TUHH
Hamburg University of Technology

UKE HAMBURG

Nagalikhitha Reddipalli

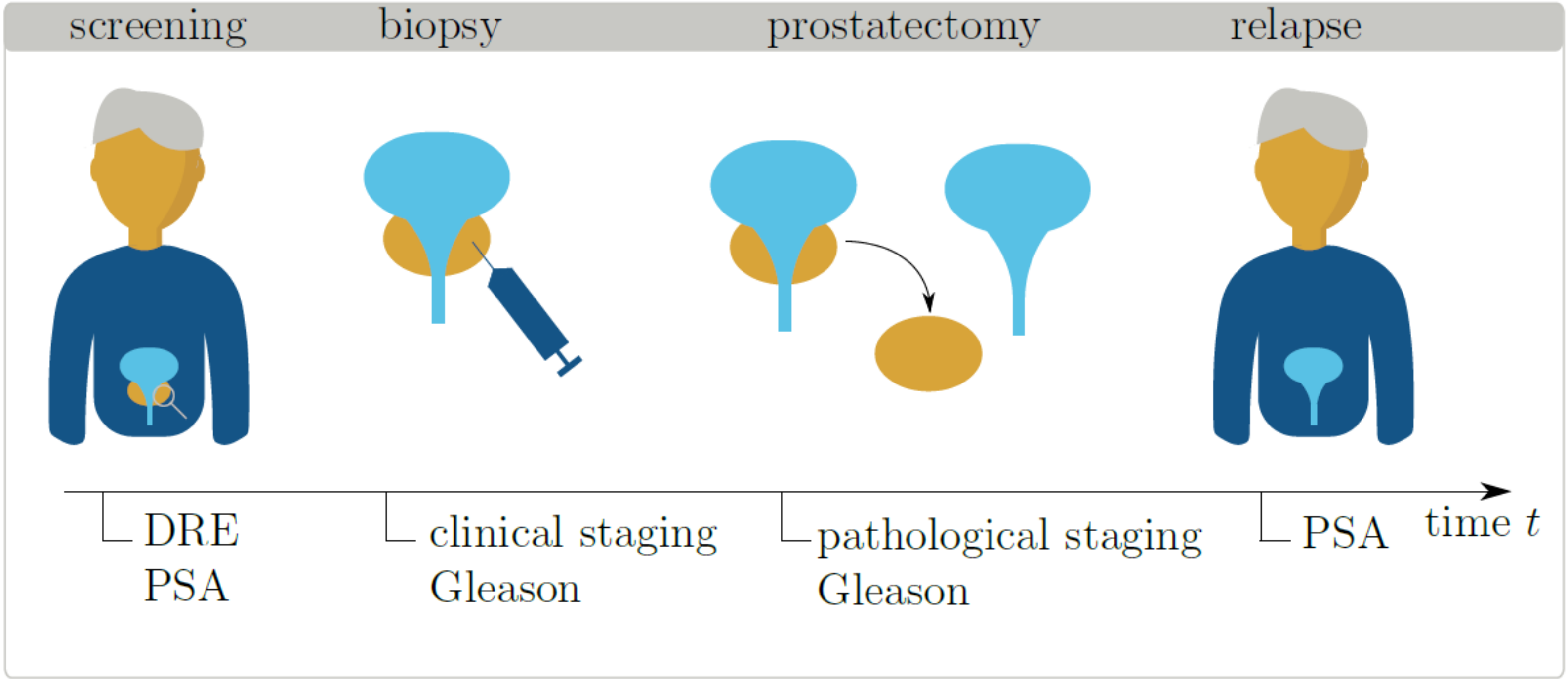
Institute of Medical Technology and Intelligent Systems

Institute of Medical Technology and Intelligent Systems

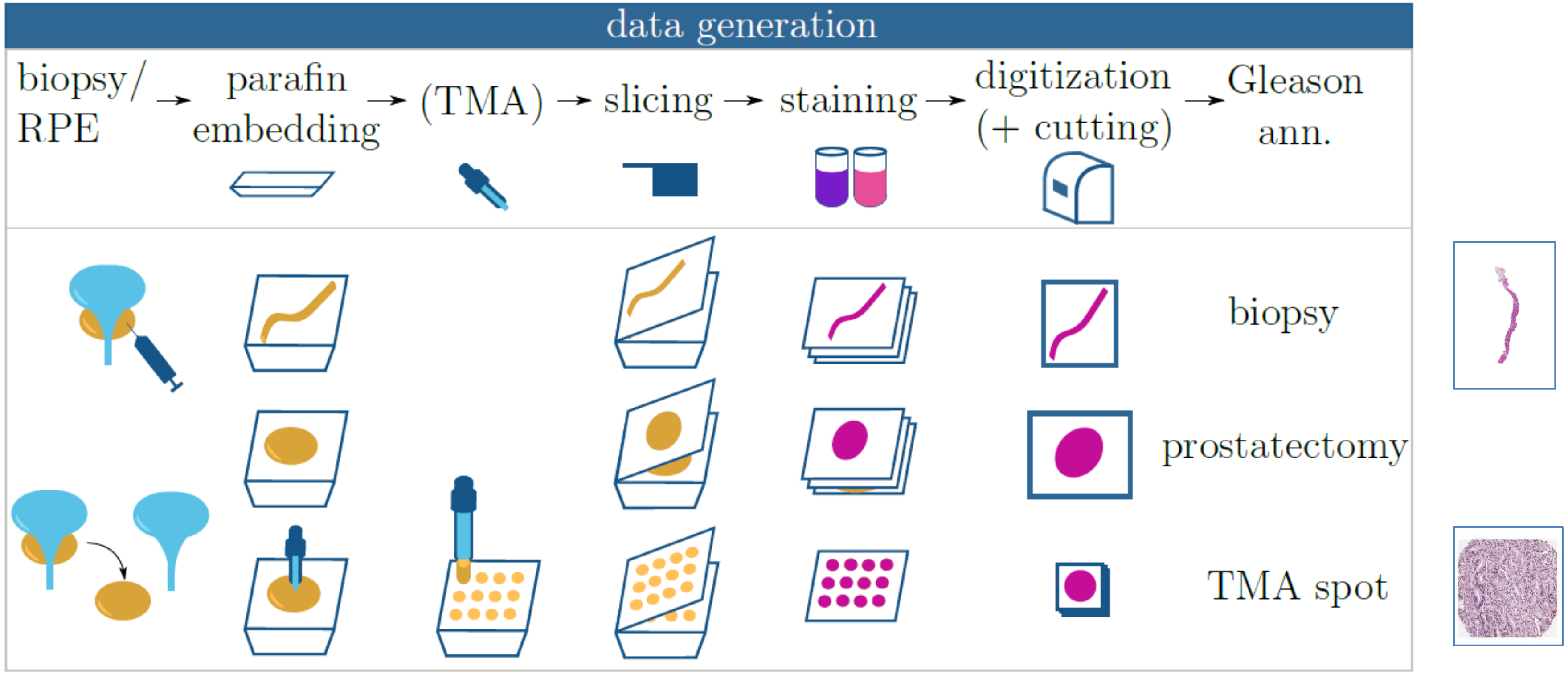
Relapse Prediction of Prostate Cancer with Histopathology Images using Vision Transformers

