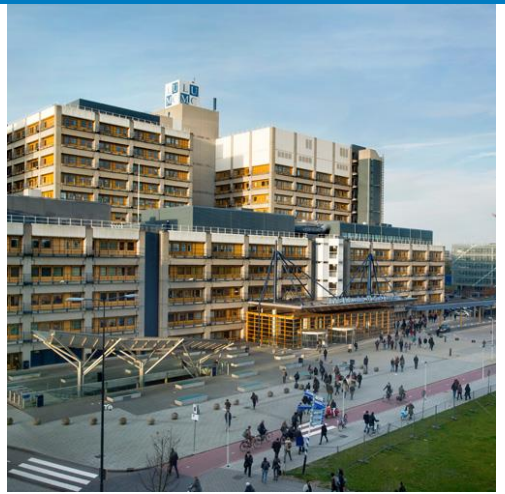
Leiden University
Medical Center

AI-based Imaging Biomarkers for Drug Evaluation Trials

Berend Stoel

Division of Image Processing, Department of Radiology

LEIDEN UNIVERSITY MEDICAL CENTER



COI: Bristol-Myers Squibb and Pfizer contributed to the presented research project on rheumatoid arthritis. BS is consultant for ClinFact, Clario, Medpace, CSL Behring, Talecris, Kamada, BeamTx and Takeda.

Imaging Biomarkers in Drug Evaluation Trials

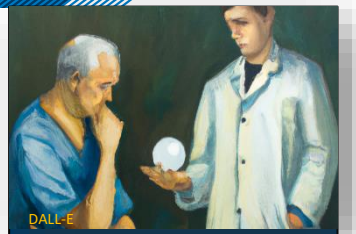
Imaging biomarker

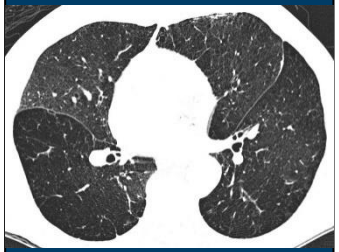
- Measurable defined characteristic
- Indicator of processes or responses to an intervention
- Numerical or categorical outcome measure
- **Based on in-vivo medical imaging**

The role of AI in producing imaging biomarkers

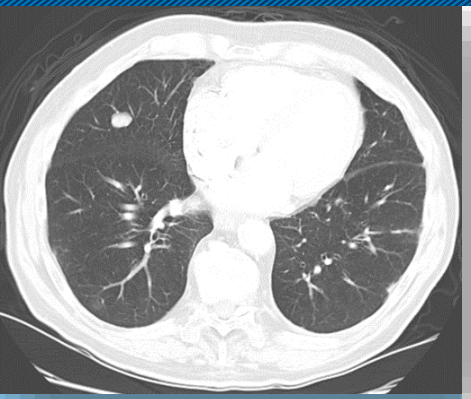
Examples in drug evaluation trials

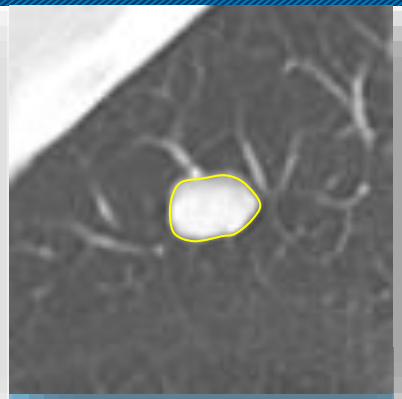
- Lung emphysema secondary to α_1-antitrypsin deficiency
- Early treatment of patients with arthralgia suspected of developing rheumatoid arthritis

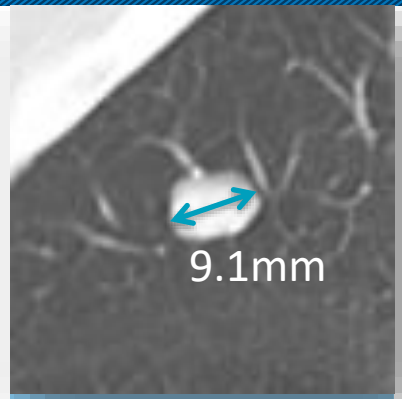




Different Routes to Produce an Imaging Biomarker
Classical versus Deep Learning







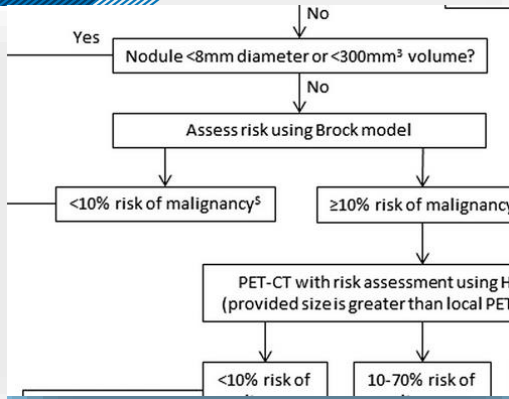


Image acquisition

Segmentation/ Detection

Handcrafted Feature(s)

