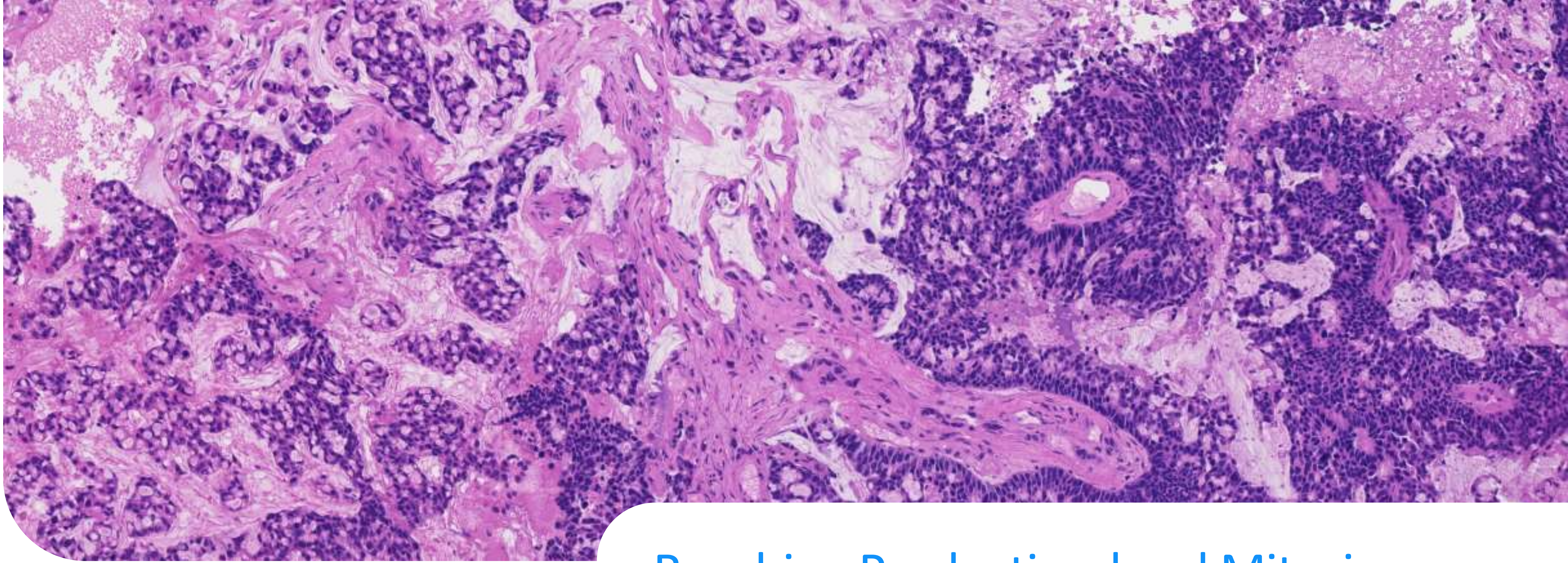


Reaching Production-level Mitosis Detection Performance through Competitive Challenges

Nikolas Stathonikos

Principal Investigator AI development & Implementation

UMC Utrecht

London – December 2023

“Holy trinity” of subsets in machine learning

Training	Validation	Test
Used to find the optimal **parameters** of the model	Used to find the optimal **model** (hyper-parameters)	Used to estimate the **performance** of the optimal model
w	$f(\cdot)$	$\|\|\hat{y} - y\|\|$

UMC Utrecht

However, poor experimental practice is difficult to detect

Training	Validation	Test

Repeatedly evaluating models on the test set
Creating new subset divisions that “work”

Different studies use different evaluation metrics and test datasets, which makes comparison difficult

Grand Challenges

Opening your data AND ensuring fair use:

Medical image analysis challenges

Friendly competitions in which researchers evaluate their solutions on the **same data** with the **same criteria**, in a **blinded manner**.

The obvious solution: open source and open datasets.

However, this puts the "burden of proof" on the community instead on the authors.

Problematic in very active fields such as machine learning and medical image analysis.

Grand challenges

Setting up a challenge

1. Define a task
2. Curate a dataset
3. Publish the dataset
4. Publish a clear evaluation procedure
5. Invite researchers to submit methods
6. (optional) Organize a challenge workshop

Assessment of Mitosis Detection Algorithms 2013

AMIDA13 | MICCAI Grand Challenge

Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge

UMC Utrecht

Mitosis challenges in the past

- MITOS12
- AMIDA13
- MITOS-ATYPIA14
- TUPAC16
- MIDOG21
- MIDOG22

Assessment of Mitosis Detection Algorithms 2013
AMIDA13 | MICCAI Grand Challenge

Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge

Mitosis challenges in the past

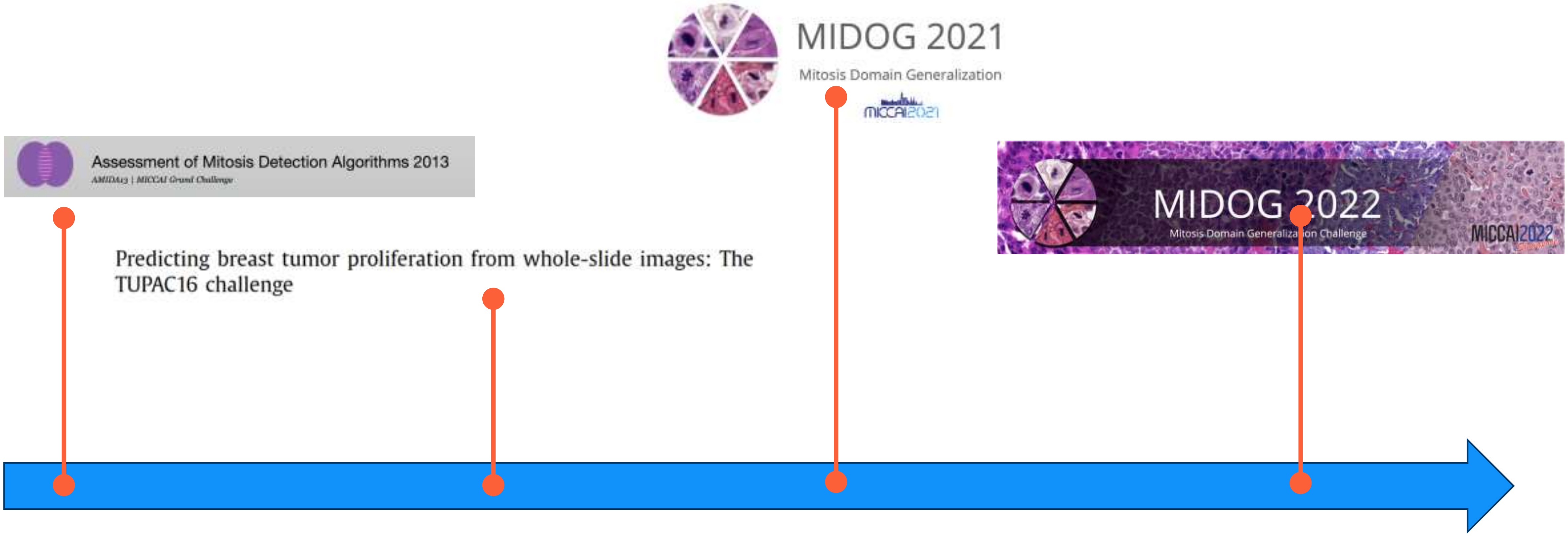
- MITOS12
- **AMIDA13**
- MITOS-ATYPIA14
- **TUPAC16**
- **MIDOG21**
- **MIDOG22**

UMC Utrecht as co-organizer

Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge

UMC Utrecht

Timeline of Mitosis challenges through the years



Timeline of Mitosis challenges through the years

First application of deep learning for medical image analysis

Assessment of Mitosis Detection Algorithms 2013

AMIDA13 | MICCAI Grand Challenge

Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge

Incorporated molecular data as well

Multi scanner dataset

MIDOG 2021

Mitosis Domain Generalization

MICCAI2021

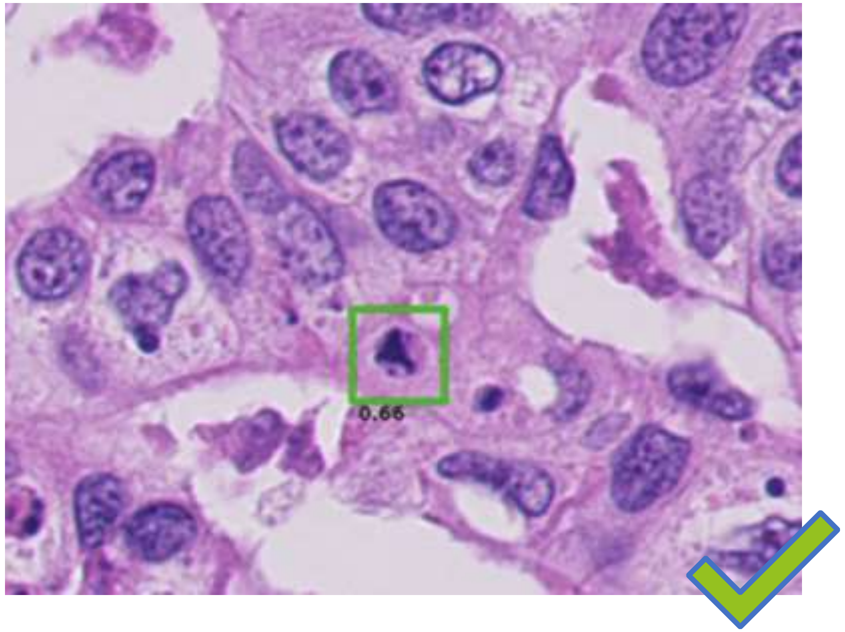
Multi scanner, multi center, multi species

