PyData

Venezia

12 DICEMBRE 2024

Conversations with an AI gynecologist: design of an explainable medical decision support tool

Rosilari Bellacosa
CTO @ SynDiag

It’s me, hi!

Former neuroscientist 🐭
Former ML specialist
Former R&D in Computer Vision
Current CTO @ SynDiag

A few facts about SynDiag:

Launched in 2018

3 founders

9 people

PoliTO Spin Off

5 IPs protected in 5+ countries

2 products

750K in sales in 2024

AI in Healthcare:

Imaging and Diagnostics
Predictive Analytics
Personalized Medicine
Virtual Health Assistants
Administrative Workflow
Remote Monitoring
Clinical Decision Support
…

Goals:

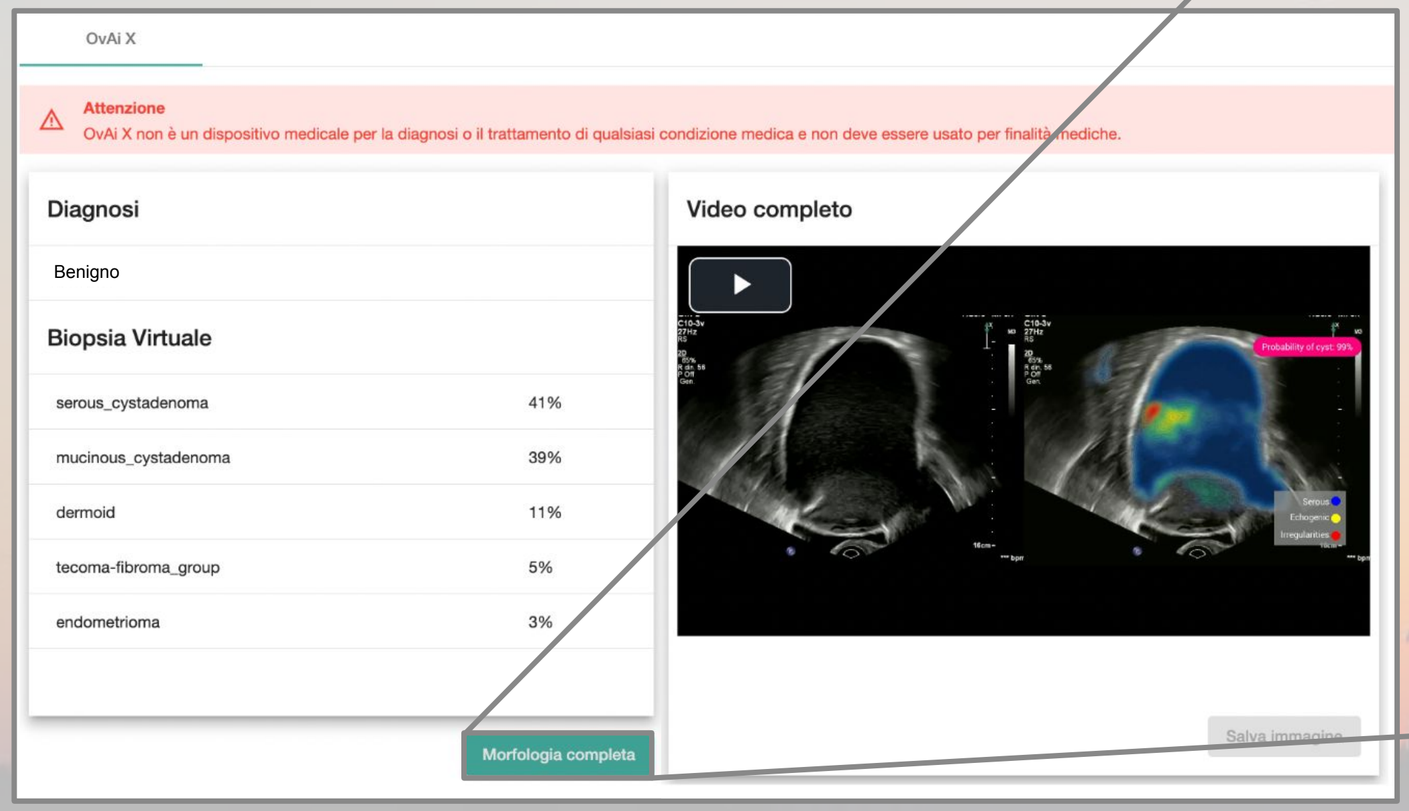
Efficacy and accuracy
Efficiency and costs reduction
Personalized care
Discovering new patterns
…

