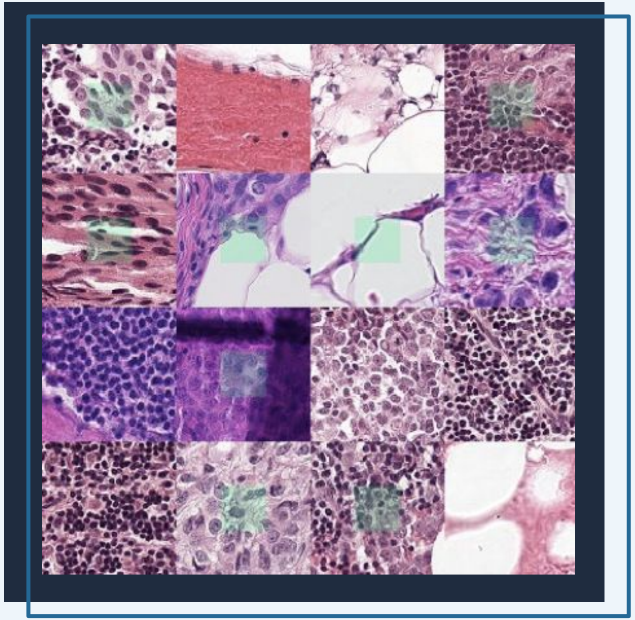
Colour images

96x96 px

Image resolution

2 classes

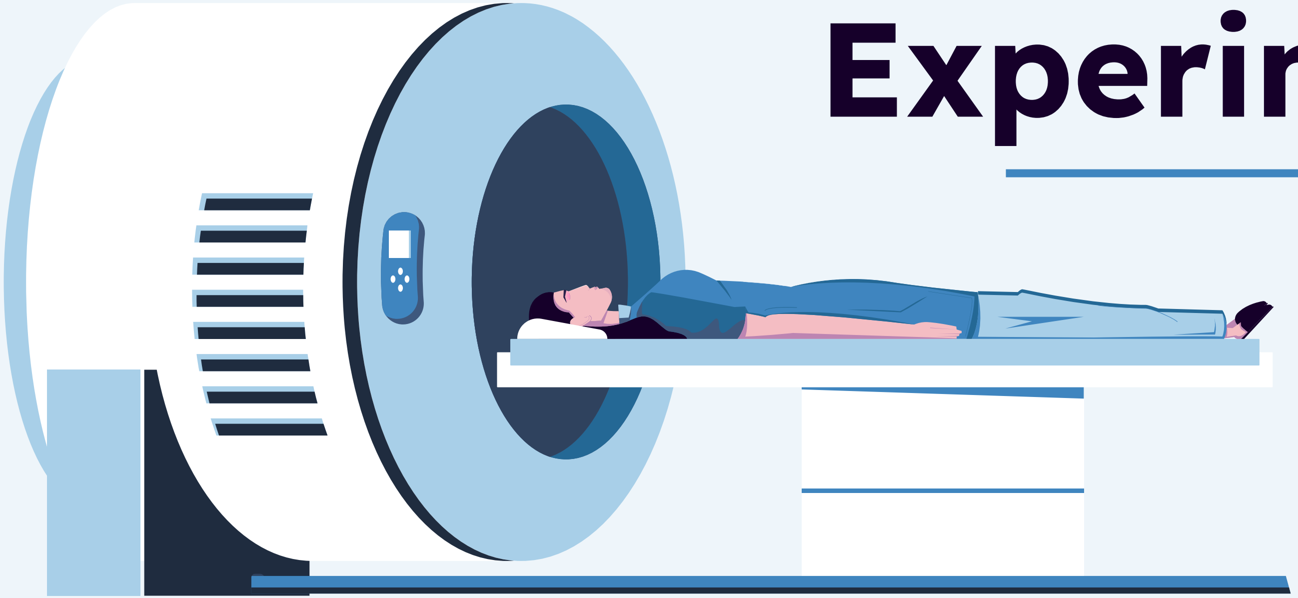
Presence of metastatic tissue or not



Lymph node sections

02

Experiments



Research Questions

1. Which pre-trained model is most suitable for the challenge based on the **Accuracy/Complexity trade-off**?
2. Can we improve the model using **Data Augmentations**?
3. Can medical expert trust the model's predictions?
    a. **Model Explainability**
    b. **Model Uncertainty**