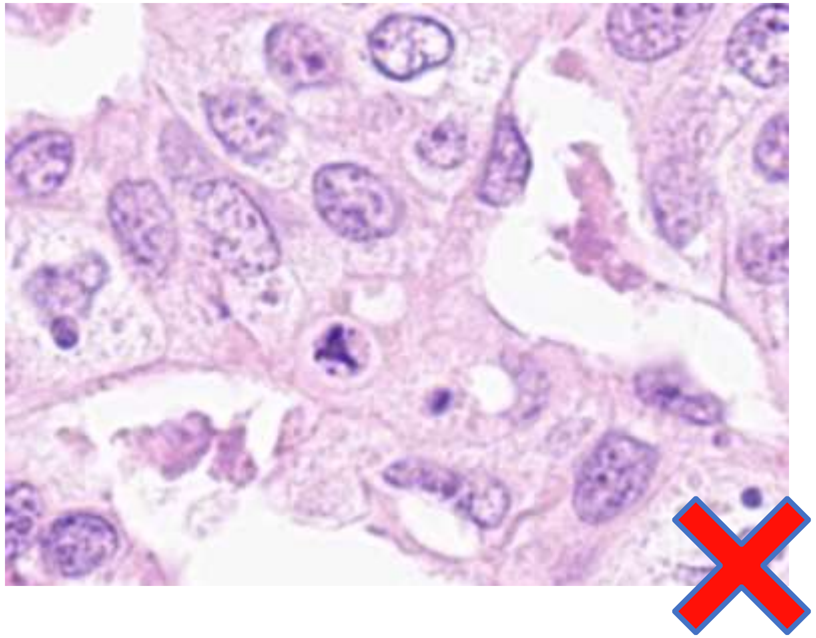
MIDOG 2022

Mitosis Domain Generalization Challenge

MICCAI2022

The task





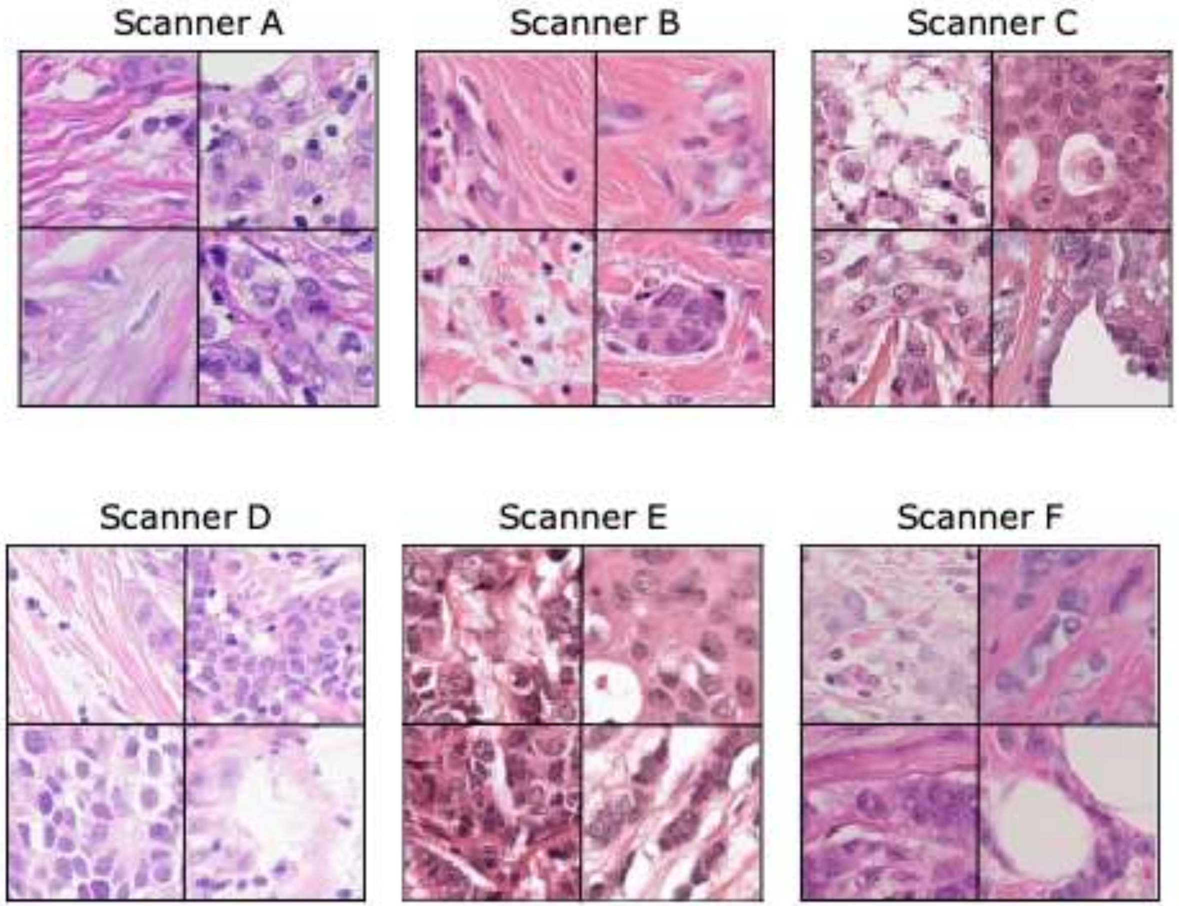
- Create a scanner-invariant mitosis detection algorithm

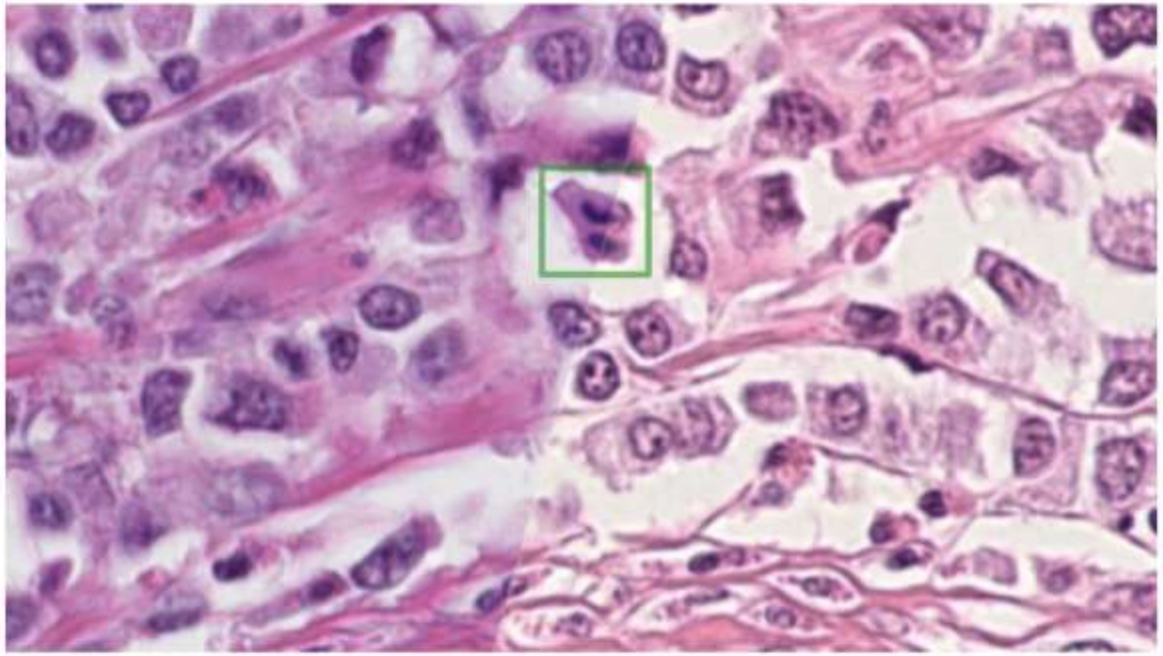
MIDOG is Born!