IMSB

Introduction: Prostate cancer



Data acquisition



Dietrich(2022)

Introduction: Relapse prediction

- Gleason grade is subjective
- High interrater variability between pathologists – ambiguity in diagnosis and therapy
- Biochemical relapse prediction has objective endpoint
- Generate binary label
  - If event occurs before 60 months : positive
  - If event does not occur : negative

RESEARCH QUESTIONS

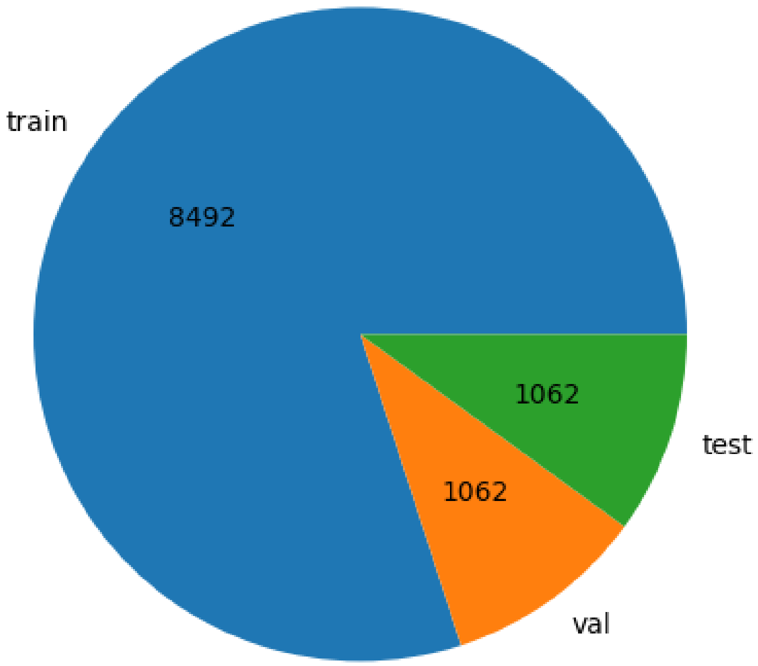
Research questions

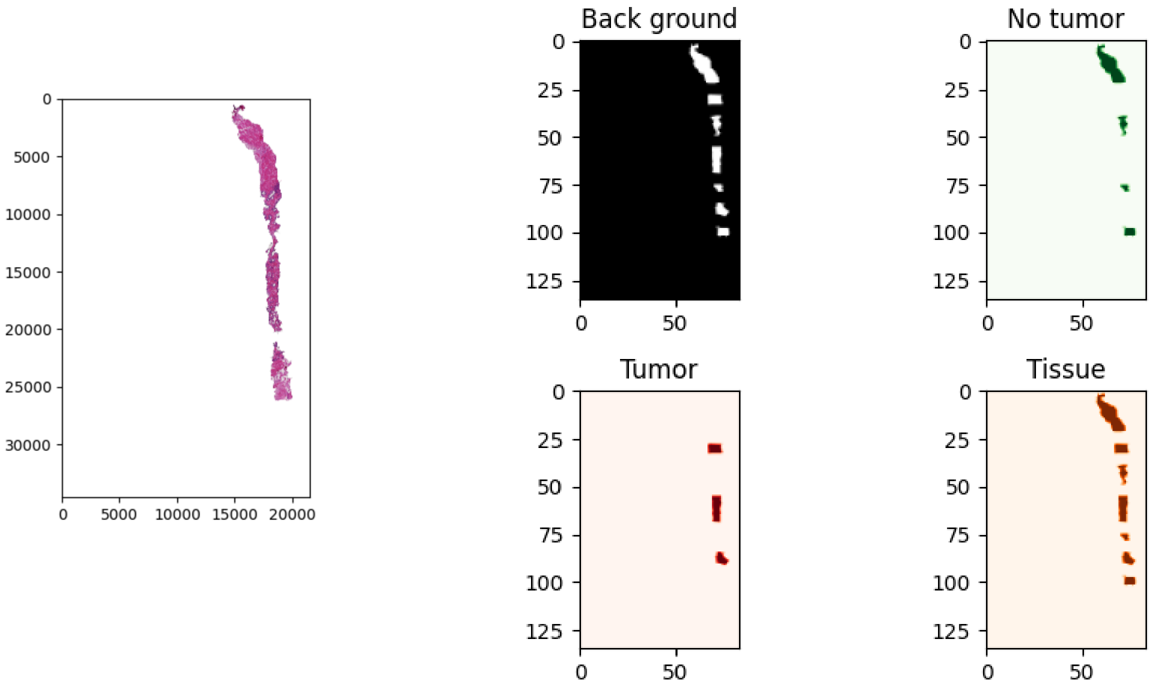
- **R1:** How well can state-of-the-art **vision transformers** (ViT) predict relapse of prostate cancer on our in-house data?
- **R2:** Does **pre-training** ViT on **domain-specific** images improve the performance of the model?
- **R3:** How good are **hierarchical vision transformers** on our in-house prostate cancer data?

DATASETS

Datasets

- PANDA challenge for cancer detection
  - Prostate cANcer graDe Assessment
  - Size: 5,000 to 40,000 pixels per dimension
  - 10,616 images