Experiment 1: Accuracy/Complexity trade-off

↪ Model Selection

2012
AlexNet
5 Conv Layers,
3 FC Layers

2014
GoogleNet
22 Conv Layers,
1 FC Layer, and
Inception Modules

2015
VGG-11
9 Conv Layers,
2 FC Layers

2015
ResNet-18
17 Conv Layers,
1 FC Layer and
Residual Connections

2015
Inception-v3
48 Conv Layers,
1 FC Layer, and
Inception Modules

2017
DenseNet-161
161 Conv Layers,
1 FC Layer and
Inception Modules

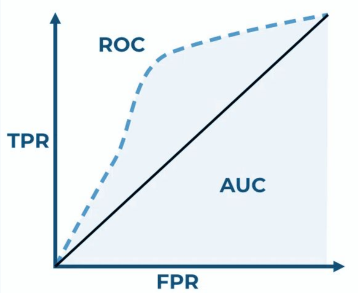
2020
ViT-Base-16
12 Transformer Layers:
16x16 Patches, Positional
Encodings, and Multi-head
Self-Attention

2022
Swin-Transformer-v2
12 Swin Transformer blocks:
MLP, LayerNorm, and
Multi-head Self-Attention

Experiment 1: Accuracy/Complexity trade-off

↪ Experiment Setup

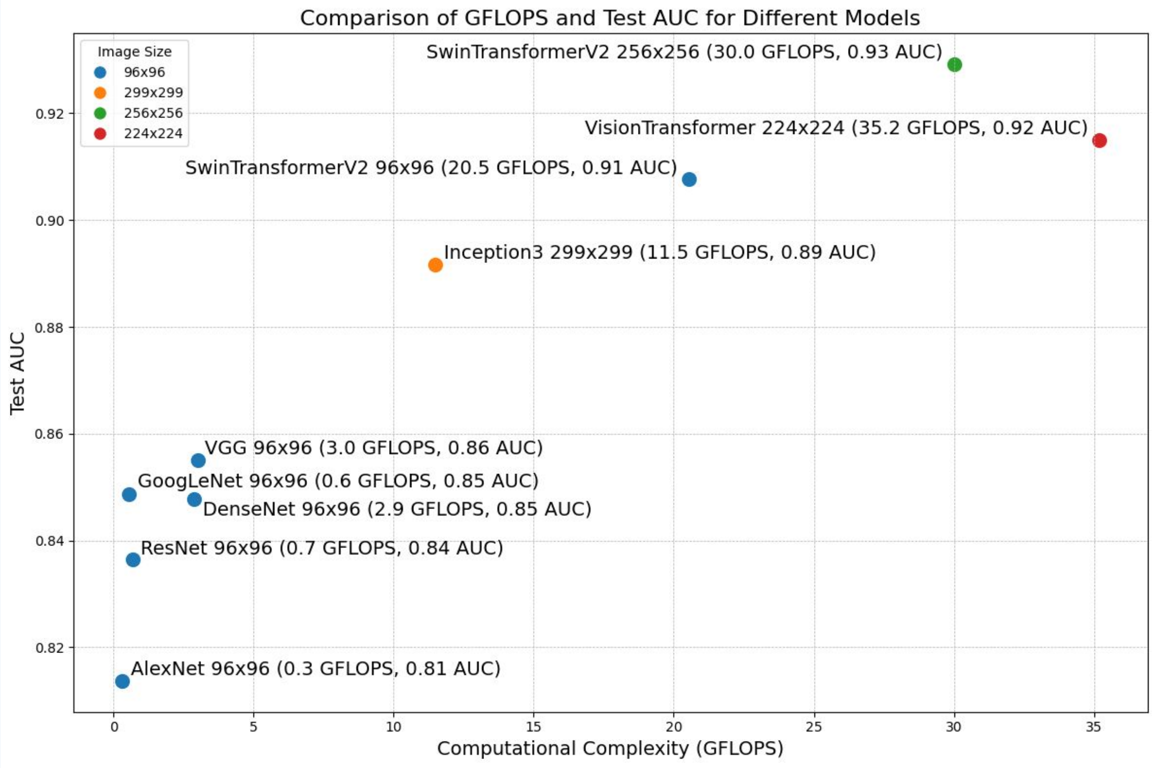
- Metrics: **AUC** for accuracy, **GFLOPs** for complexity
- All models pre-trained on **ImageNet 1k**
- Training on **last fully connected layer** only
- Hyperparameters:
  - **5** Epochs, Batch size = **256** (128 for Inception)
  - Optimizer: **Adam** (learning rate = 0.01)
- Loss function: **Cross Entropy Loss**