Organizers

Marc Aubreville	Technische Hochschule Ingolstadt, Germany
Christof Bertram	Institute of Pathology, University of Veterinary Medicine, Vienna, Austria
Mitko Veta	Medical Image Analysis Group, TU Eindhoven, The Netherlands
Nikolas Stathonikos	Pathology Department, UMC Utrecht, The Netherlands
Robert Klopfleisch	Institute of Veterinary Pathology, Freie Universität Berlin, Germany
Katharina Breininger	Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander- Universität Erlangen-Nürnberg, Germany
Natalie ter Hoeve	Pathology Department, UMC Utrecht, The Netherlands
Francesco Ciompi	Computational Pathology Group, Radboud UMC Nijmegen, The Netherlands
Andreas Maier	Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

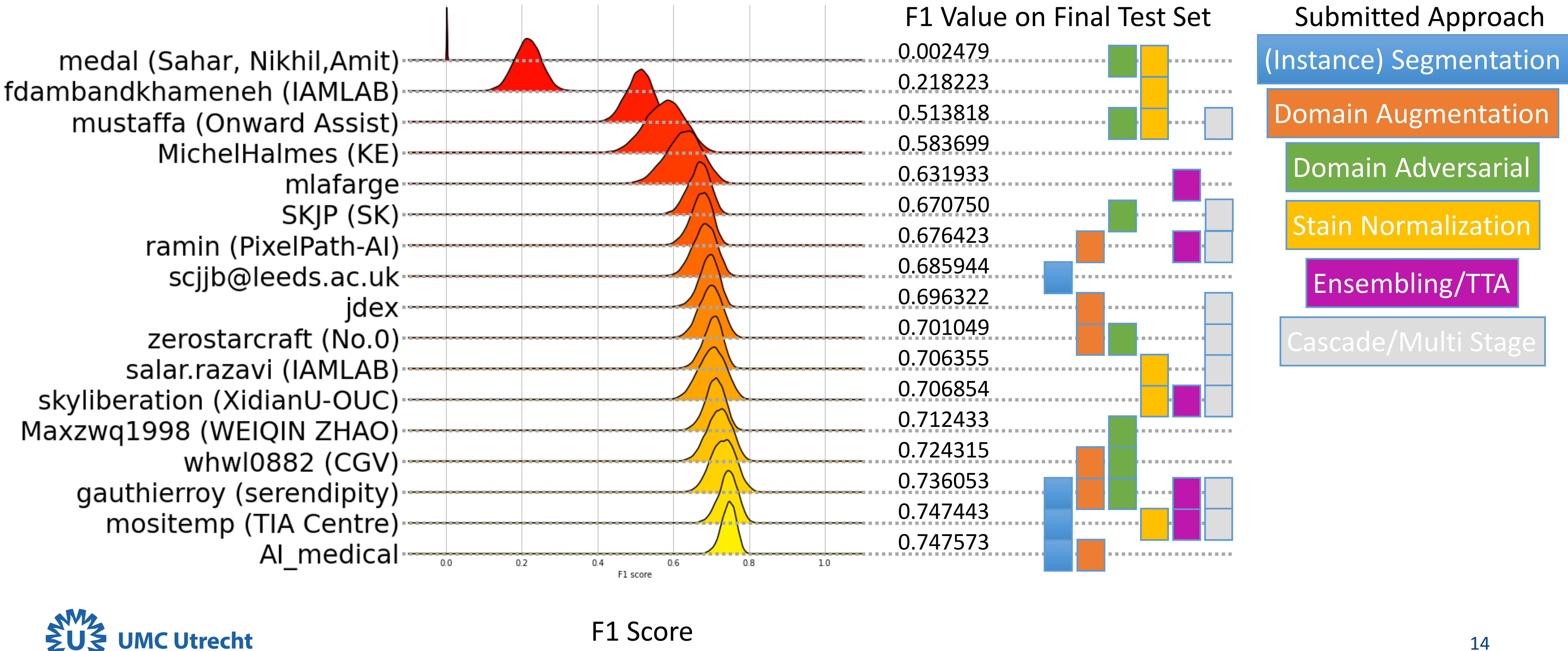
Overview of dataset





Same image – different scanners

Results



UMC Utrecht

Examples of detections

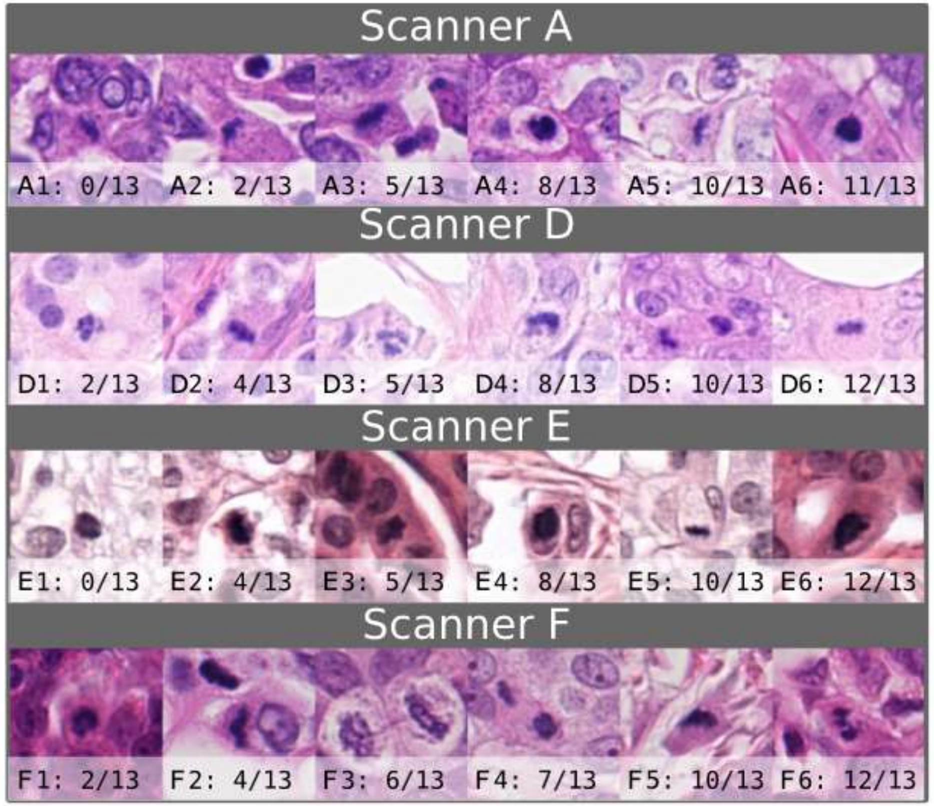


Fig. 8. Examples of ground truth mitotic figures (true positives and false negatives), ordered by the count of models voting for it. The numbers (x/13) indicate, how many models voted for this cell to be a mitotic figure. The rows are stratified by the number of models to give examples for the complete distribution in Fig. 7.

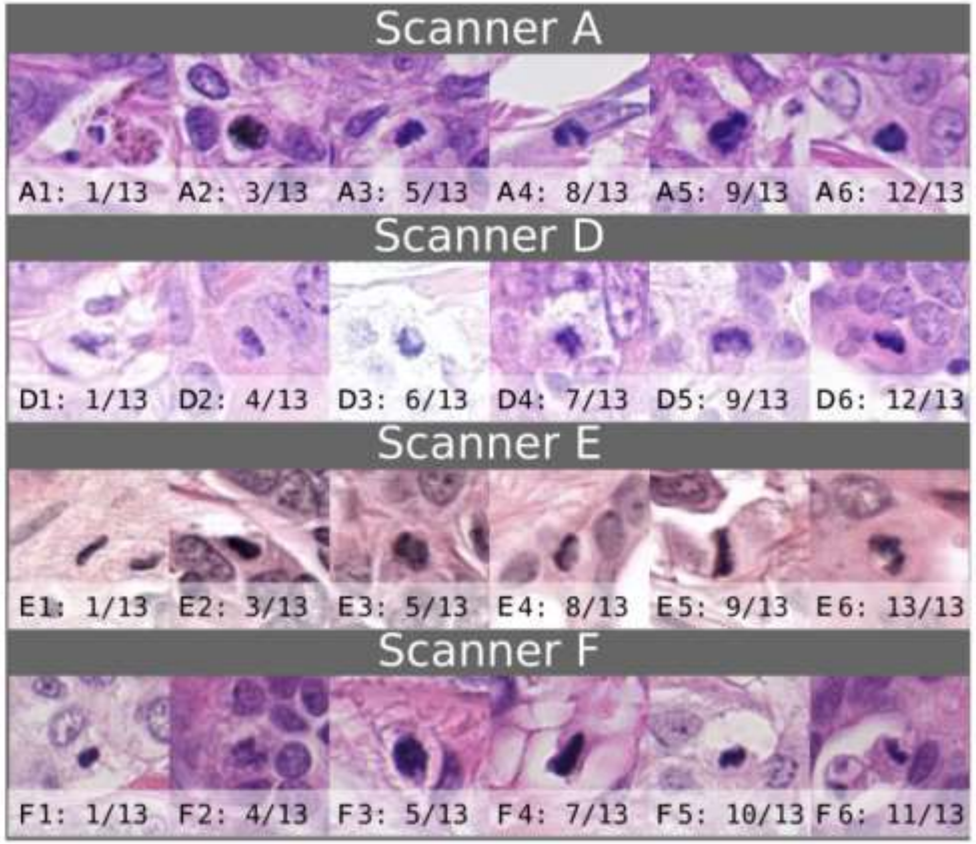
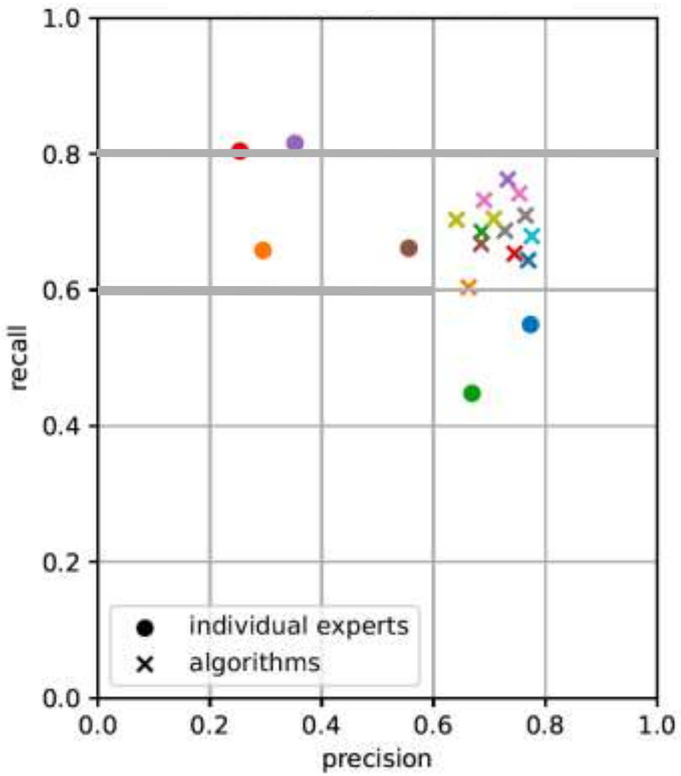