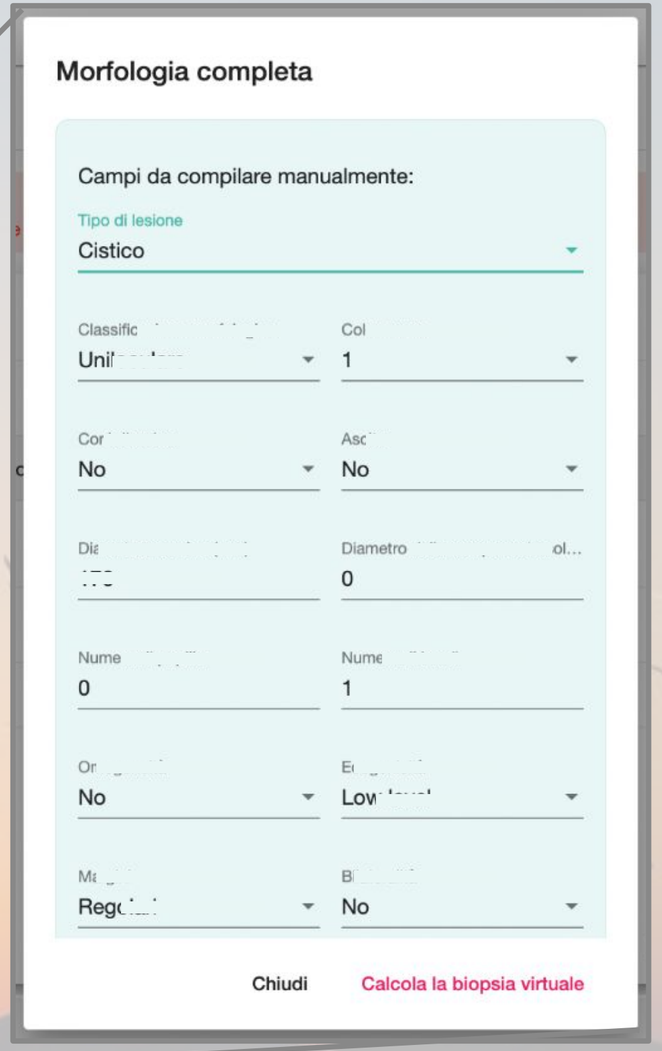
AI in Healthcare:

Andrew Ng on Coursera


SynDiag
United against ovarian cancer.

OvAi case study





OvAi case study:

Diagnostic support for gynecological tumors

Visual tumor description

Diagnostic suggestion

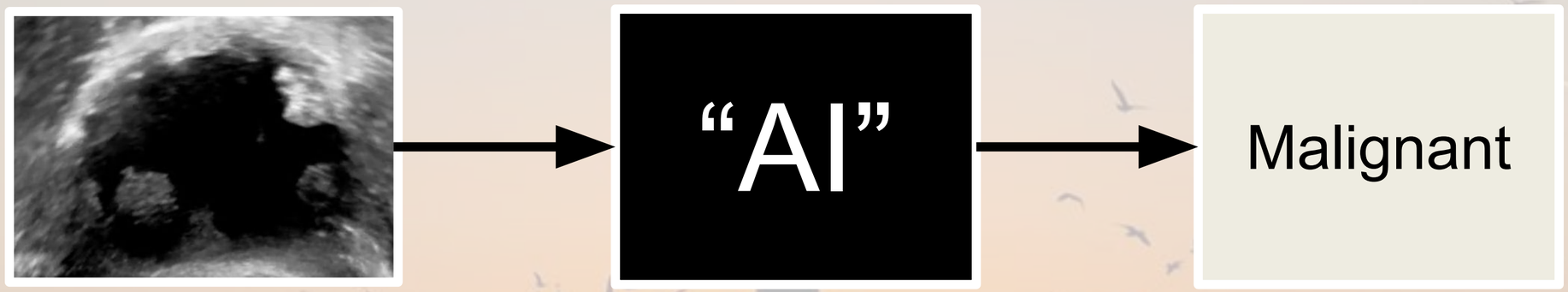
Printable ultrasound report

Main features of OvAi:

Transparency

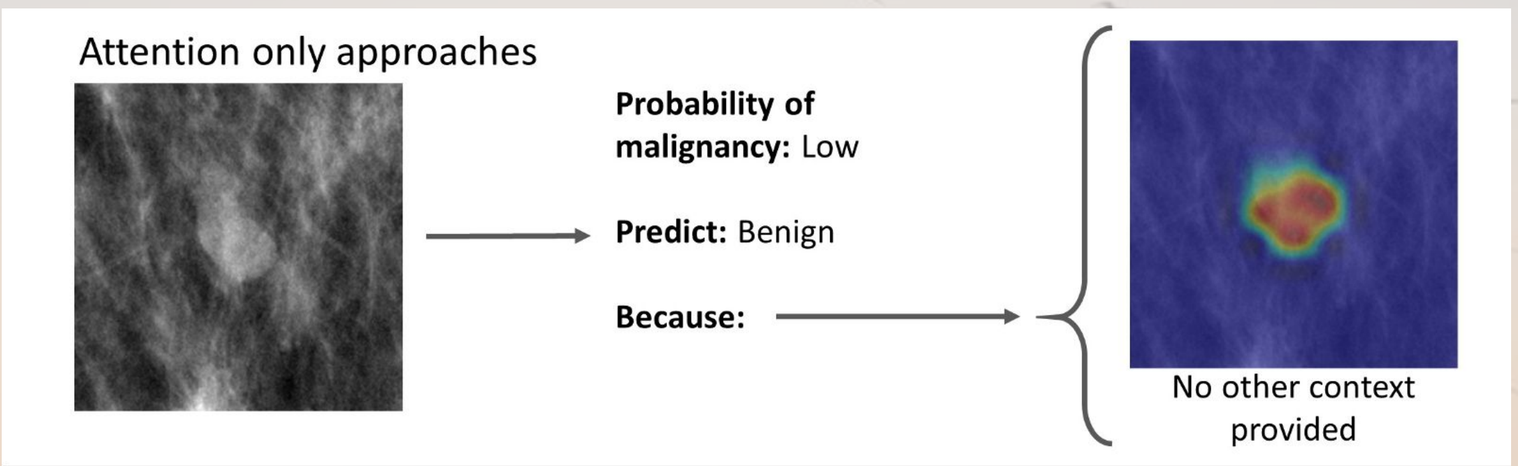
Modularity

Transparency by design



Transparency by design

Limits of XAI



Citation

Transparency by design

Real Data

Official diagnostic guidelines

Visual tumor description

Diagnostic suggestion

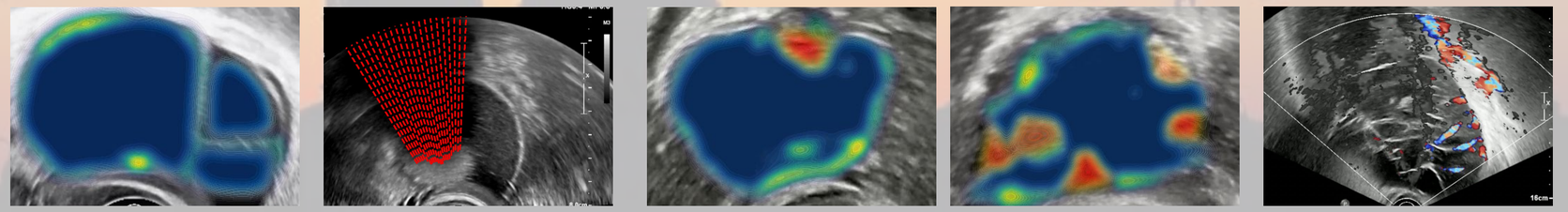
Printable ultrasound report

Transparency by design

Real Data

Official diagnostic guidelines

OvAi Explainable output



Pillars of responsible AI

- Privacy and security
- Fairness and inclusion
- Robustness and safety
- Transparency and control
- Accountability and governance

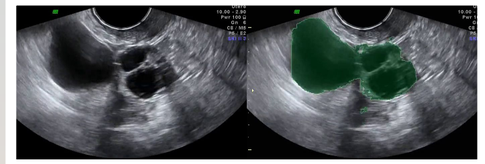
Summary from Meta, Microsoft, Google, IBM, OpenAI, AWS resources

Modularity in OvAi

Visual tumor description

Diagnostic suggestion

Printable ultrasound report

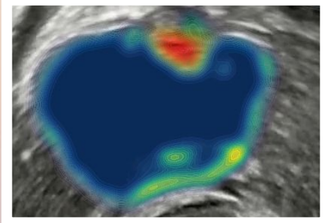


ROI detection and navigation

Diagnostic class suggestion

Benign vs Malignant

Visual description



Tumor identification

Mucinous Cystadenoma

Modularity in OvAi

Pros:

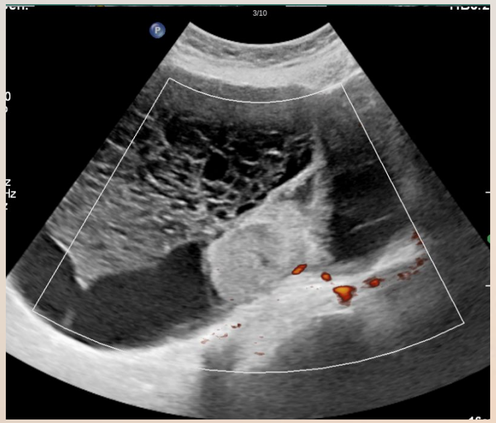
- Specific tasks and technologies
- Work with limited resources
- Reuse modules
- Regulatory requirements

Cons:

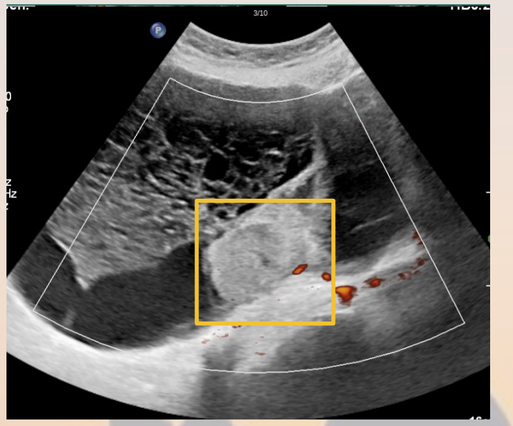
- Integration
- Overall performance
- Maintenance
- Coupling

Module performance: segmentation

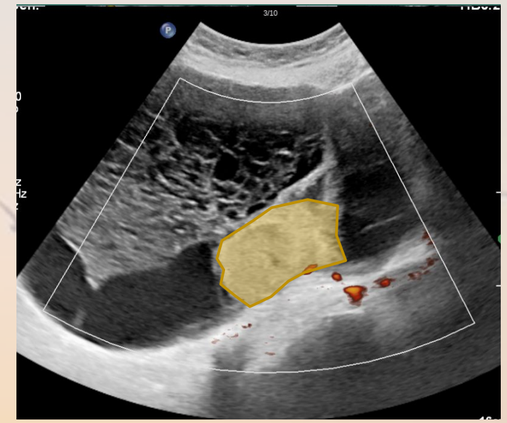
Do physicians really need pixel-level information?



Classification
Is there a tumor component?



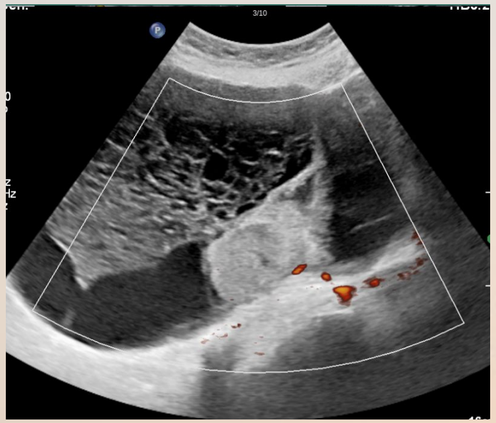
Detection
classification + where is it?