$$\text{Loss} = -\sum_{i=1}^{\substack{\text{output}\\ \text{size}}} y_i \cdot \log \hat{y}_i$$

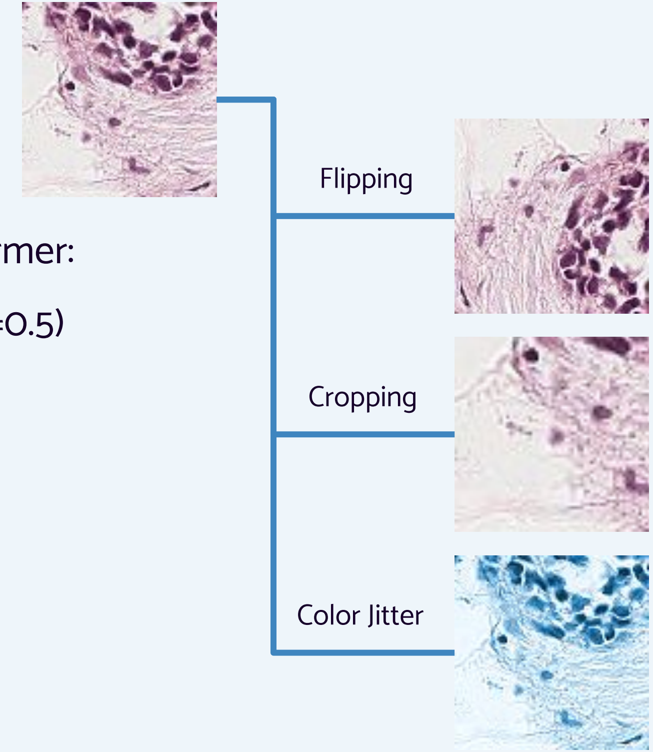
Experiment 1: Accuracy/Complexity trade-off

↪ Results



Experiment 2: Data Augmentations

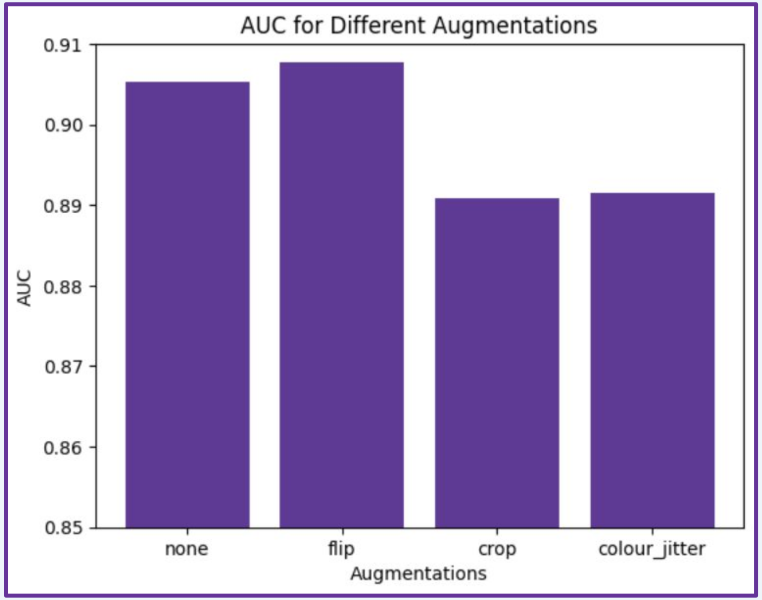
↪ Ablation Study



- ➔ Compared three augmentations on Swin Transformer:
  - Random horizontal and vertical **flipping** (p=0.5)
  - Random **cropping** (p=1, random location)
  - **Color Jitter** (p=1, random fluctuations)
- ➔ Hyperparameters:
  - **10** Epochs, Batch size = **512**