Fig. 9. Examples of false positives, ordered by the count of models voting for it. The numbers (x/13) indicate, how many models voted for this cell to be a true mitotic figure.

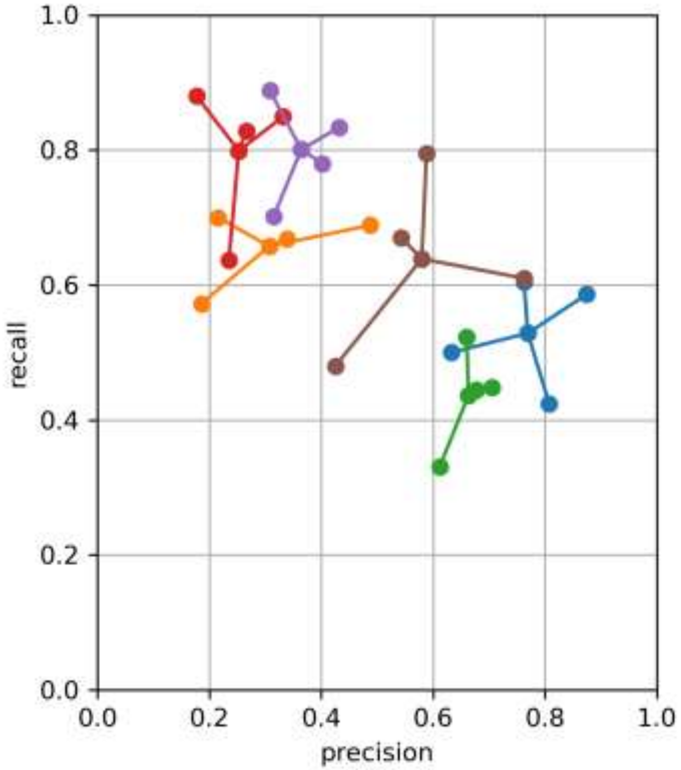
Comparison to human experts

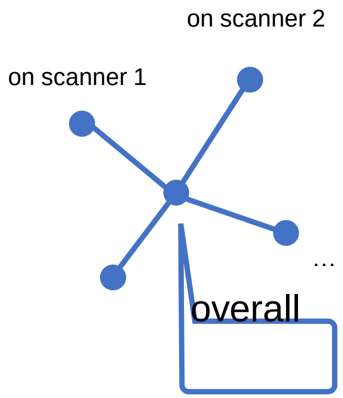
Results of experts and algorithms



Human experts were much less consistent.

Scanner-dependency of experts

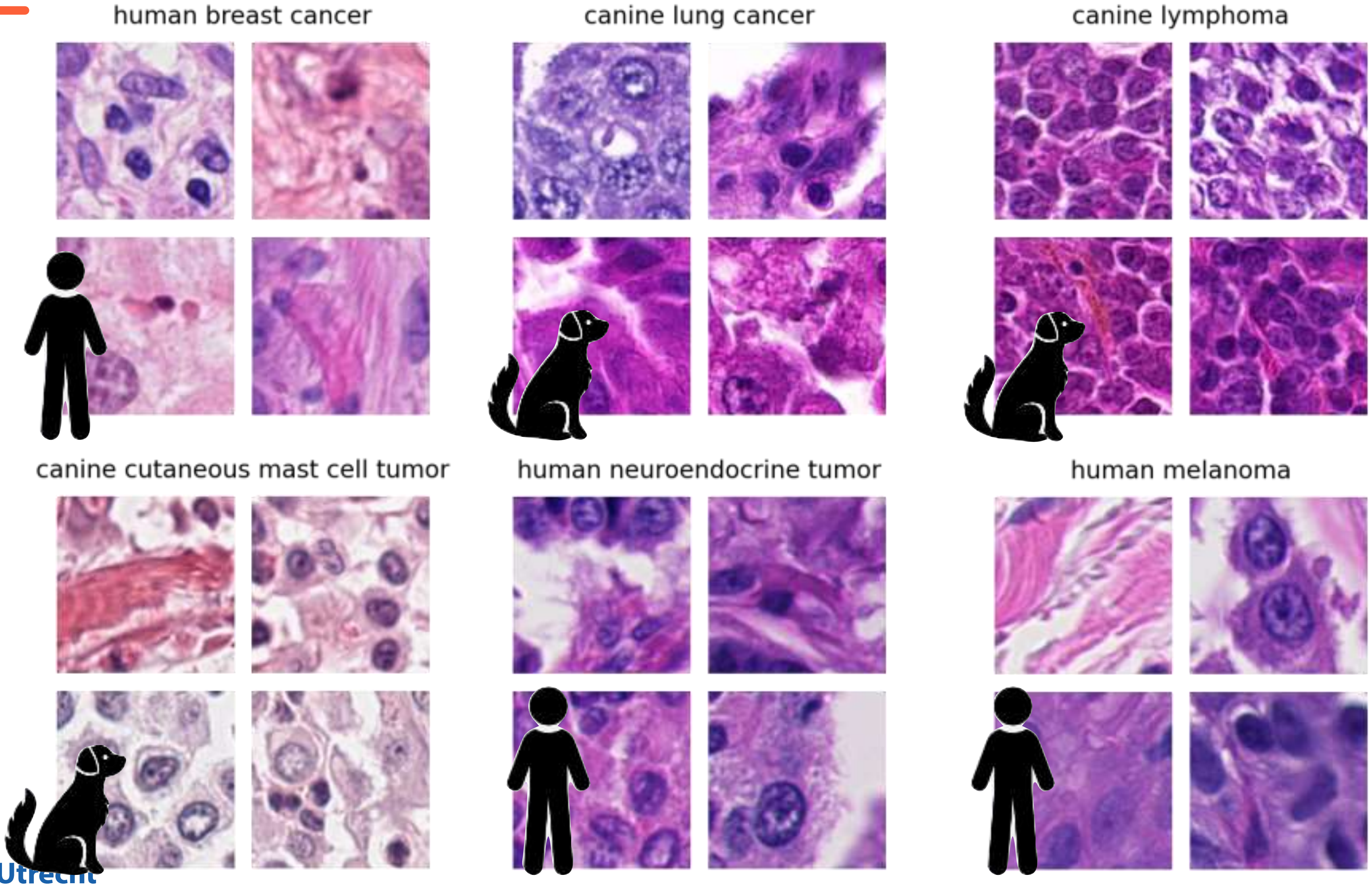




… and were also strongly domain dependent.