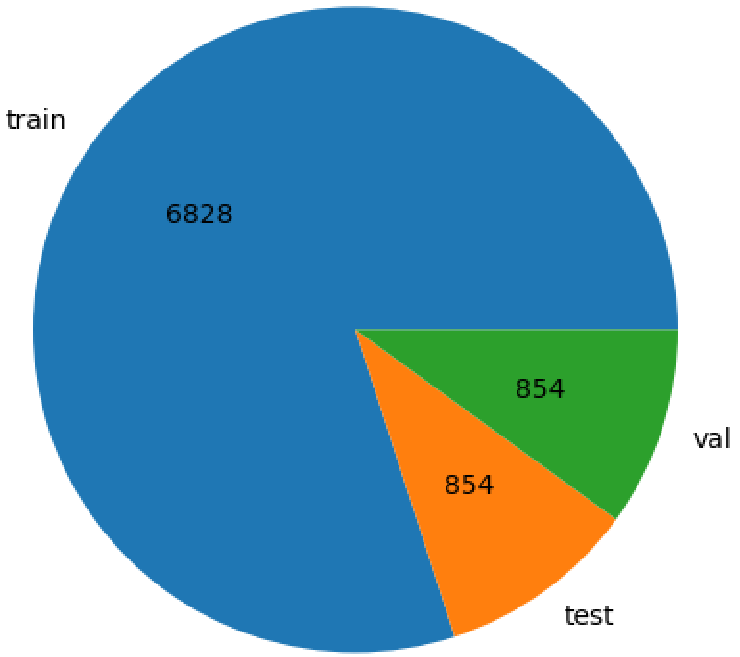


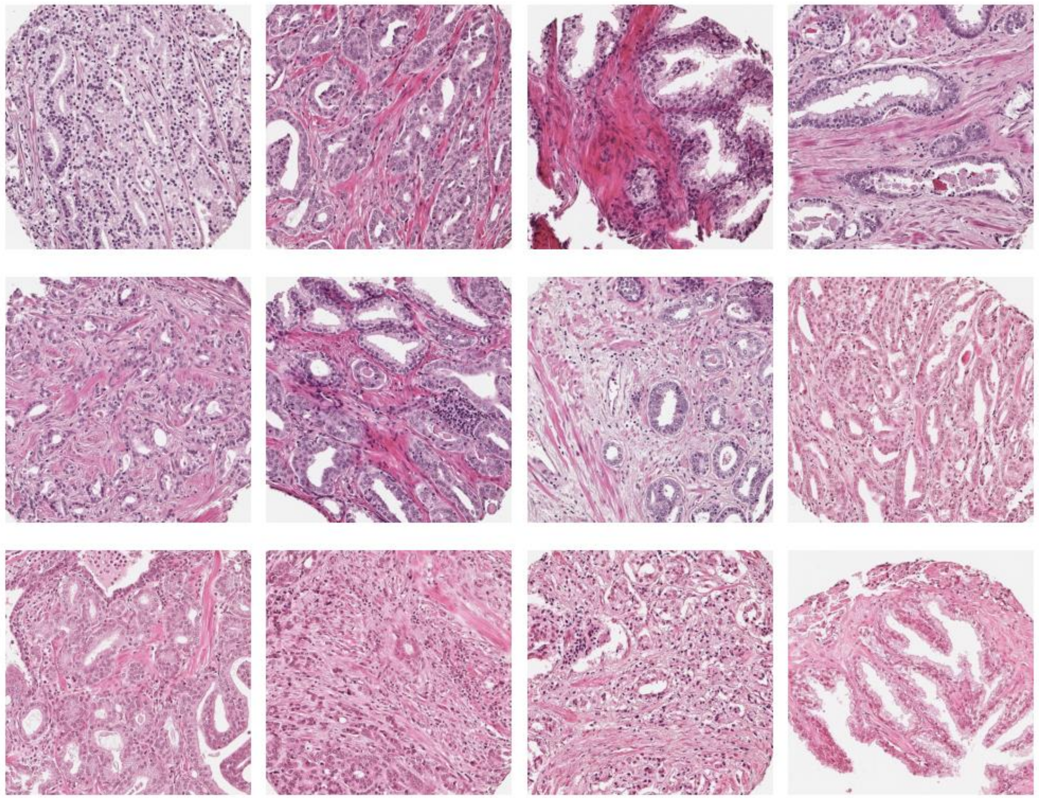


https://www.kaggle.com/competitions/prostate-cancer-grade-assessment/data

Datasets

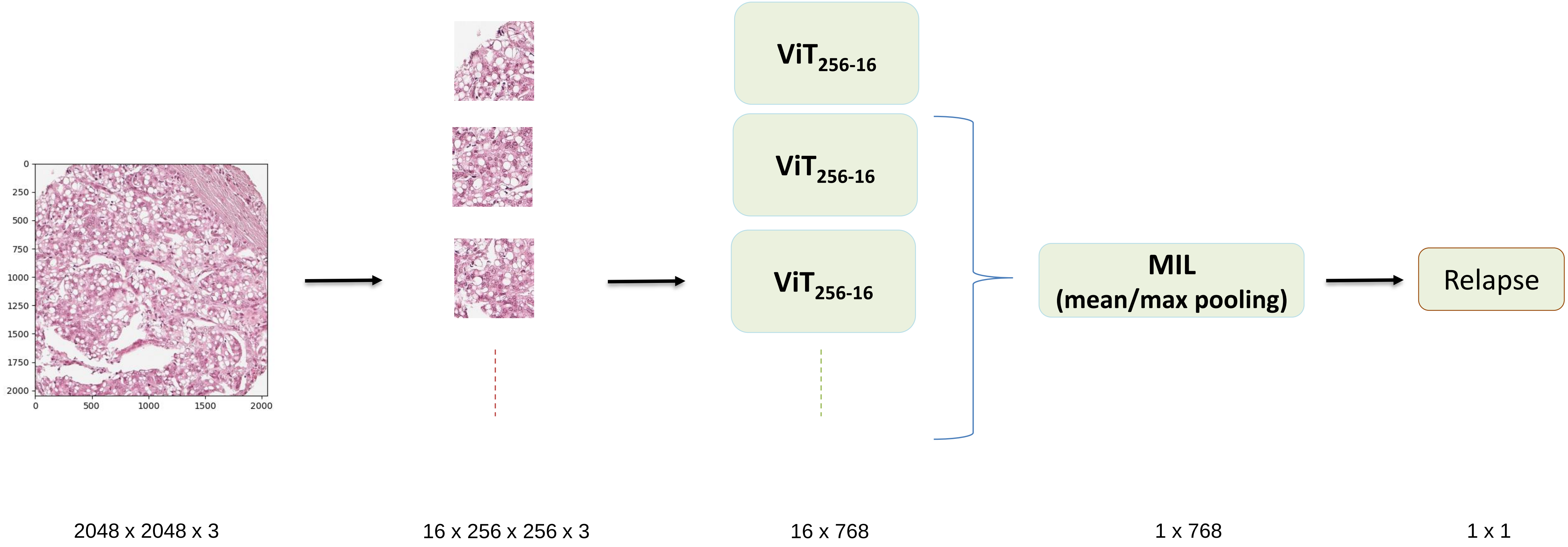
- UKE dataset for relapse prediction
  - TMA spots
  - Size: 2048 x 2048 pixels
  - 8536 images