Learnable Feature(s)

Statistical Analysis/ Classification

Deep Learning

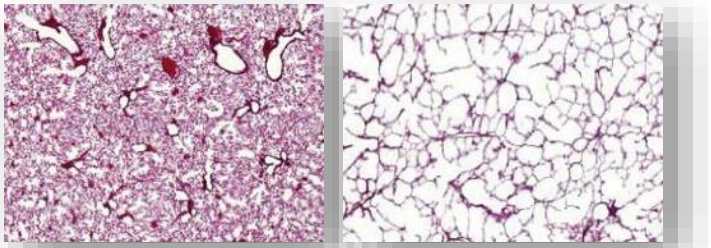
Regression / Prediction

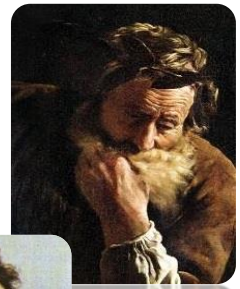
Stoel, BC et al. Nat Rev Rheumatol 20, 182–195, 2024

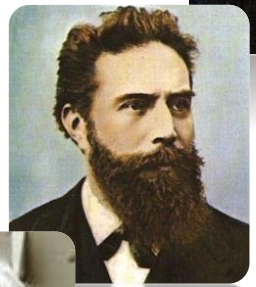
Classical Route: Lung CT Densitometry

Pulmonary emphysema is defined in pathological terms:

Permanently enlarged distal airspaces and destruction of their walls[1]