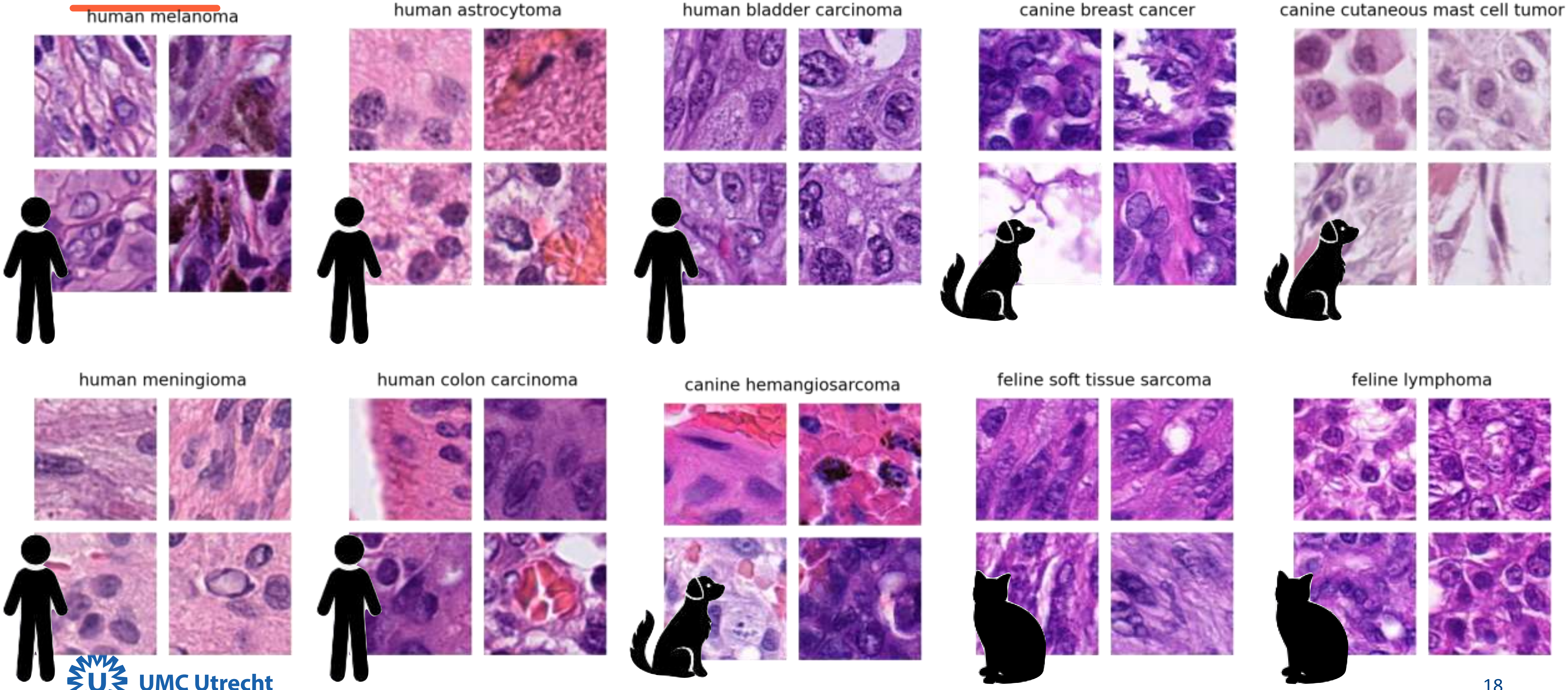
UMC Utrecht

MIDOG 2022: Training set



UMC Utrecht

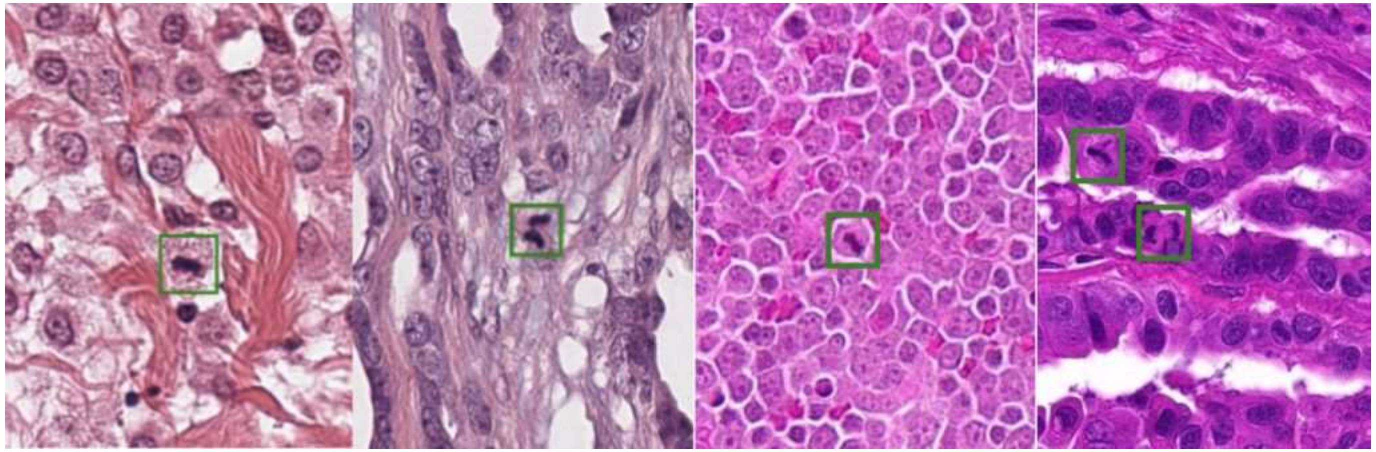
MIDOG 2022: Test set



UMC Utrecht

Performance on different tissue

MIDOG 2022
Mitosis Domain Generalization Challenge
MICCAI2022



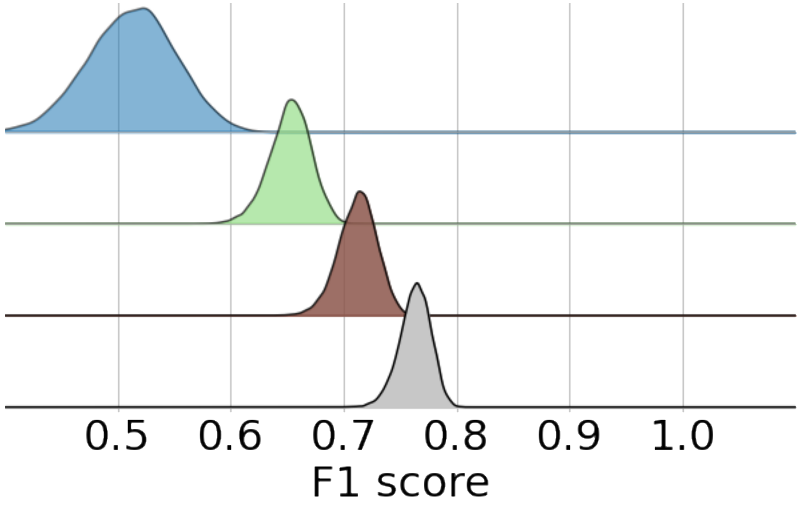
Note: Also includes differences in tissue processing, species, and scanners.

Trained on breast cancer (4 scanners)

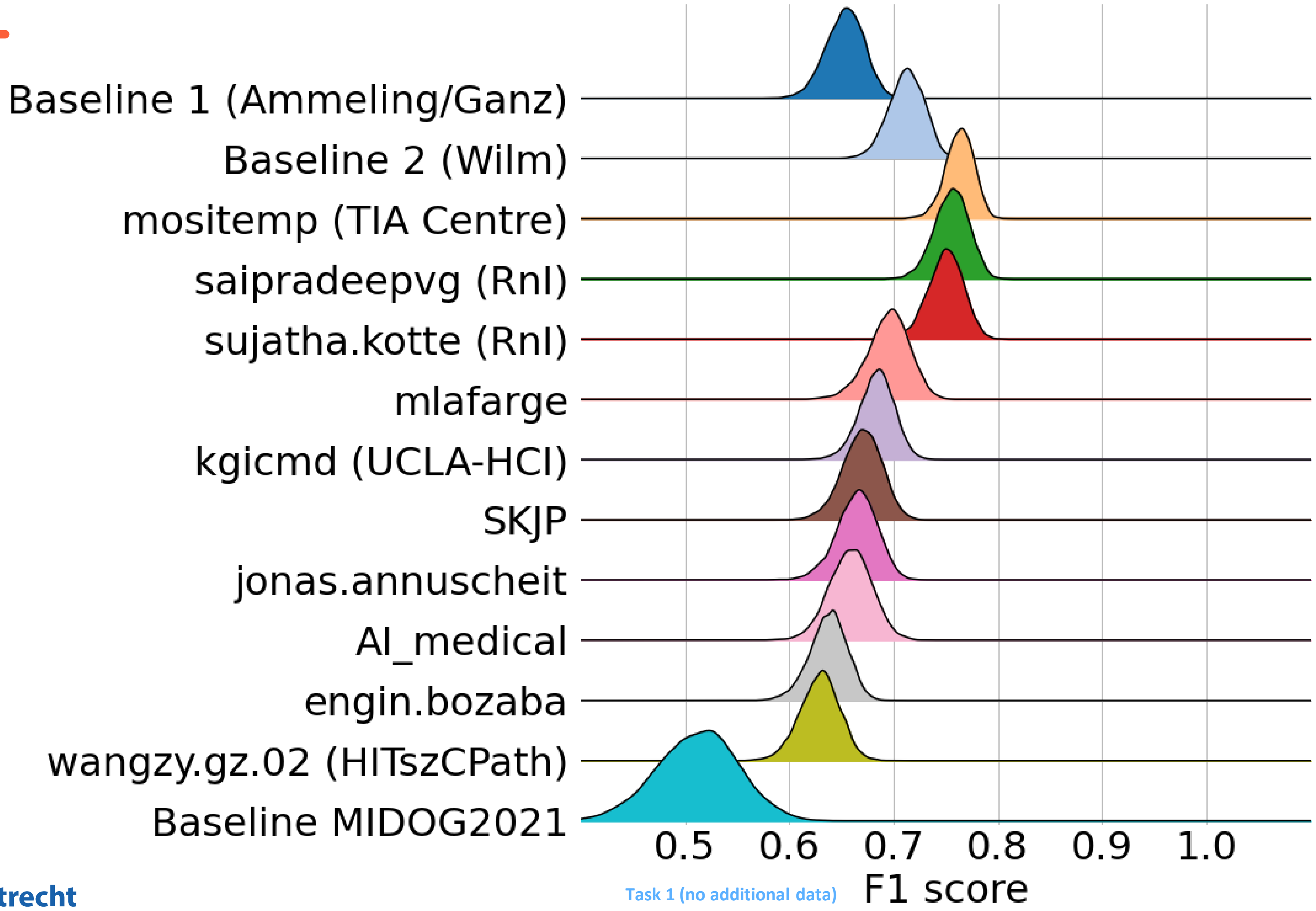
Trained on 6 tumor types

Trained on 6 tumor types + adversarial

Winner (augmentation+ensembling)