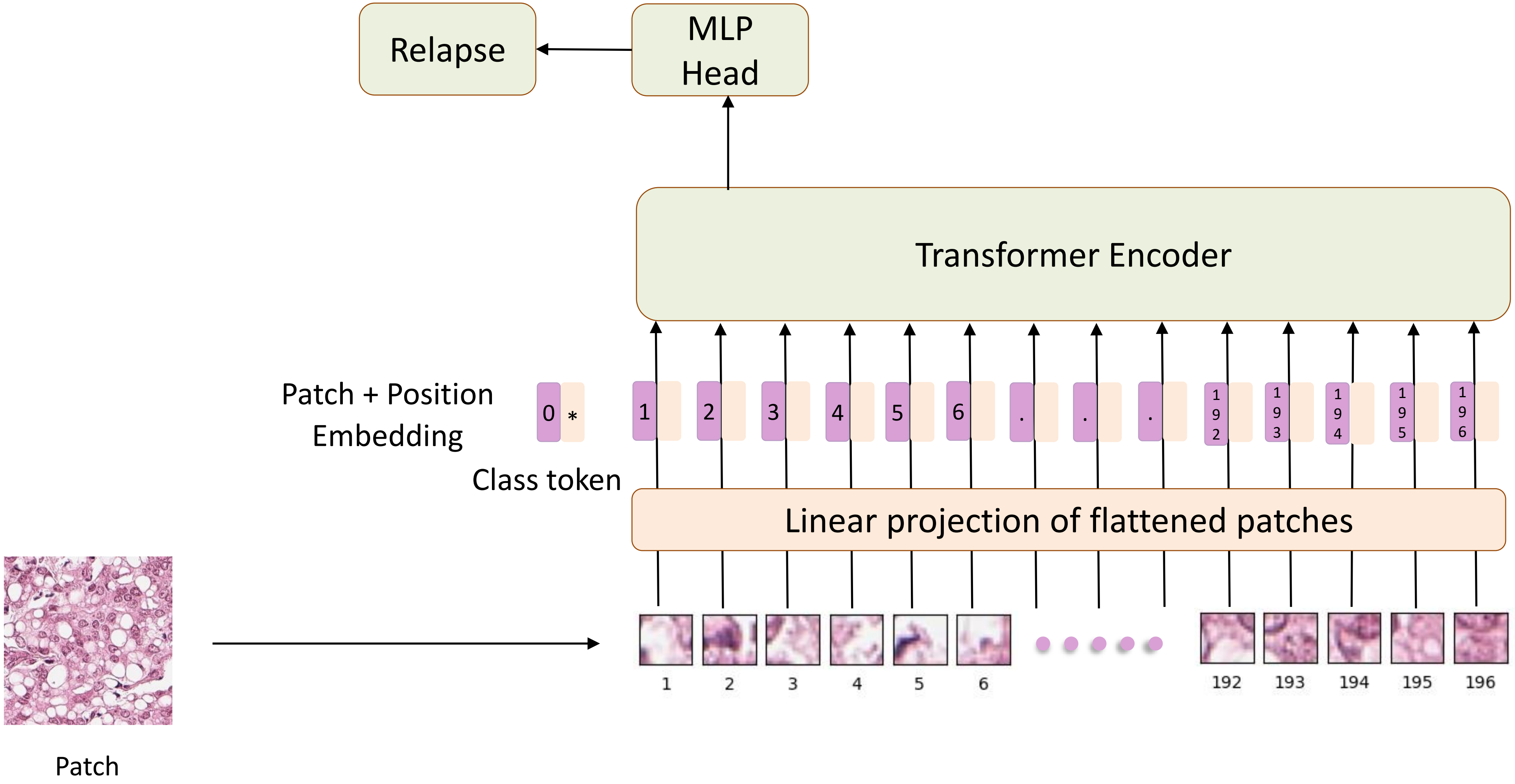


METHODS

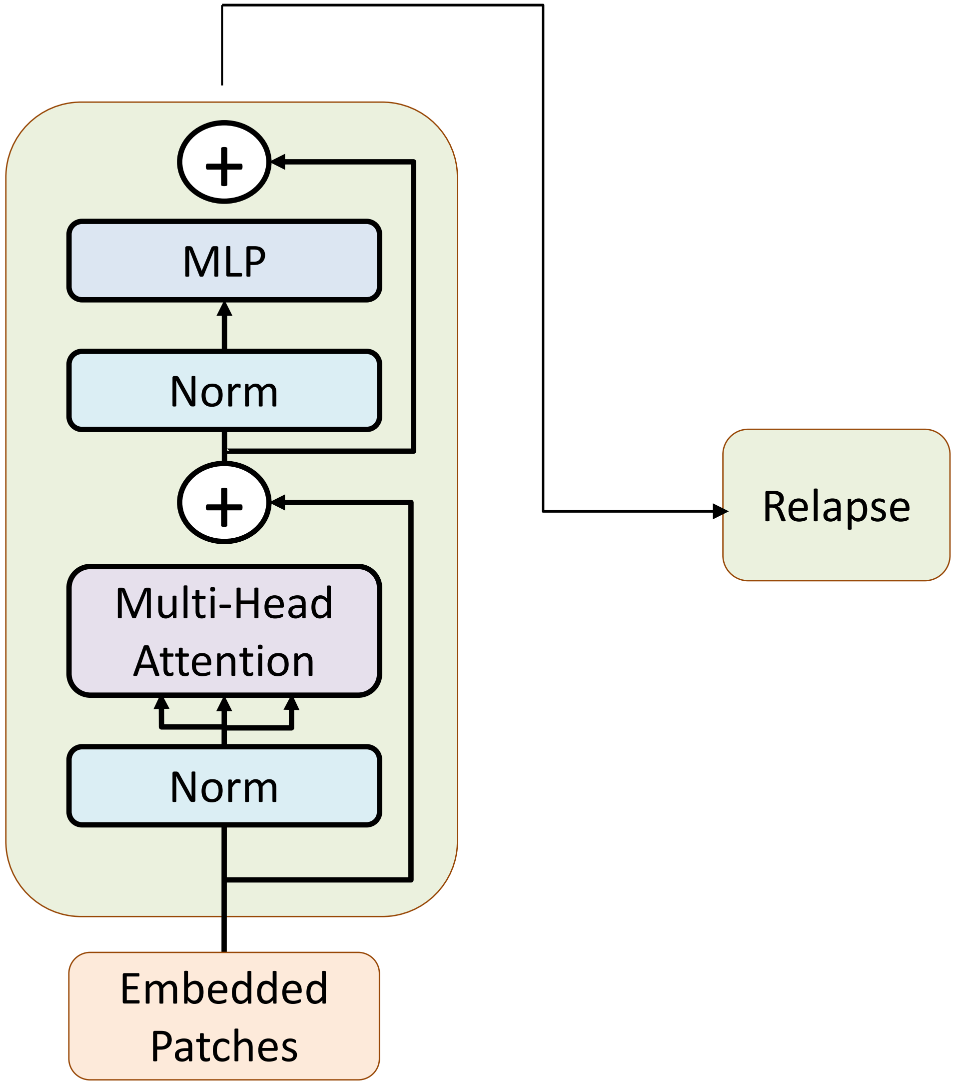
Vision transformers with multiple instance learning