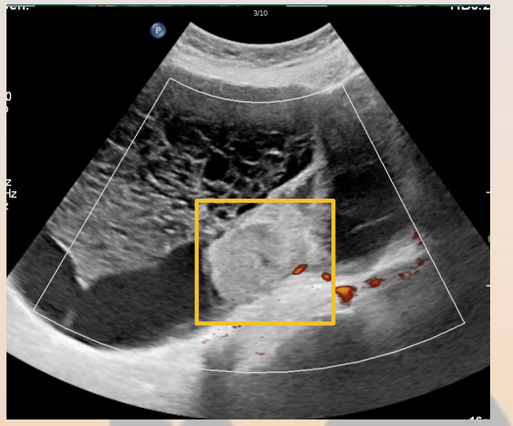
Segmentation
detection + object boundaries

Module performance: segmentation

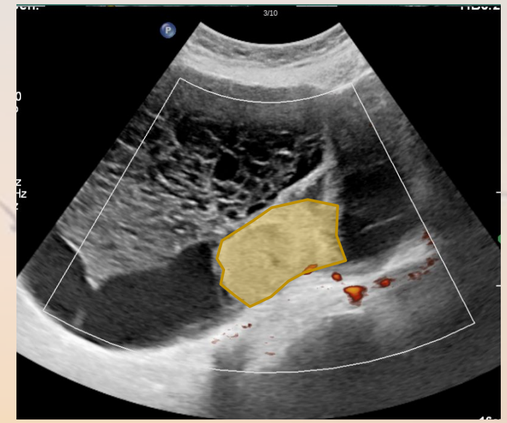
Do physicians really need pixel-level information?



Classification
NAVIGATION (TUMOR Vs NON TUMOR)