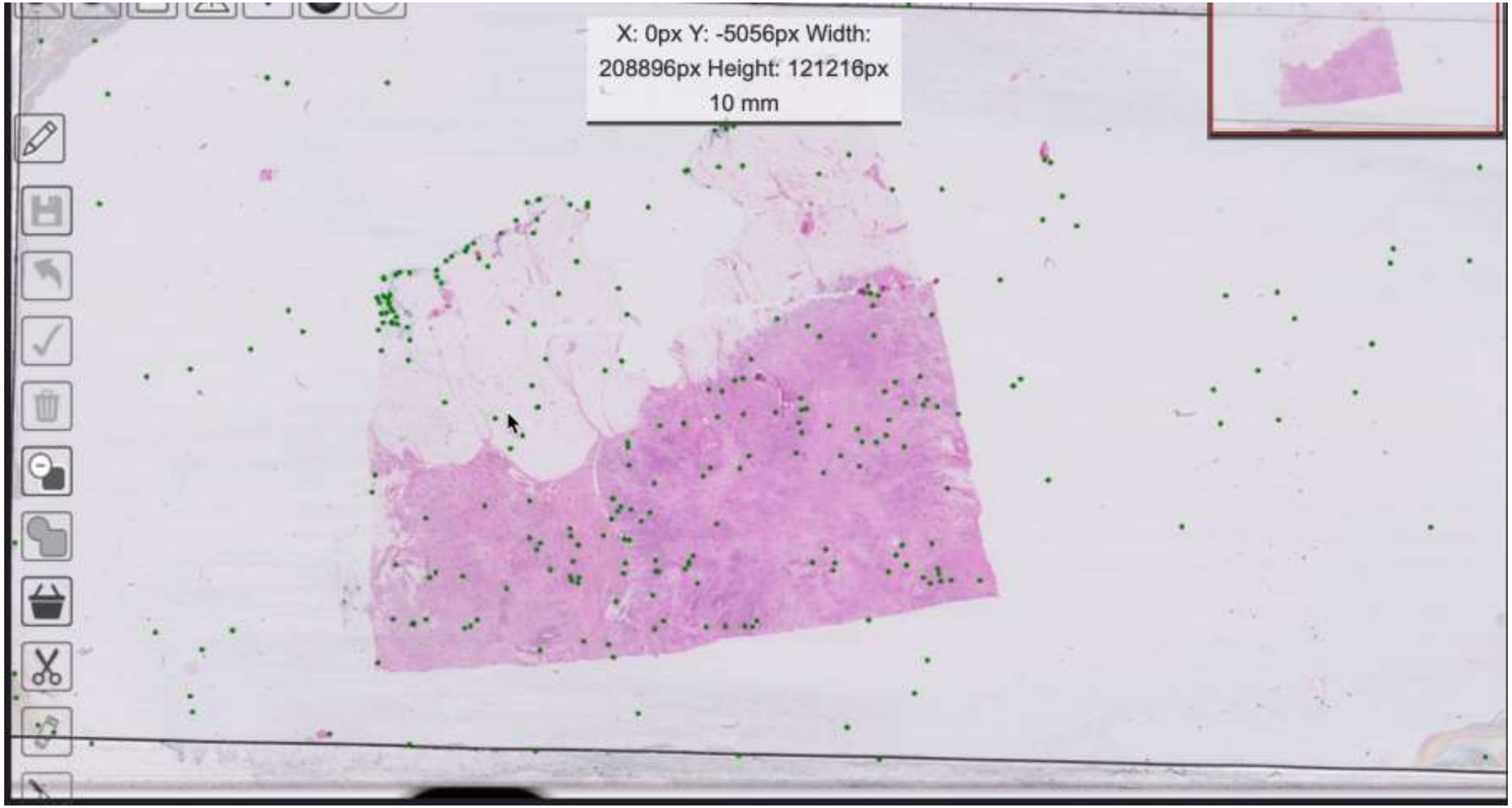
Results MIDOG22



Task 1 (no additional data)

UMC Utrecht

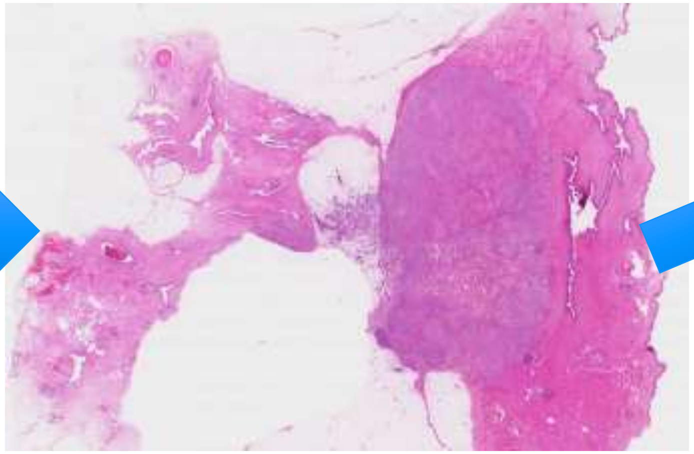
Robustness on WSIs



Grand challenges can stimulate the production of nice models!

And then?

Mitosis detector in clinical practice



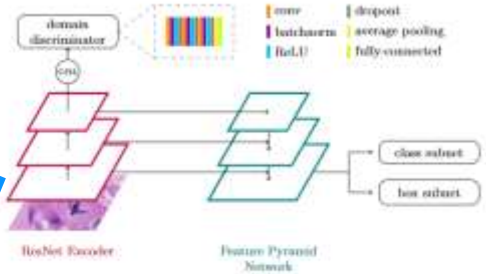
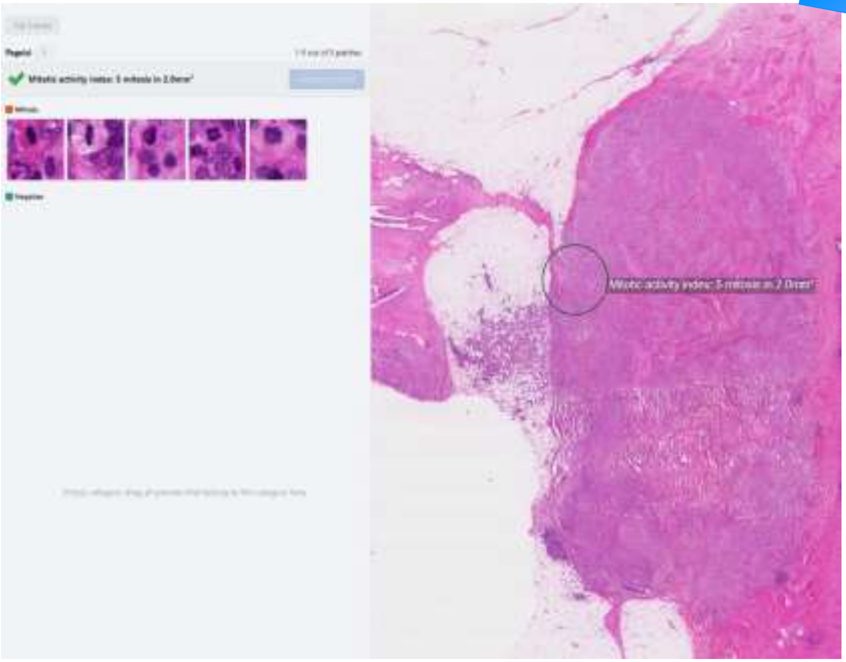


Fig. 2: Domain adversarial RetinaNet architecture.



Mitosis detector in clinical practice

Machine learning models in PACS Pathology. Using PACS API we have implemented an AI model that can detect mitotic cells in breast cancer slides.