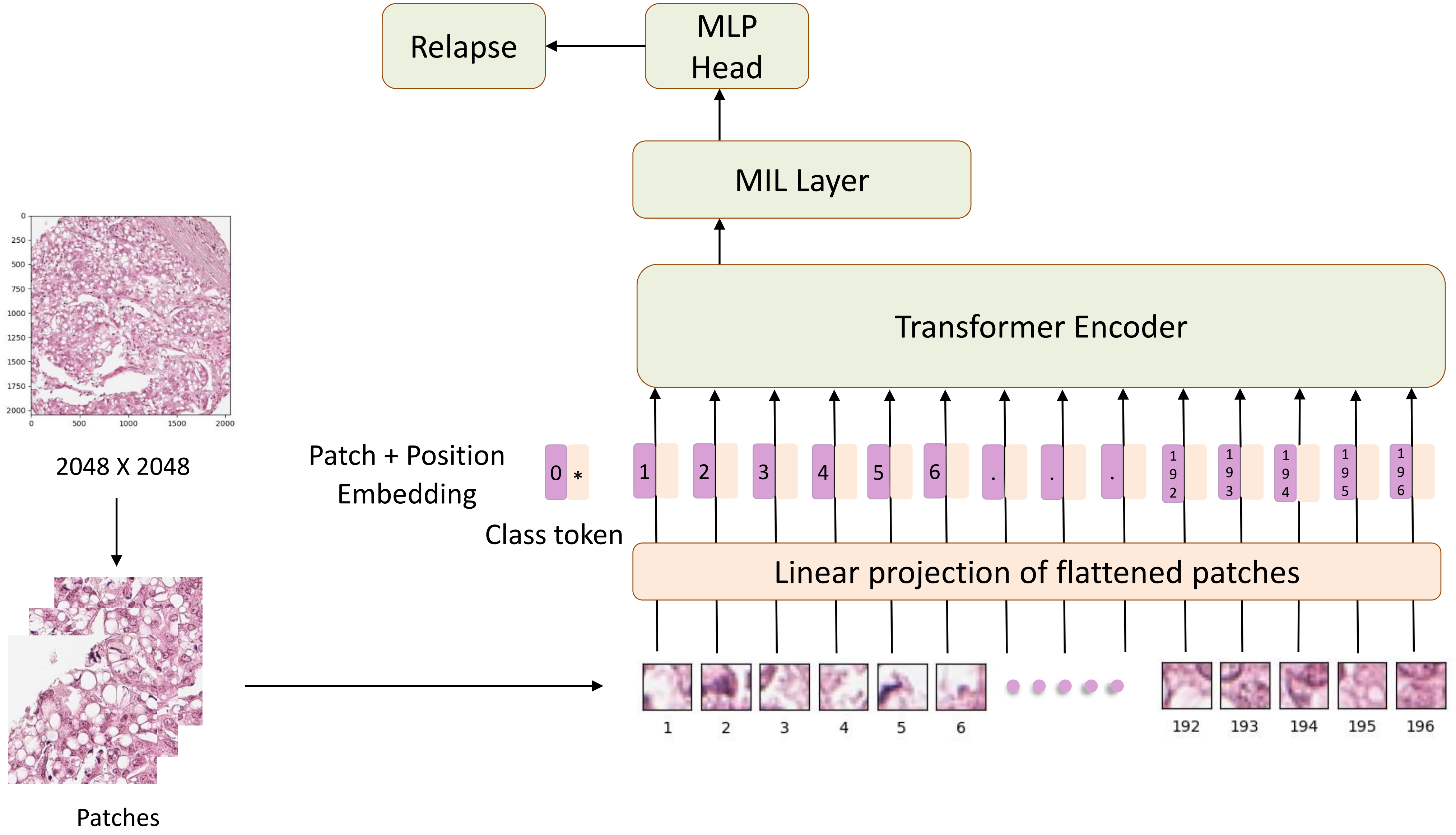
ViT on single patch



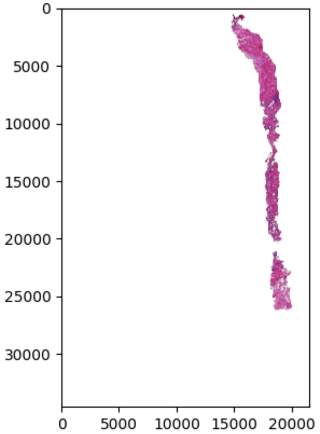
ViT encoder


MLP
Norm
Multi-Head
Attention
Norm
Embedded
Patches
Relapse

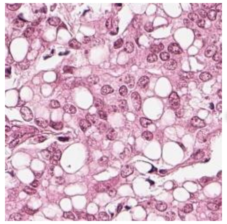
ViT for TMA spot



Pre-training of ViT on cancer detection



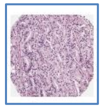
→ Patch selection →

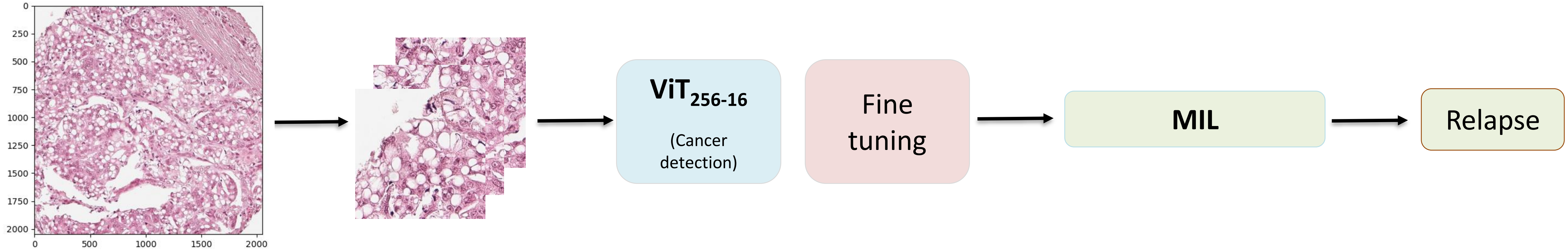


→ $\mathbf{ViT}_{256\text{-}16}$ → Cancer detection

PANDA image

Relapse prediction on PANDA pre-trained ViT





2048 x 2048 x 3

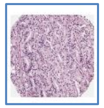
16 x 256 x 256 x 3

16 x 768

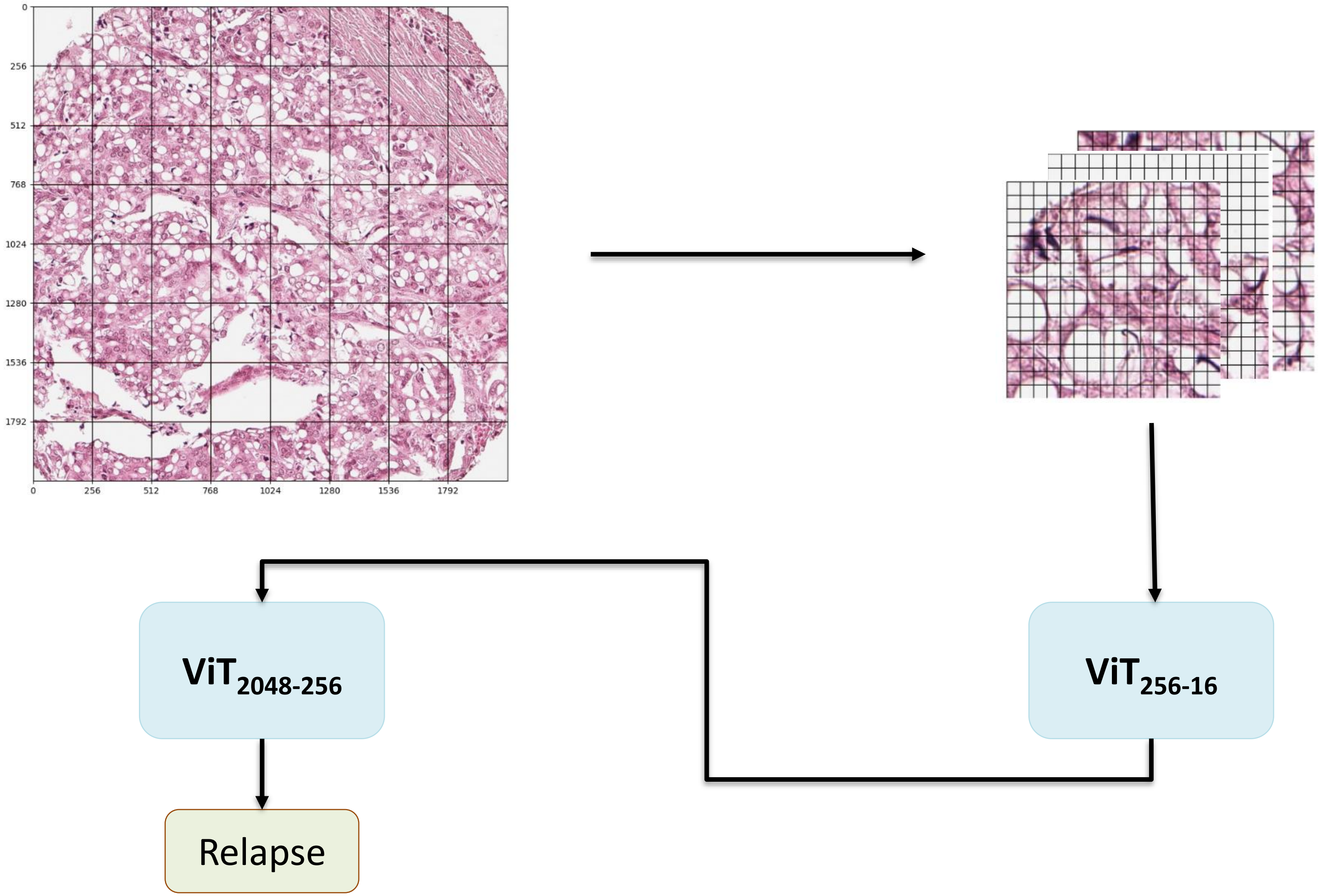
1 x 768

1 x 1

Hierarchical ViT