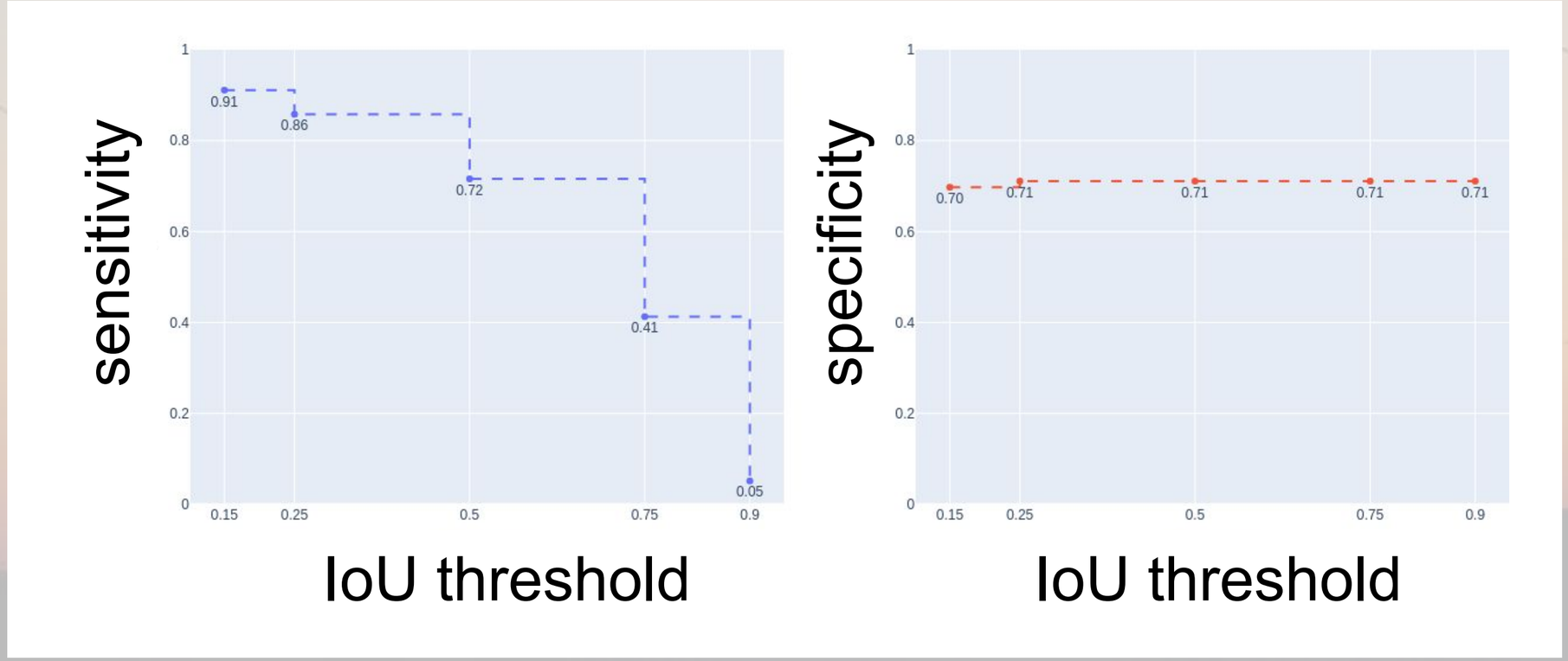
Detection
ROI (TUMOR) DETECTION



Segmentation
TUMOR FEATURE EXTRACTION

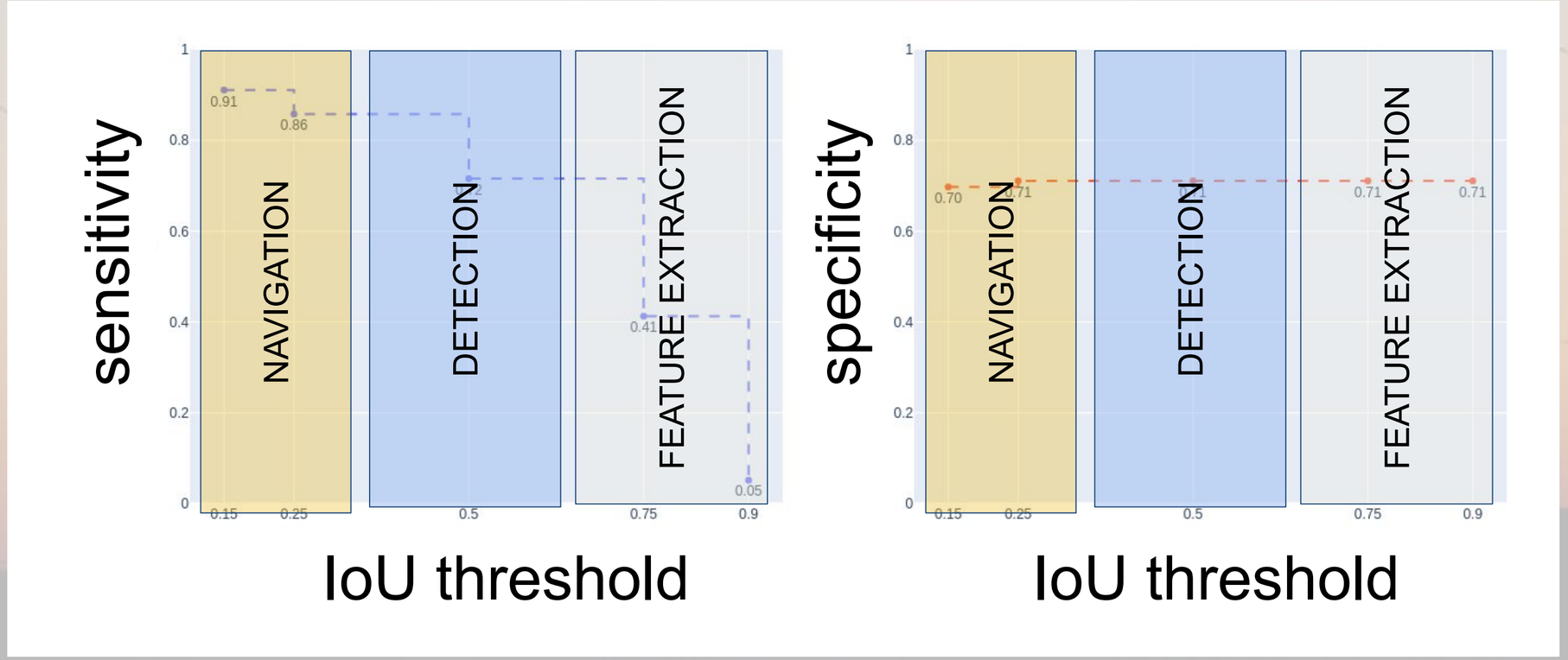
Module performance: segmentation