Model in the background

The model is monitoring all incoming scans. When a scan fits the criteria, a job is sent to the HPC to analyze that scan. Average time from scan to results: **7 min**

UMC Utrecht

Iceberg of AI implementation

AI in pathology

- Improve reproducibility
- Quantitative diagnosis
- Specialist support
- Improve turnaround

AI result on screen

Heatmaps, coordinates, annotated regions
Structured reporting
Case worklist prioritized

Iceberg of AI implementation

AI in pathology

- Improve reproducibility
- Quantitative diagnosis
- Specialist support
- Improve turnaround

Validation

IVDR, MDR, ISO, Quality registry

Processes

Model retraining, registry,
Experiment registry
AI model triggering, workflow, job scheduling

Hardware

GPUs, CPUs, storage, databases to run AI models, save results
Nvidia vs ASICs

AI result on screen

Heatmaps, coordinates, annotated regions
Structured reporting
Case worklist prioritized

Design choices

Workflow or user initiated?
Heatmap or coordinates?

Software

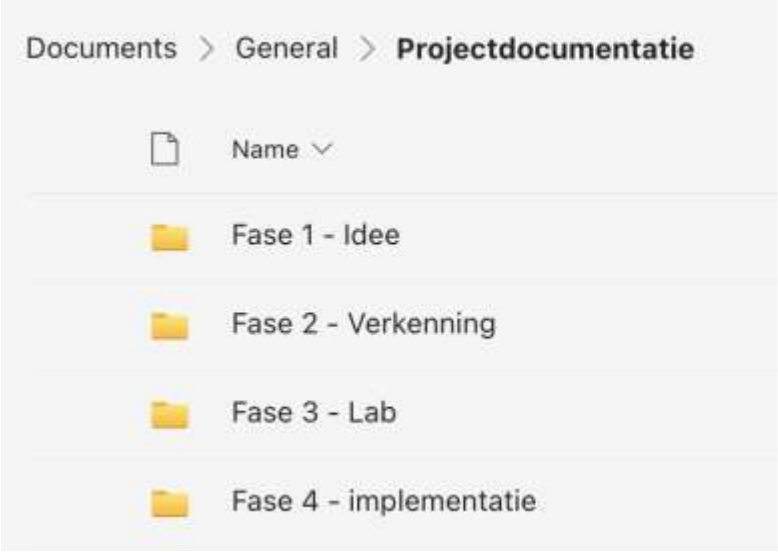
Code repositories, integrations with various LIMS

Platform choice

Tensorflow, PyTorch, MXnet
Tensorflow serving,
NVIDIA Triton Inference Server,
Ray [Serve], Docker, Kubernetes

Validating for clinical use

- Performance in clinical setting?
- IVDR or MDR?
- Safety?