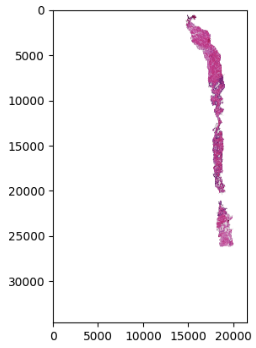


- As UKE dataset is of size 2048 x 2048, the intended approach has two levels of hierarchy



Hierarchical ViT

- Two pre-trainings : $\mathbf{ViT_{256\text{-}16}}$, $\mathbf{ViT_{1024\text{-}256}}$

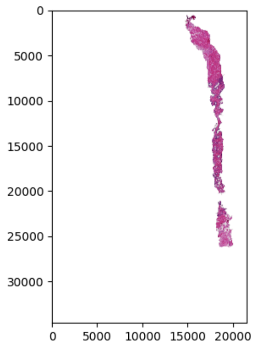


$\downarrow$

$\mathrm{ViT}_{256\text{-}16}$

$\downarrow$

Cancer



$\downarrow$

$\mathrm{ViT}_{1024\text{-}256}$

$\downarrow$

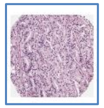
Cancer

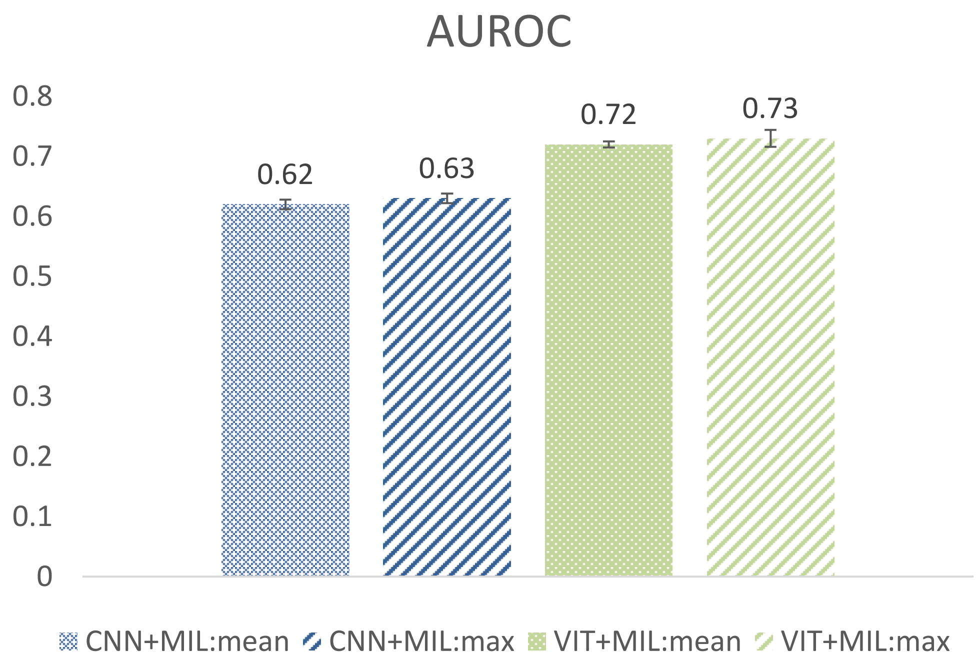
Methods

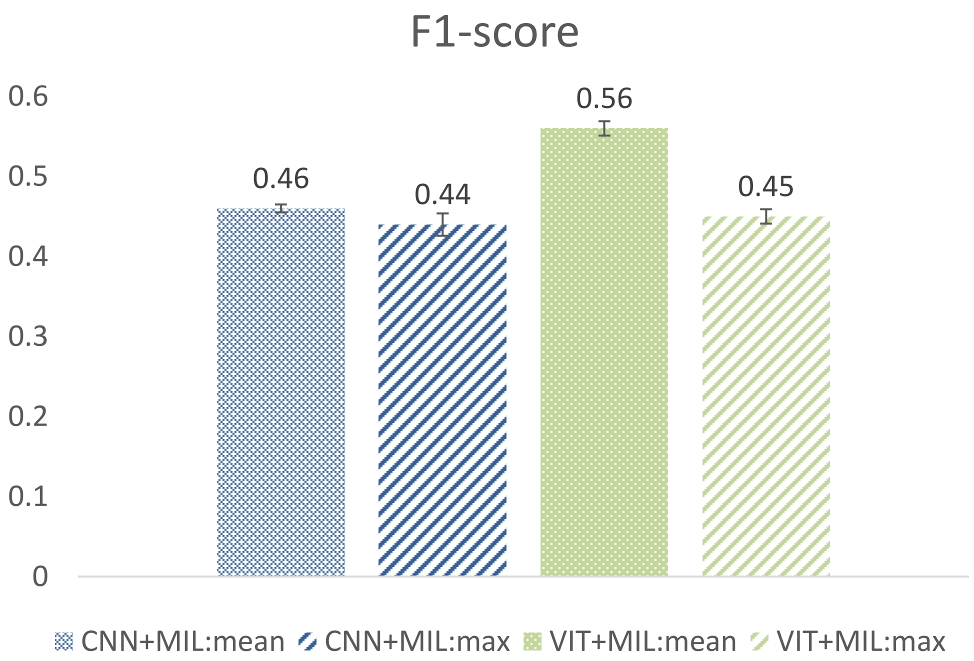
- Baseline model : CNN combined with MIL
- ViT+MIL: mean and max pooling for ImageNet and PANDA pre-training
- Hierarchical ViT