Experiment 2: Data Augmentations

↪ Results



Swin + Flipping	AUC (%)
96x96	90.77
256x256	92.92

Experiment 3: Explainability

↪ Grad-CAM

➔ **Idea**

- Find parts of the image that have more influence on the model’s decision
    - Idealistically the metastatic parts of the tissue
    - Want to see if we change an area in the input how much the output differs
    - More change → More influence on the decision
- Can be used as a guide for doctors and radiologists to locate the disease easier

Experiment 3: Explainability

↪ Grad-CAM

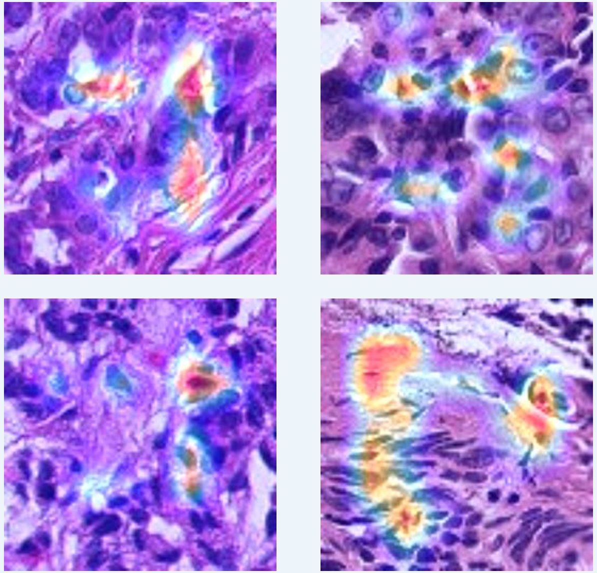
→ **Method**

- Take the last block of the last layer (before linear layers) in Swin Transformer
- Take the gradient of the class logit with respect to those activation maps
- Pool the gradients
- Weight the channels of the map by the corresponding pooled gradients
- Interpolate the heat-map

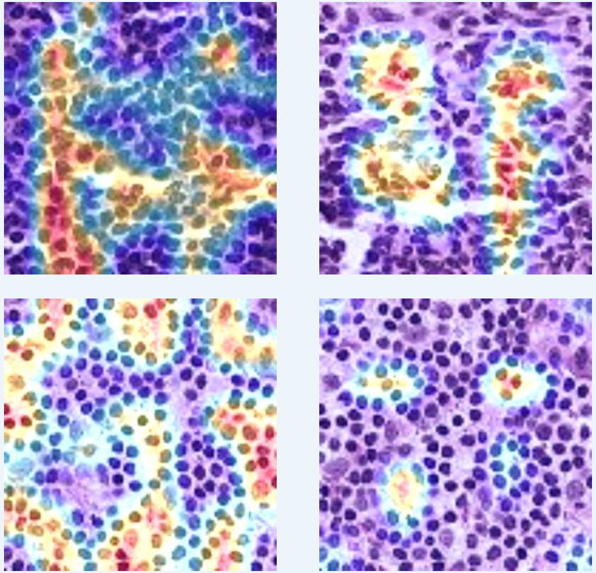
Experiment 3: Explainability

↪ Results

Metastatic (P)



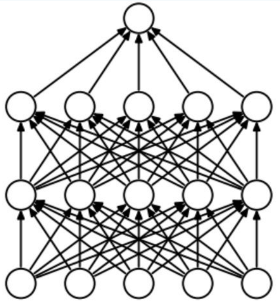
Non-Metastatic (N)



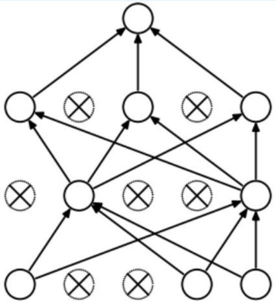
Experiment 3: Uncertainty

↪ Monte-Carlo Dropout

- ➔ Apply **dropout** during **testing** (p=0.3)
- ➔ Perform **multiple test runs** on each image (10 runs).
- ➔ Each run will generate a **different output probability** due to randomized dropout.
- ➔ The **uncertainty** on a sample is measured as the **variance** of its output probability across the multiple runs.