Classification x IoU (overlap)



Module performance: segmentation

Classification x IoU (overlap)



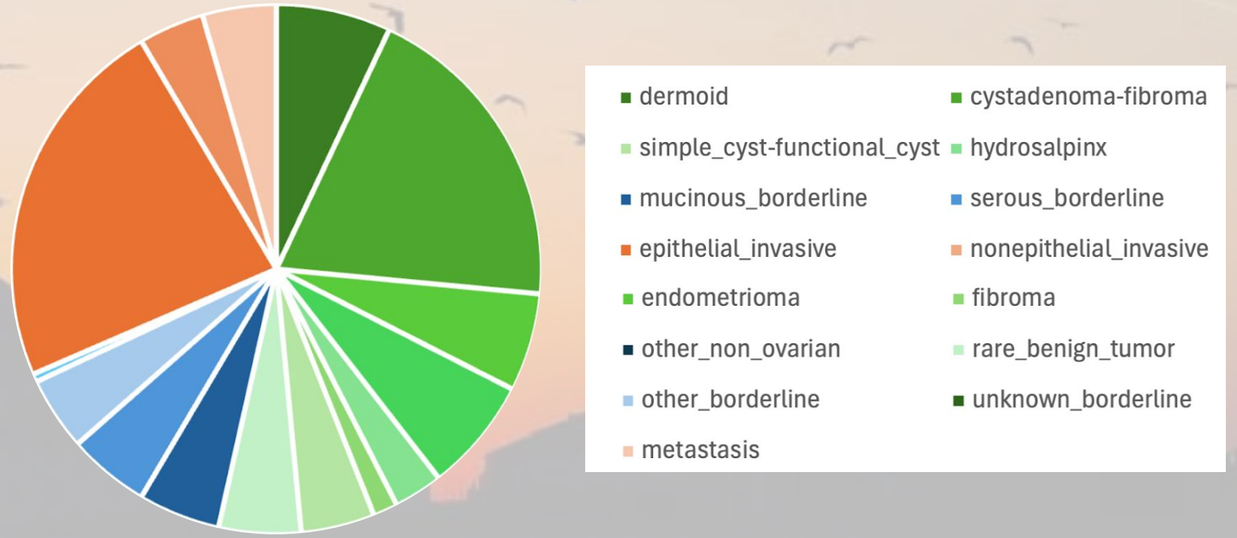
Used dataset:

~800 clinical cases from 7 hospital partners

55% benign, 33% malignant, 12% BOT cases

~5 videos, ~6 images per case

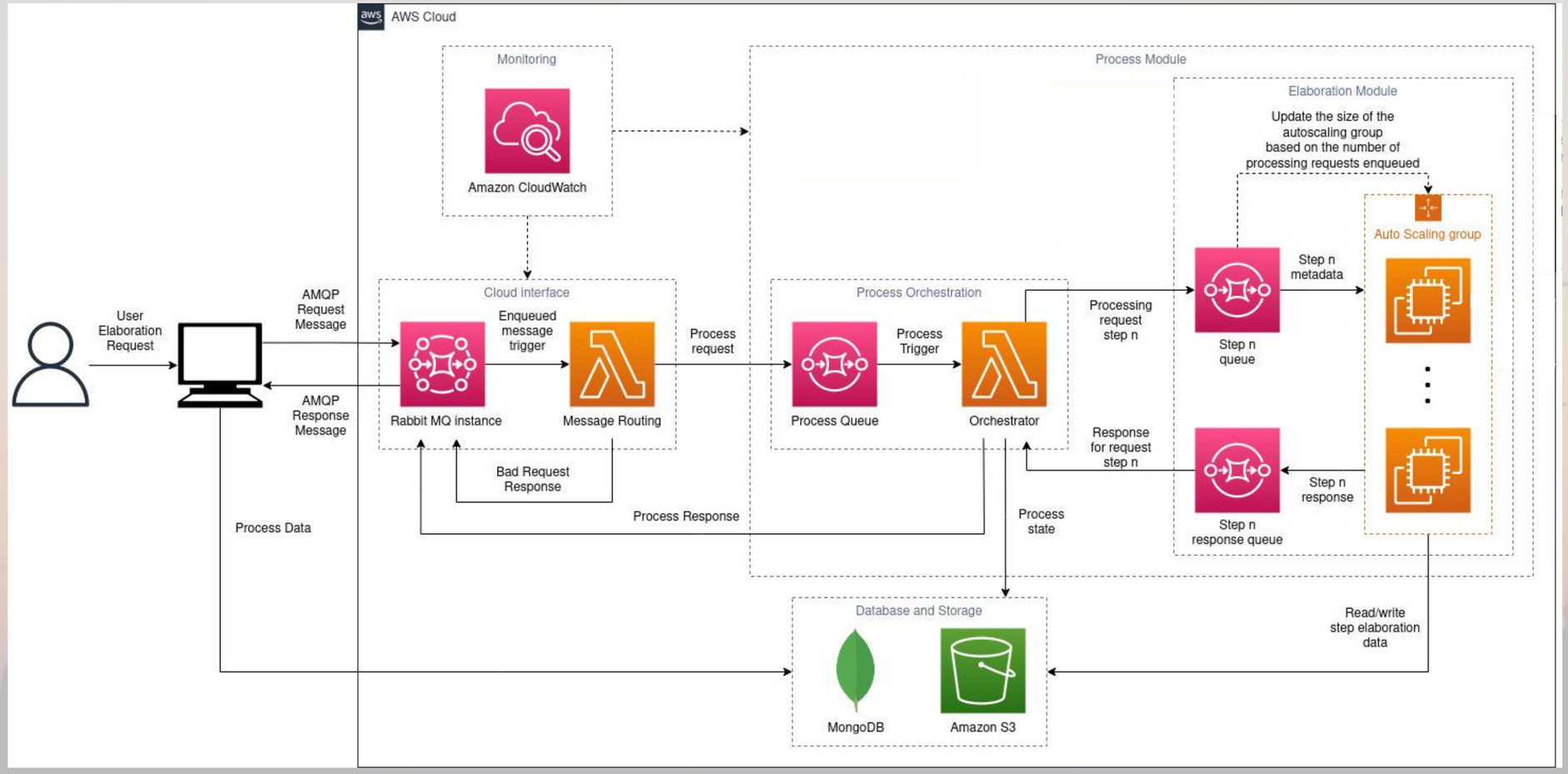
15 histotypes



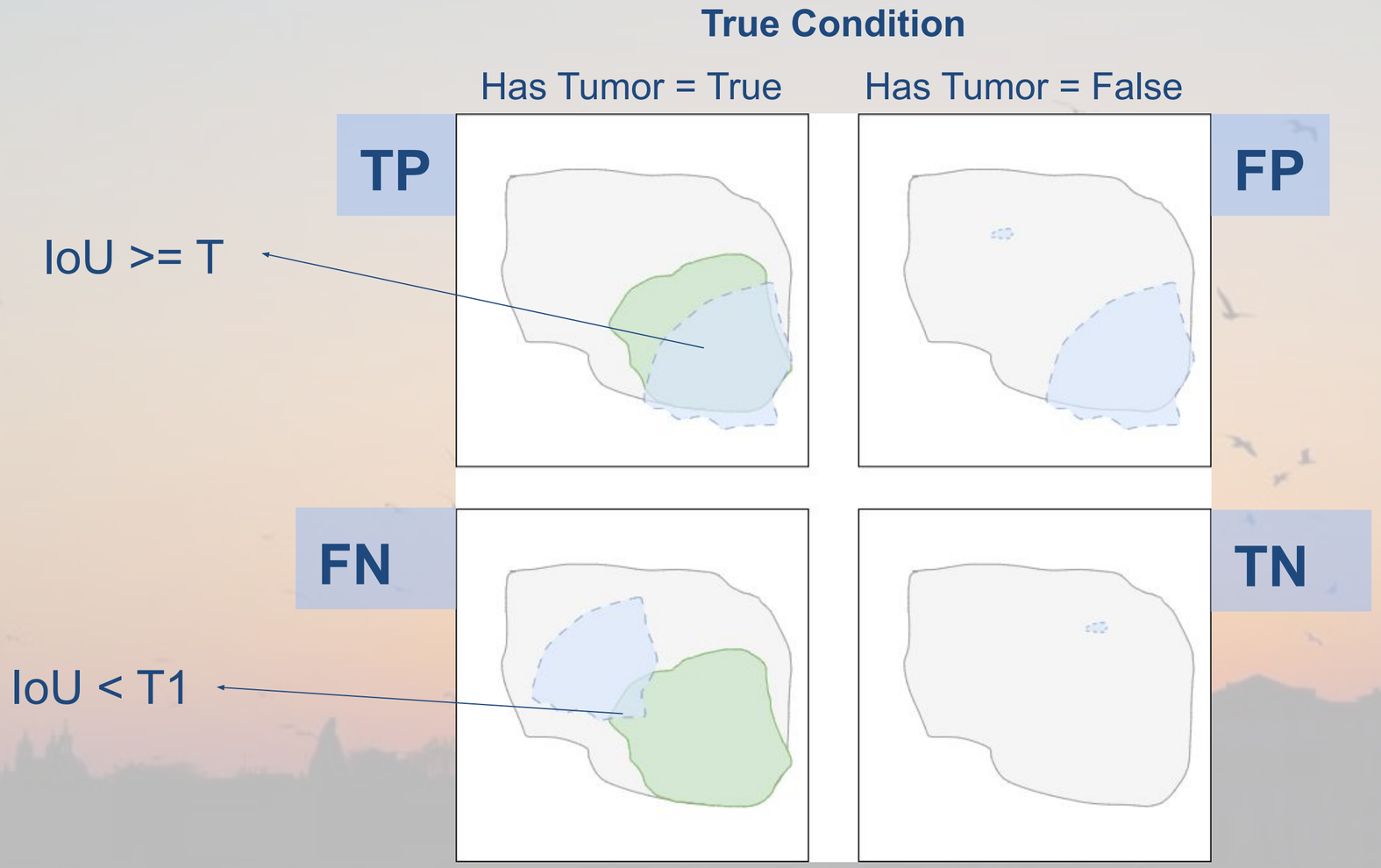
Solution development

- Python/ TensorFlow, PyTorch, Scikit-learn
- Algorithms: UNet, CNN, random forest, …
- Cross validation on AWS batch
- Gitlab, Weight&Biases