RESULTS

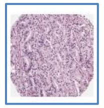
Results of CNN+MIL and ViT+MIL (ImageNet weights)

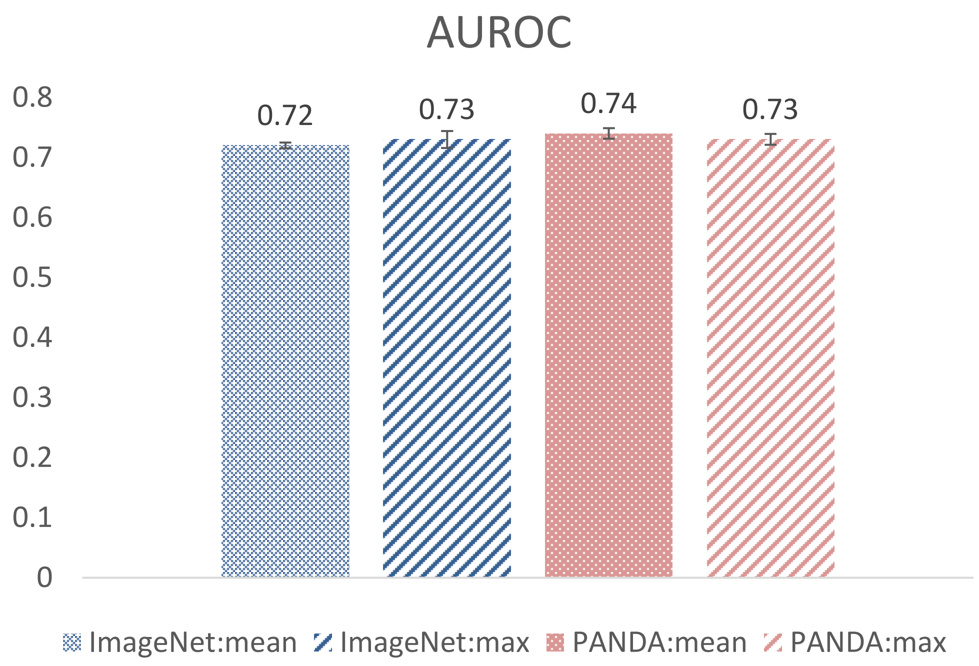


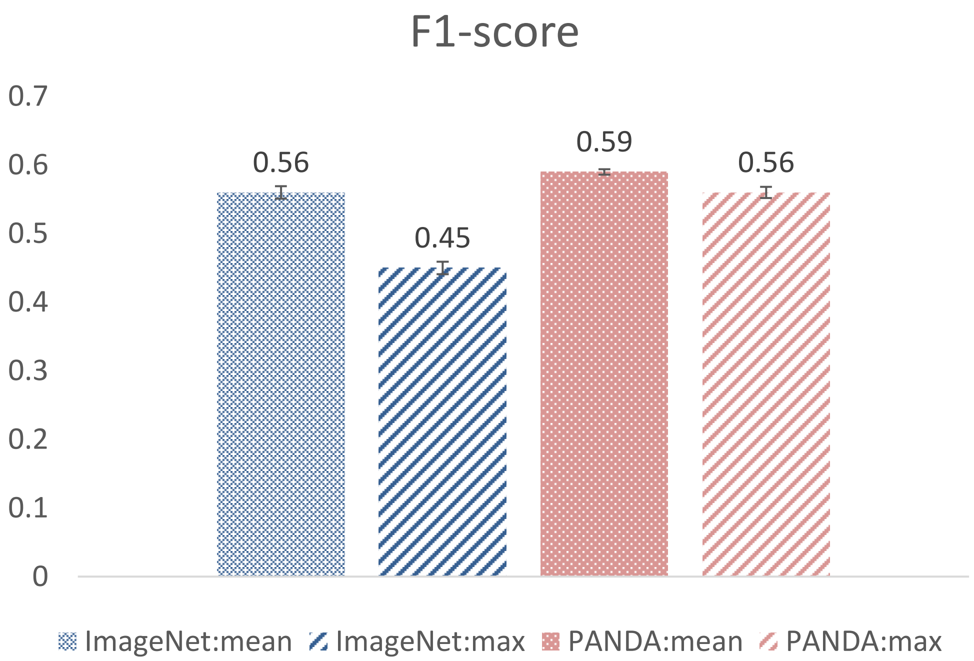




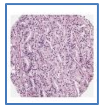
Results of ViT+MIL (ImageNet and PANDA weights)