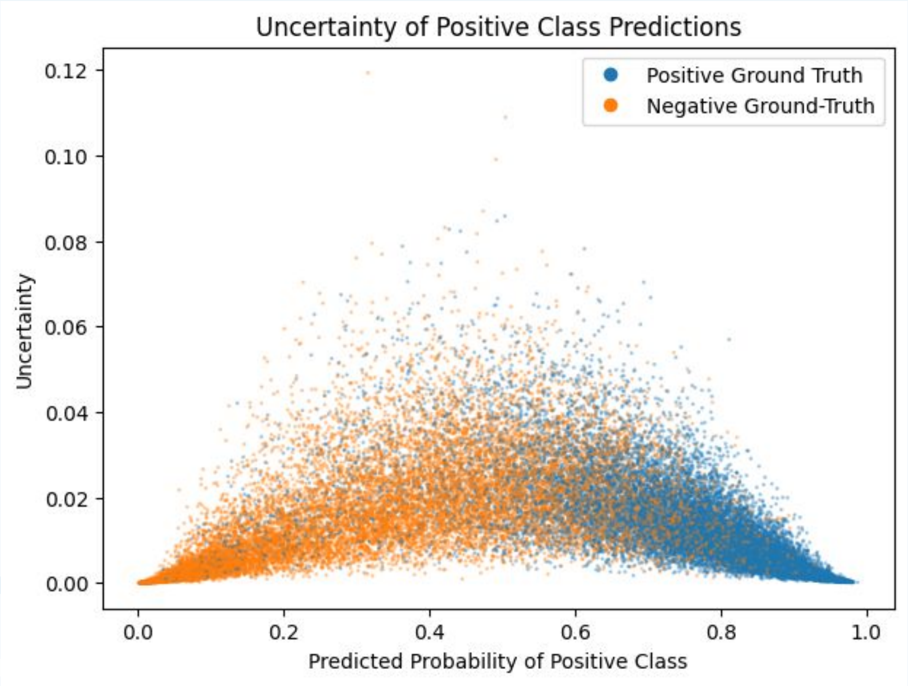
(a) Standard Neural Net



(b) After applying dropout.