Product architecture



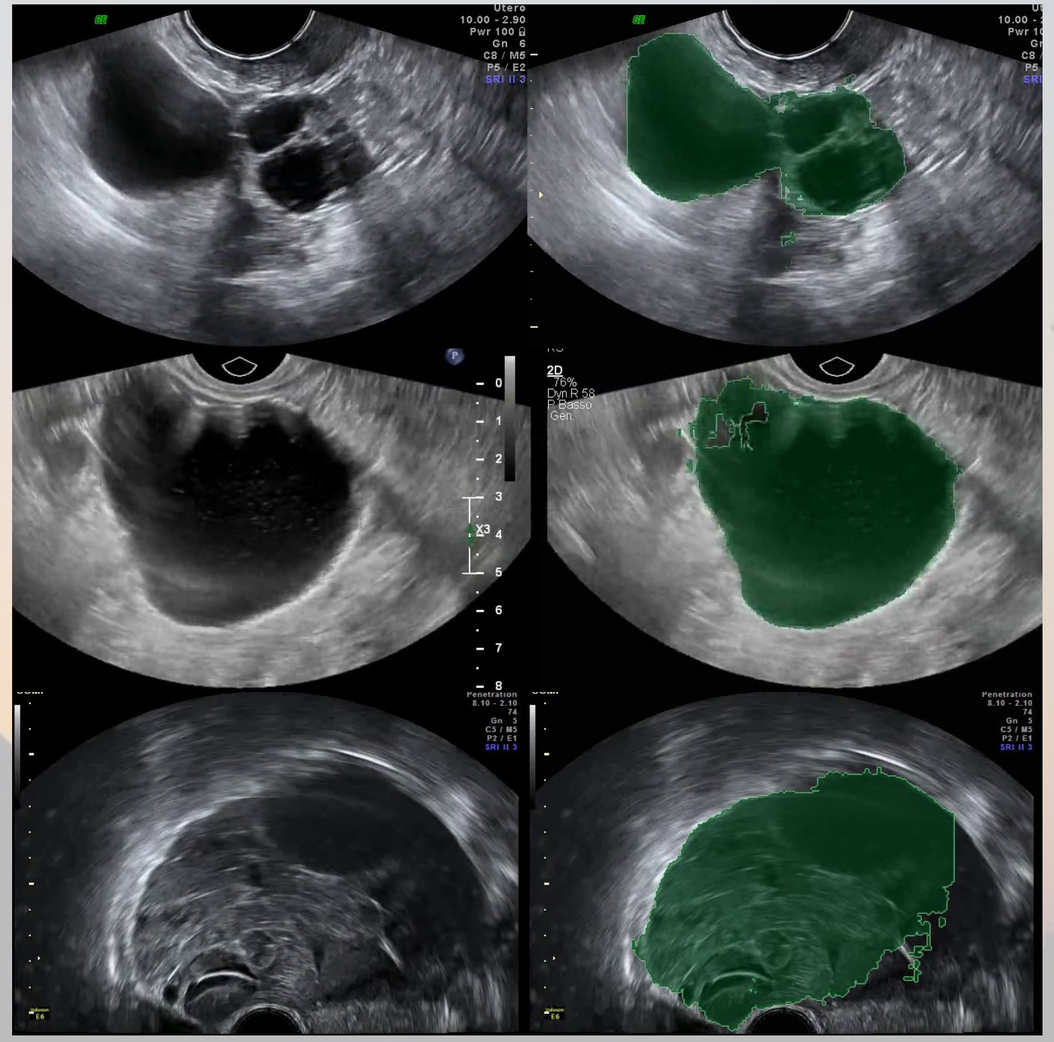
Modules performance



Modules performance

ROI detection

Segmentation of lesion
DICE: 85%
IoU: 77%
precision 85%, recall: 86%



Modules performance

Tumor identification

Accuracy: 77%
Sensitivity: 88%
Specificity: 95%

Table 2. Accuracy, sensitivity, specificity, positive and negative LR with regard to malignancy of subjective evaluation of static ultrasound images by observers with varying levels of ultrasound experience

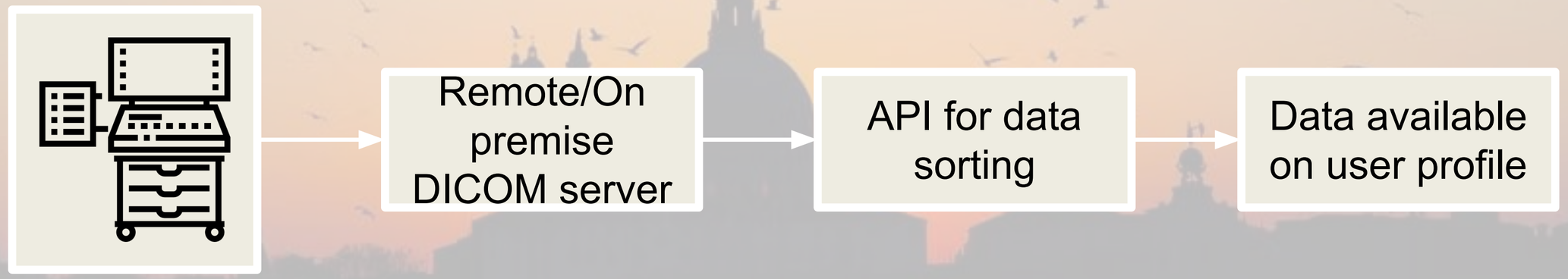
Sonologist	AUC	Accuracy n (%)	95% CI	p	Sensitivity n (%)	95% CI	p	Specificity n (%)	95% CI	p	LR+ (95% CI)	LR– (95% CI)
Experts												
A	0.92247	89 (147/166)	83–93		86 (60/70)	76–92		91 (87/96)	83–95		9.14 (5.03–17.25)	0.16 (0.09–0.27)
B	0.86109	82 (136/166)	75–87		86 (60/70)	76–92		79 (76/96)	70–86		4.11 (2.81–6.23)	0.18 (0.10–0.31)
C	0.88199	83 (138/166)	77–88		80 (56/70)	69–88		85 (82/96)	77–91		5.49 (3.41–9.12)	0.23 (0.14–0.36)
Consensus opinion		85 (141/166)	79–90		83 (58/70)	72–90		86 (83/96)	78–92		6.12 (3.74–10.36)	0.20 (0.12–0.32)
Senior trainees												
D	0.84189	80 (133/166)	73–85	0.1441	84 (59/70)	74–91	0.7630	77 (74/96)	68–84	0.0389	3.68 (2.56–5.45)	0.20 (0.12–0.34)
E	0.85506	81 (134/166)	74–86	0.1779	70 (49/70)	58–79	0.0201	89 (85/96)	81–93	0.5637	6.11 (3.52–10.95)	0.34 (0.23–0.47)
Junior trainees												
F	0.78586	78 (129/166)	71–83	0.0455	70 (49/70)	58–79	0.0290	83 (80/96)	75–89	0.4913	4.20 (2.67–6.81)	0.36 (0.24–0.51)
G	0.72560	72 (120/166)	65–79	0.0014	74 (52/70)	63–83	0.1336	71 (68/96)	61–79	0.0039	2.55 (1.83–3.62)	0.36 (0.23–0.54)
H	0.72664	70 (117/166)	63–77	0.0004	86 (60/70)	76–92	0.6171	59 (57/96)	49–69	<0.0001	2.11 (1.65–2.77)	0.24 (0.12–0.47)
I	0.79464	75 (125/166)	68–81	0.0114	73 (51/70)	61–82	0.0896	77 (74/96)	68–84	0.0606	3.18 (2.18–4.77)	0.35 (0.23–0.51)

p value refers to the comparison with the consensus opinion. Consensus opinion is defined as the diagnosis suggested by at least 2 of the 3 experts.

Citation

Deployment in clinical environment

- Access through a webapp
- Data can be dragged and drop or sent from the ultrasound machine



Challenges in AI in Healthcare:

Data quality and availability
Model biases
GDPR and data management
Regulatory requirements
Cost and resource requirements
Integration into existing systems (e.g. where will it be deployed)