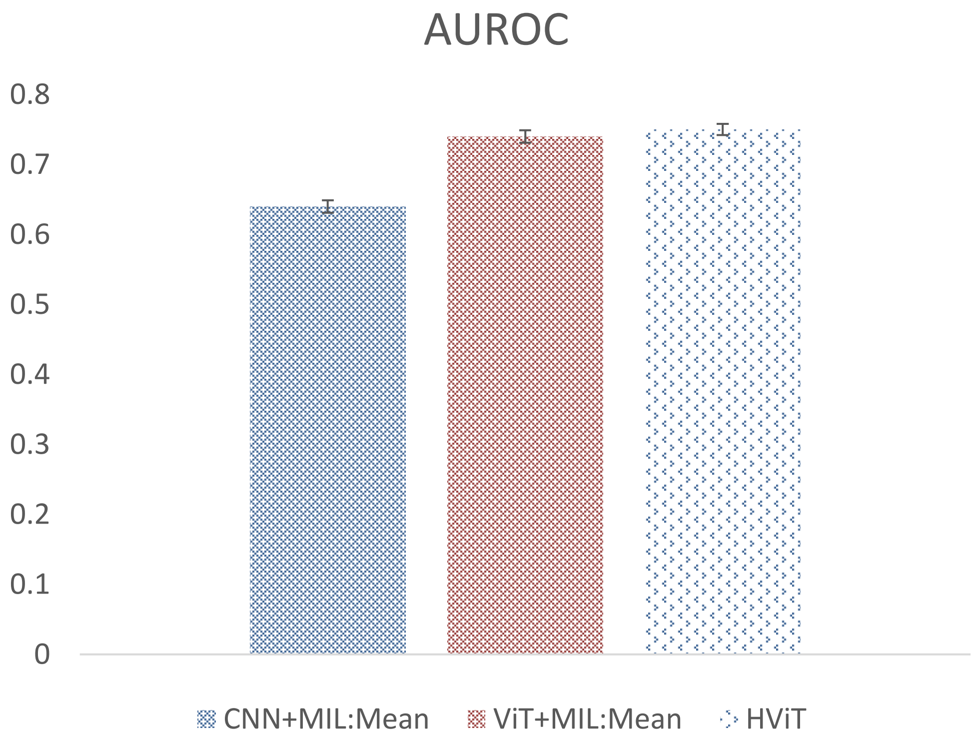


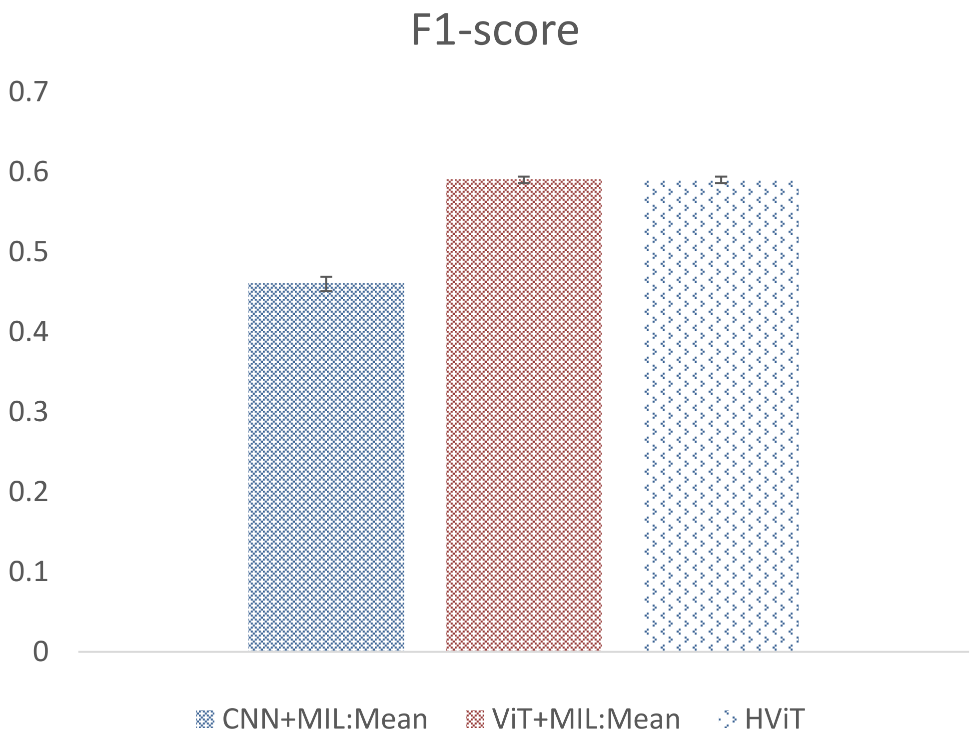




Results of HViT







Summary

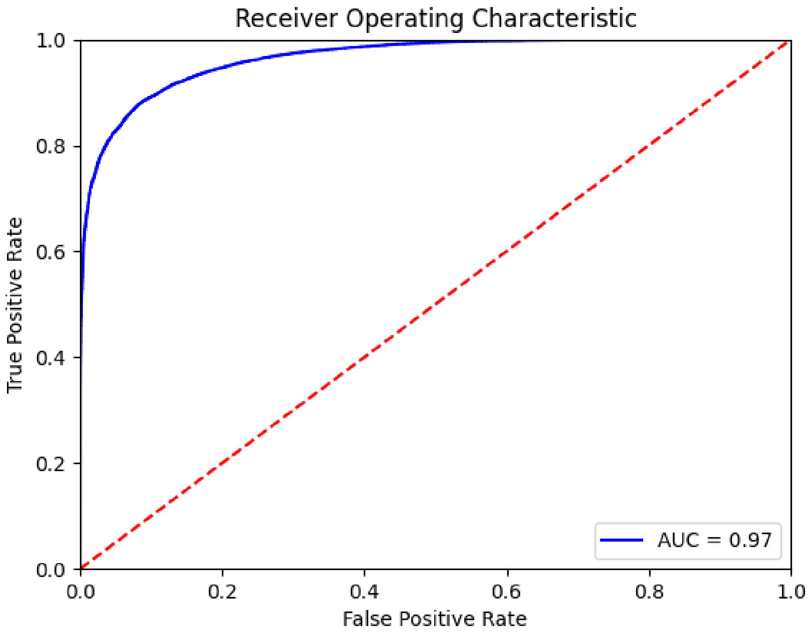
- **R1:** ViT outperformed CNN by 10 percent on AUROC and F1-score
- **R2:** Pre-training of ViT on domain-specific data did not have significant impact
- **R3:** Hierarchical ViT provided similar results compared to ViT+MIL

Further work

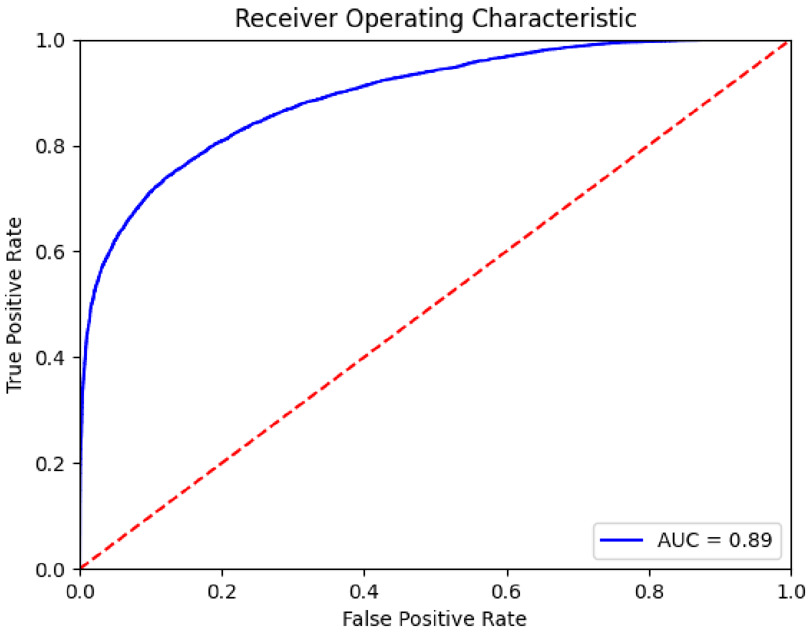
- Pre-train on more data
- Decrease model complexity (e.g. less encoder layers)
- Finetuning in multiple layers instead of only classification layer
- Other pre-training strategies (e.g. self supervised)

Thank you

Results of cancer detection

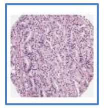


ViT



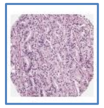
CNN

Results of CNN+MIL and ViT+MIL (ImageNet weights)



Model	Pooling strategy	AUROC	F1-score
CNN+MIL	mean	0.64±0.008	0.46±0.005
CNN+MIL	max	0.63±0.008	0.44±0.014
ViT+MIL	mean	0.72±0.005	**0.56±0.009**
ViT+MIL	max	**0.73±0.014**	0.45±0.009

Results of ViT+MIL (ImageNet and PANDA weights)



Model	Pretraining	Pooling strategy	AUROC	F1-score
ViT+MIL	ImageNet	mean	0.72±0.005	0.56±0.009
ViT+MIL	ImageNet	max	0.73±0.014	0.45±0.009
ViT+MIL	PANDA	mean	**0.74±0.009**	**0.59±0.004**
ViT+MIL	PANDA	max	0.73±0.009	0.56±0.008