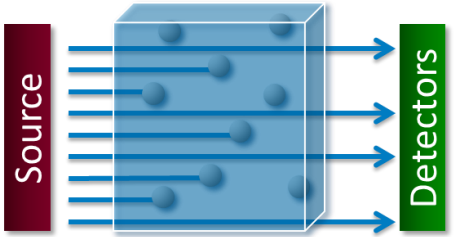






Direct way of measuring emphysema: quantification of tissue density

X-ray absorption ≡ physical density

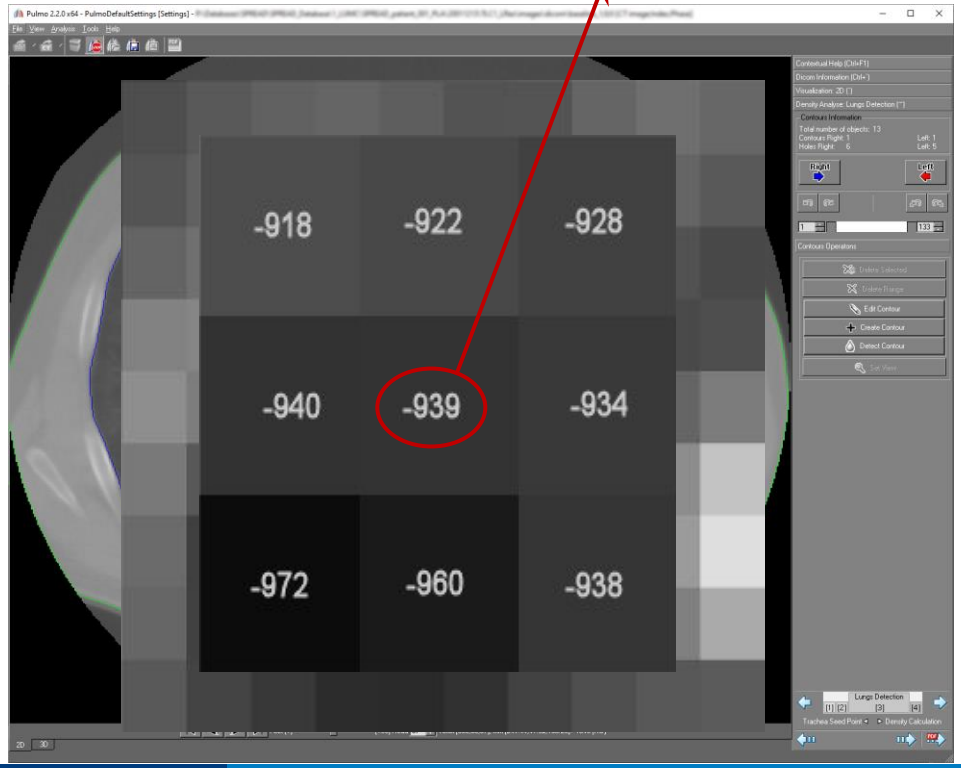


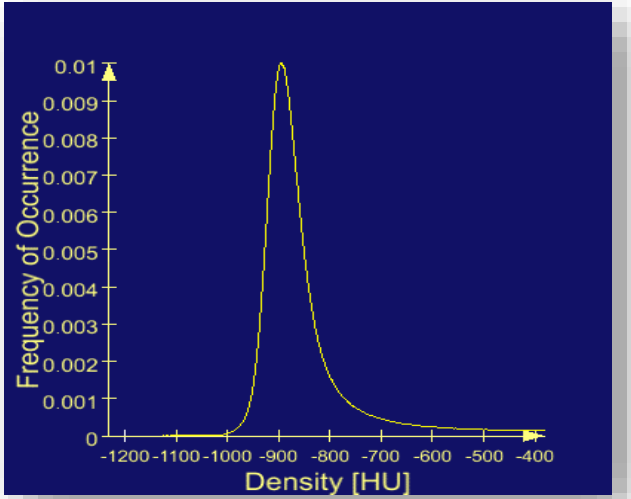
Only in the lungs: 3D imaging data from Computed Tomography

[1] Am Rev Respir Dis. 1985;132(1):182-5

Lung CT = Densitometry

Hounsfield Unit (HU) + 1000 ≈ density in gram/liter



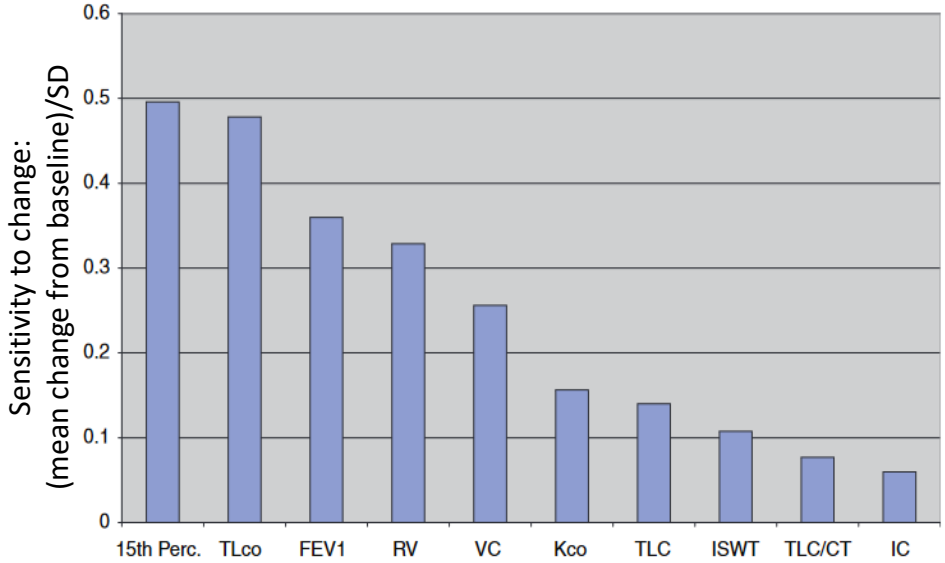


From image processing we obtain:

- Lung Volume (V)
- 15th percentile density

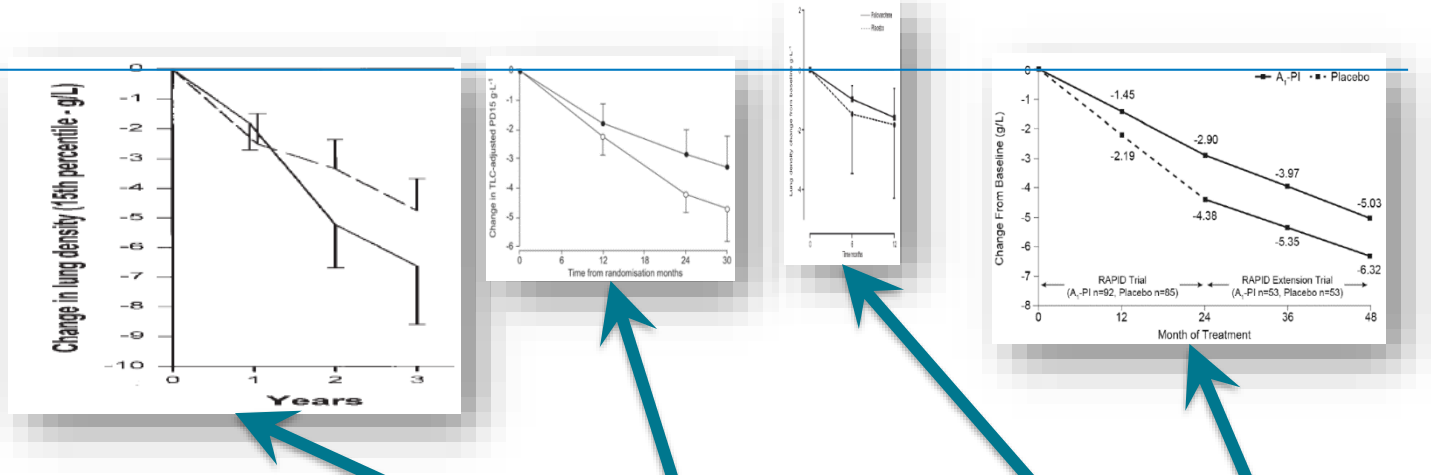
Classical Route: Lung CT Densitometry to Quantify Emphysema

Image acquisition → Segmentation/ Detection → Handcrafted Feature(s) → Statistical Analysis / Classification



J. Stolk et al., Respir. Med. 101, 1924-1930, 2007

Lung Densitometry in Drug Evaluation Trials on Emphysema Secondary to α1-Antitrypsin Deficiency



Annual Tissue Loss (g/L/year)	DD Study[1] 1999 (n=27/27)	EXACTLE[2] 2009 (n=39/38)	Combined[3] 2010 (n=59/60)	REPAIR[4] 2012 (n=133/129)	RAPID[5] 2015 (n=87/84)
Placebo	-2.6	-2.2	-2.7	-1.8	-2.2
Active	-1.5	-1.4	-1.7	-1.6*	-1.5

* RARgamma

1. Dirksen A et al. AJRCCM, 160:1468-1472 1999
2. Dirksen A et al. Eur Respir 33: 1345-1353, 2009
3. Stockley RA et al. Resp. Res 11:136, 2010
4. Stolk J et al. Eur Respir 40: 306-312, 2012
5. Chapman KR et al. the Lancet, 386:360-8, 2015

Putting Things into Perspective

Natural progression of AATD emphysema

- Each liter of lung tissue looses 2 gram (HU) per year

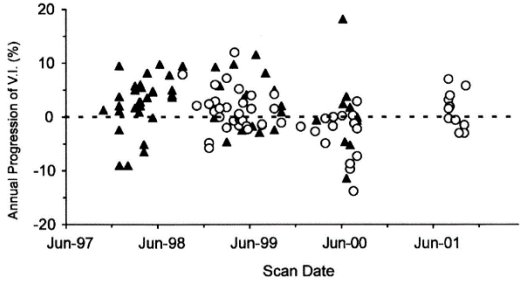
Treatment effect of AAT augmentation therapy

- Of each liter of lung tissue 1 gram (HU) is preserved each year

Tolerance levels for CT number accuracy in clinical practice[1]:

- ± 5 HU (water)
- ± 10 HU (other material)