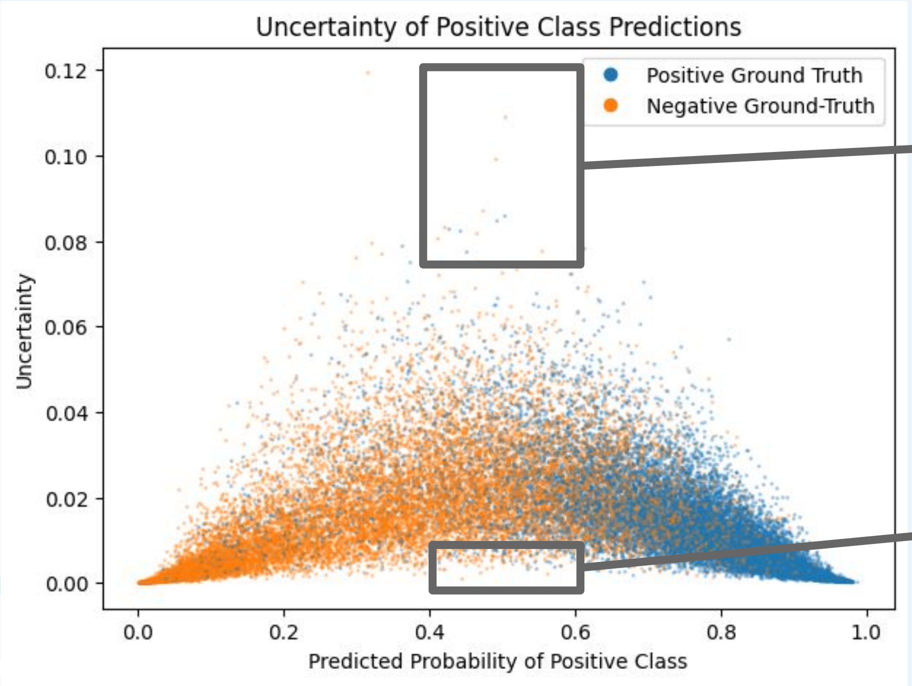
Experiment 3: Uncertainty

↪ Results

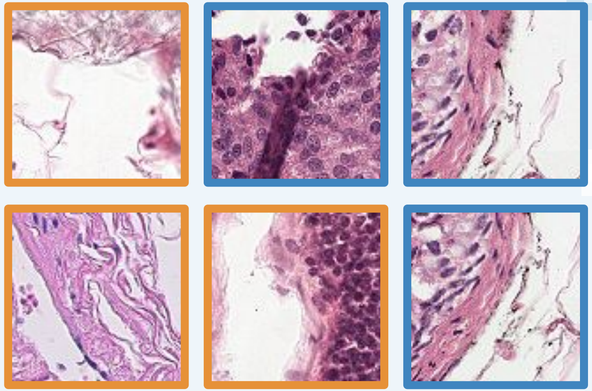


Experiment 3: Uncertainty

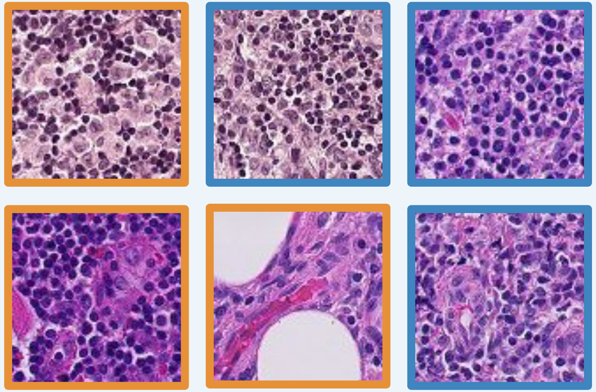
↪ Results



Uncertain Samples

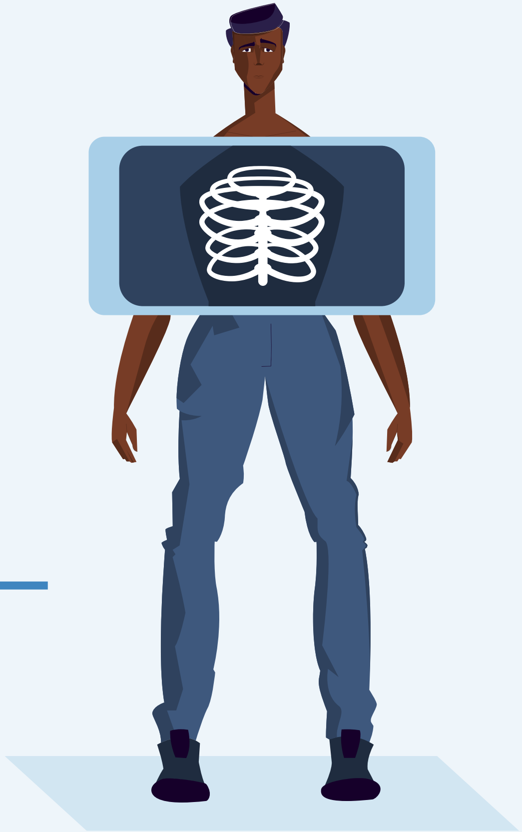


Certain Samples



03

Conclusion



Conclusion

- ➔ The **Swin Transformer** performed best compared to other classifiers in terms of both **accuracy** and **complexity**.
- ➔ The data augmentation of **flipping** was useful for the PCAM Challenge.
- ➔ Our Swin Transformer model with data augmentations **ranked 62th out of 179** models with **92,92% AUC** in the leaderboard. The top models had a perfect AUC score.
- ➔ Using **explanations** and **uncertainty**, medical expert can **reliably understand model predictions** and filter down the patients for manual diagnosis.

Future plans

- Test out bigger Transformer models, due to our current resource limitations.
- Improve performance with other techniques
  (e.g. model ensembling, stain normalization, complex data augmentations, etc.).
- Test and compare more explainability methods.
- Understand the pattern behind uncertain and certain predictions.

Thanks!

Do you have any questions?