Validation dossier

Validating for clinical use

Phase 1

Classification medical device/software

Business case

Phase 2

Data report

Model acceptation criteria

AI impact assessment

Software development according to quality system

Phase 3

Product development

Architecture diagram

Risk assessment

End user engagement

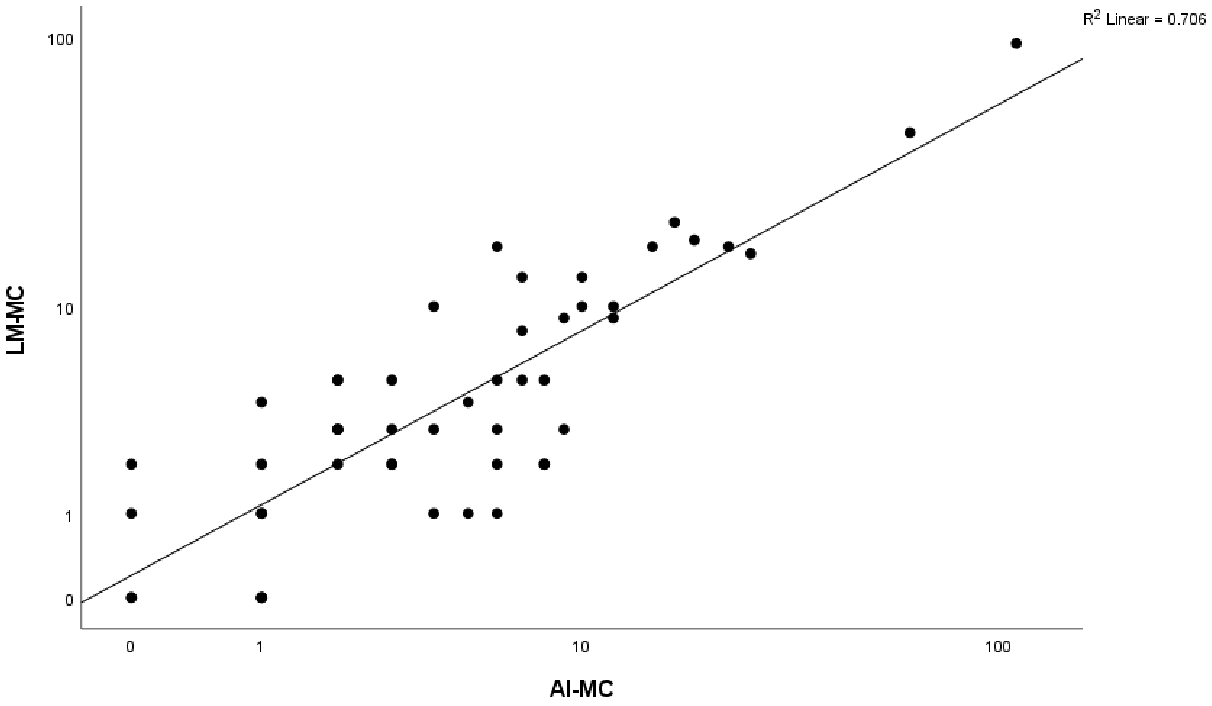
1st Pilot

Phase 4

2nd pilot

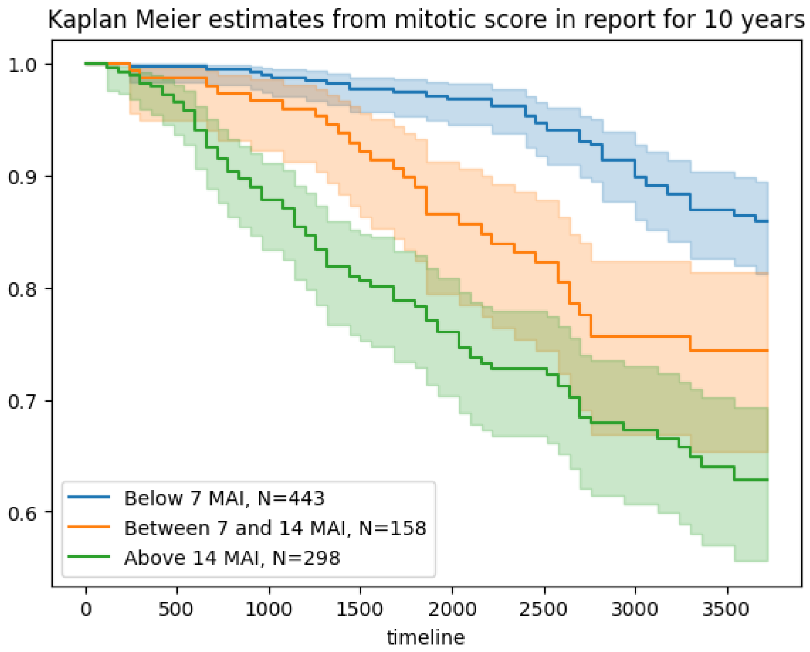
Technical implementation

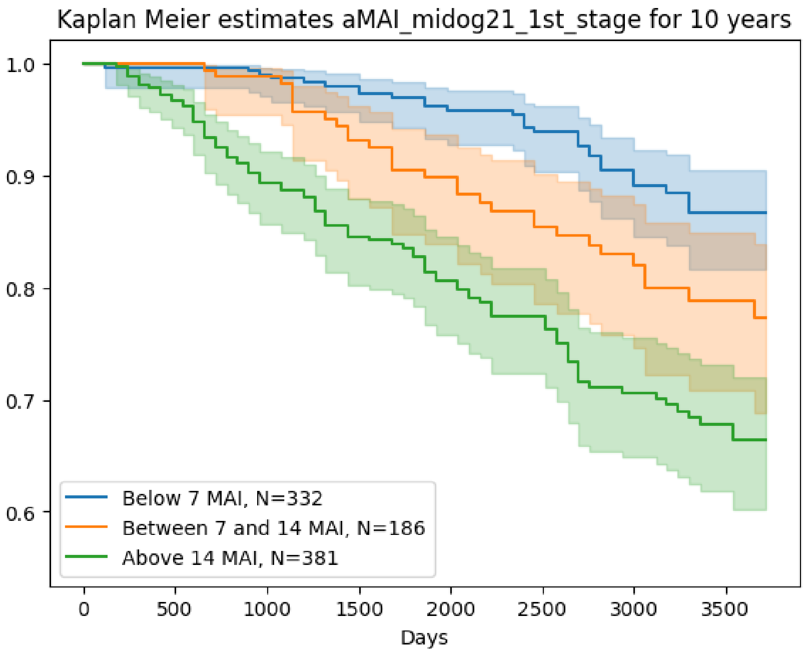
Internal Validation



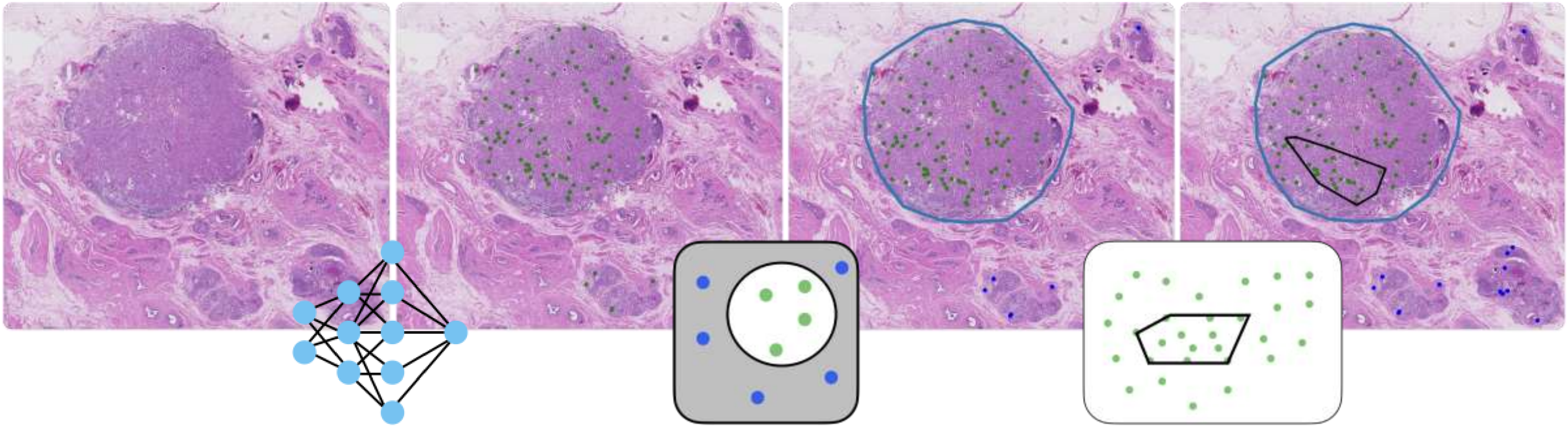
- Compare it to current standard
- Use multiple observers

Bonus validation





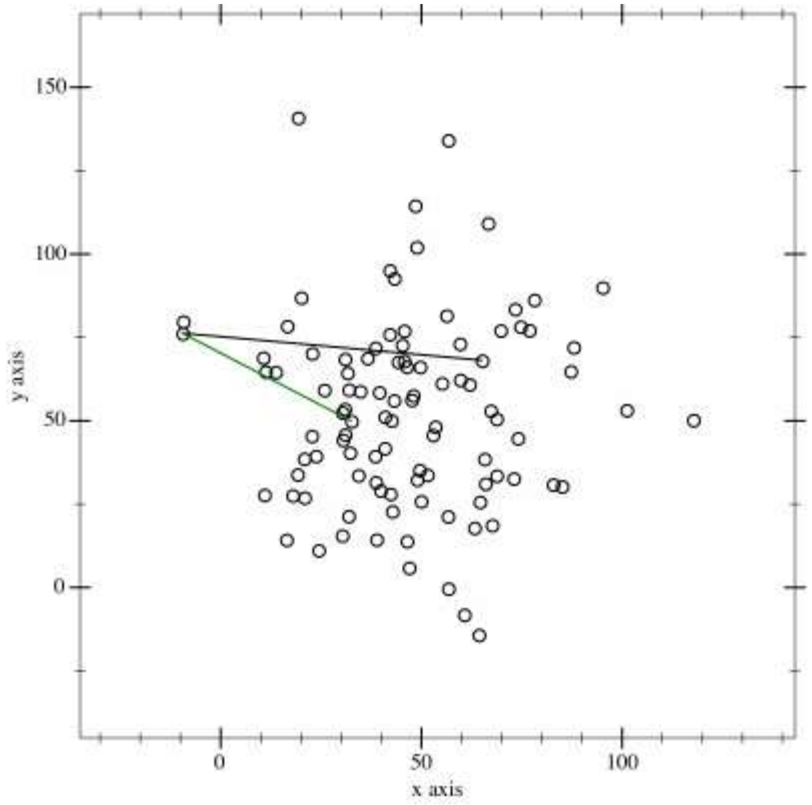
Workflow of mitotic detector in production



Automatic area selection - aMAI

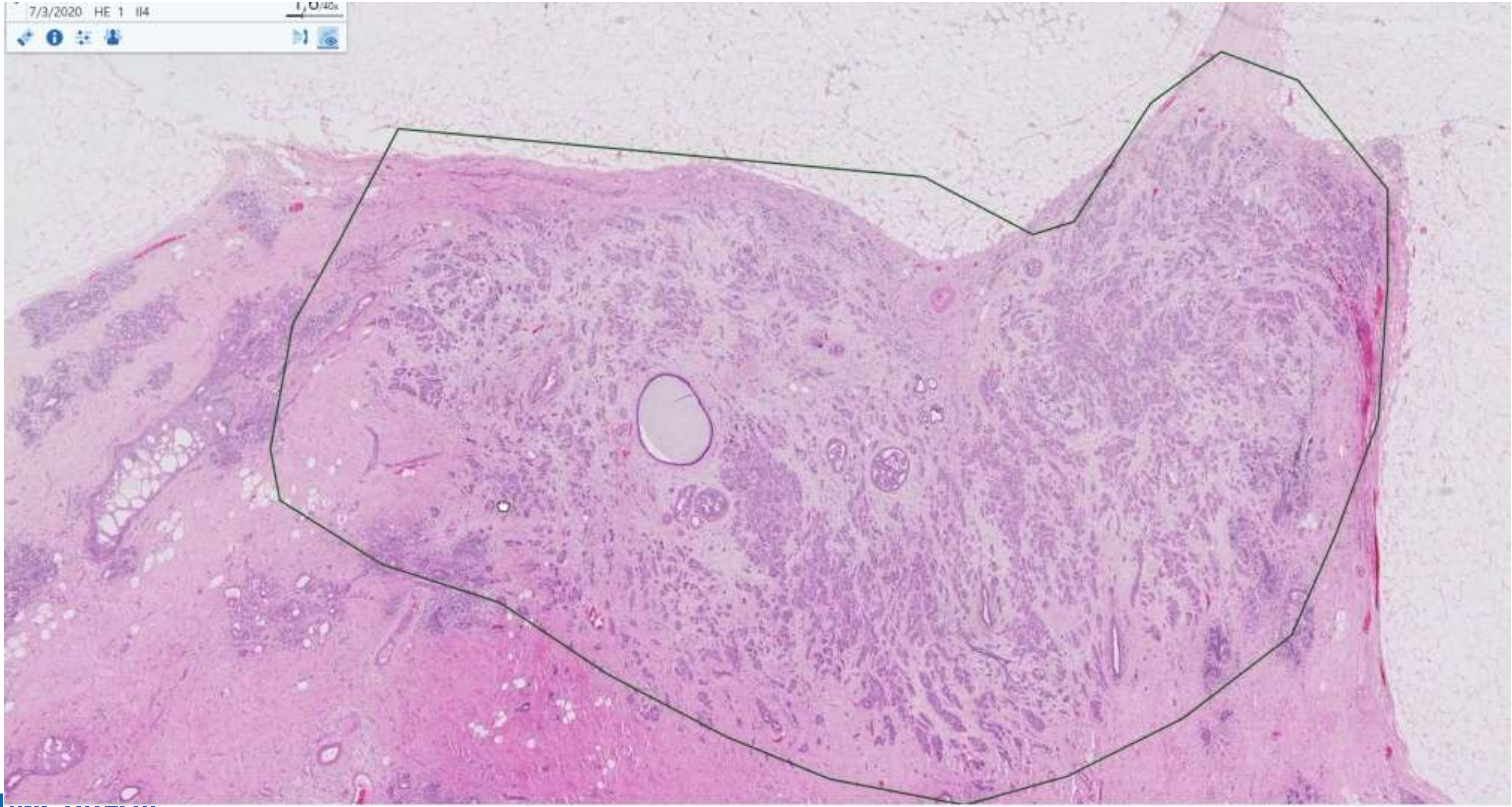
To find the MAI area, we developed an algorithm that can find a fixed area of $2mm^2$ with the most mitosis.
Sounds simple but it is surprisingly difficult.

Most published research relies on fixed shapes (rectangles, circles etc)



UMC Utrecht

Detector in action



PACS integration

- Runs in background – results waiting for pathologist
- Goal – analysis <10’
- Pathologist can correct the output – ultimately responsible

- MongoDB running in the background
- Capturing all results + delta
- Model degradation monitoring
- Service monitoring

Future planning

- Expand from breast only - to generic mitosis detection
- Improve model based on analyzed cases
- Add different metrics/improve diagnostic standards – (mitosis per 1000 cells instead of area)
- Add more classification options (atypical vs typical)

Thank you for your attention!