Parr DG et al. Am J Respir Crit Care Med. 2004



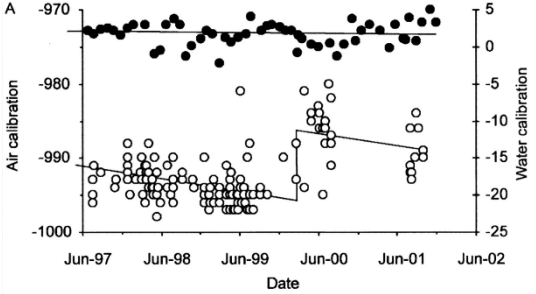
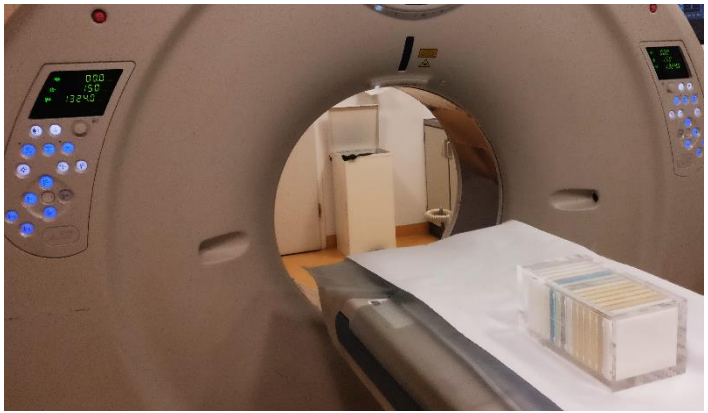


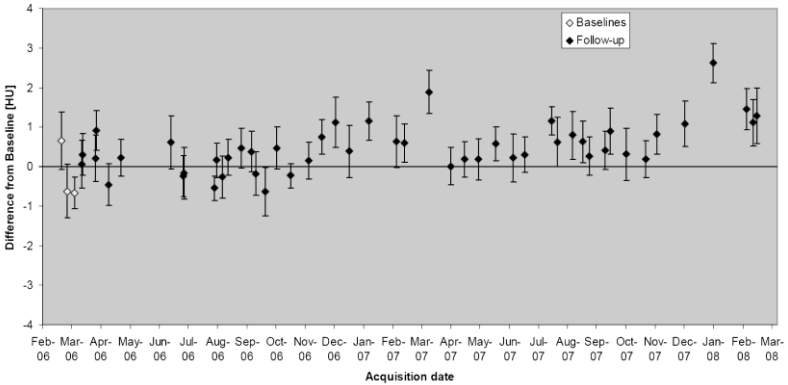
Image acquisition → Segmentation/ Detection → Handcrafted Feature(s) → Classification/ Statistical Analysis

[1] International Atomic Energy Agency. Quality Assurance Programme for Computed Tomography: diagnostic and therapy applications. Vienna: IAEA; Human Health Series No 19; 2012

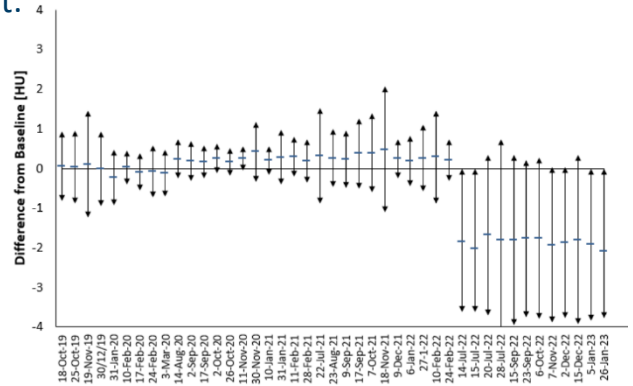
Quality Control of CT Scanners, More Strictly than Clinical Tolerance Levels



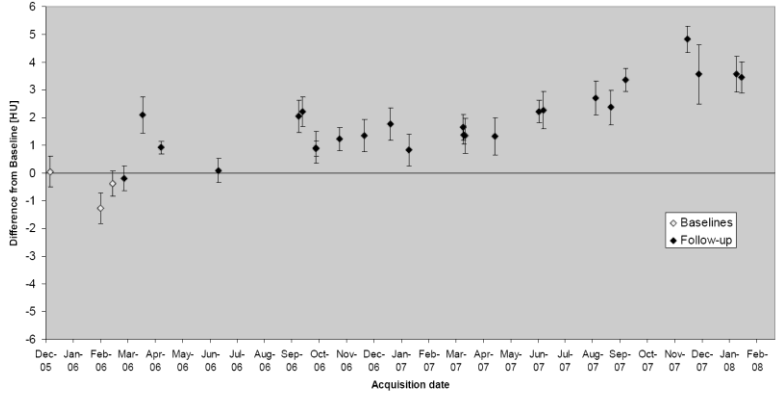
Stable:



Shift:



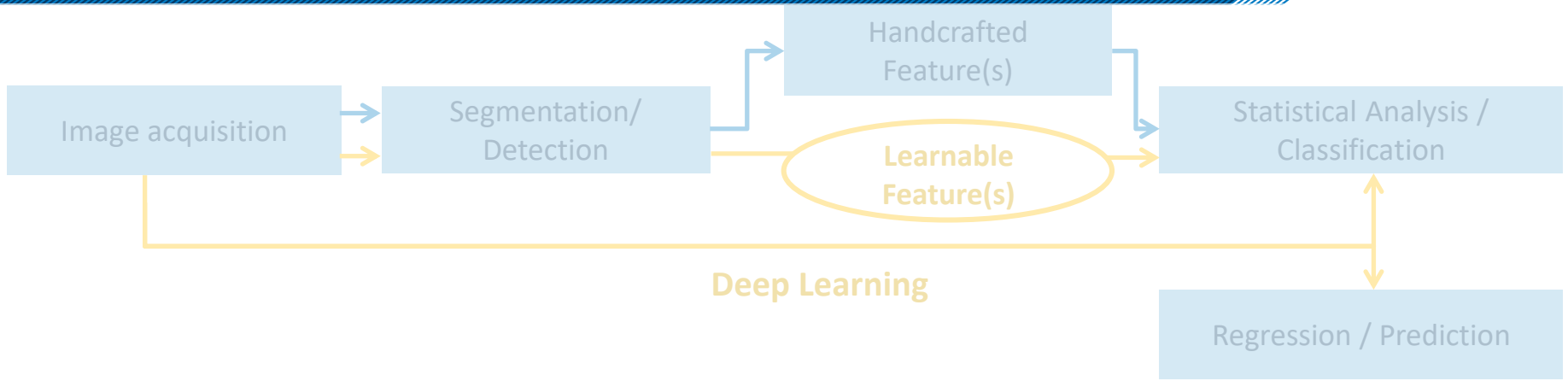
Drift:



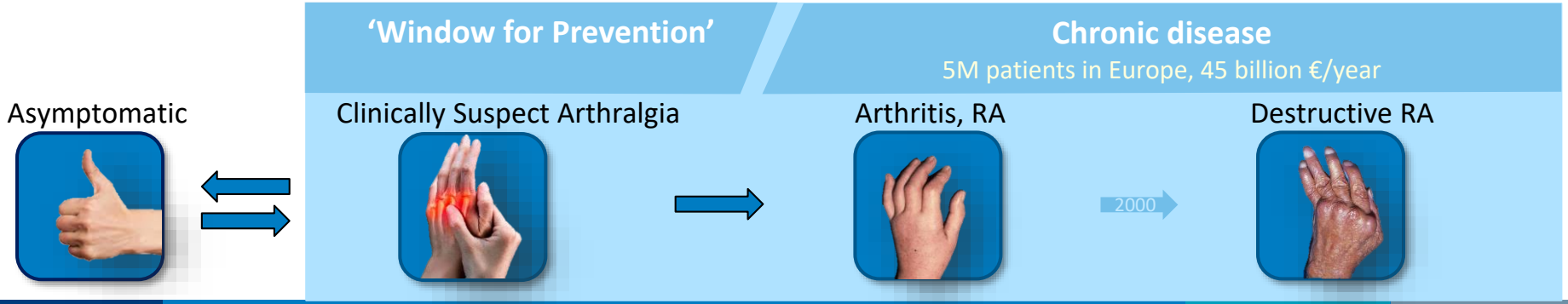
To correct this, internal recalibration during image analysis

Stoel BC, et al. Proc Am Thorac Soc. 15;5(9):929-33 2008

Different Routes to Produce an Imaging Biomarker
Classical versus *Deep Learning*



Rheumatoid arthritis (RA) is the most common autoimmune diseases



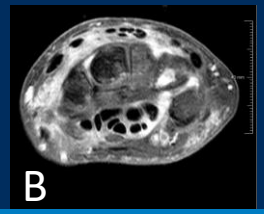
TREAT EARLIER Trial

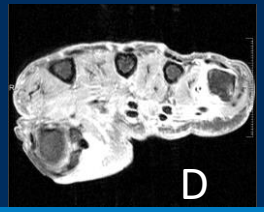
Early, preventive treatment in Clinically Suspect Arthralgia (CSA) patients:

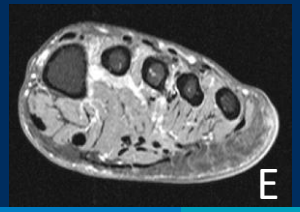
- Placebo (n=117)
- Single intramuscular glucocorticoid-injection and a one-year course of oral methotrexate (n=119)

Location		High resolution plane (T1 fat-suppressed + contrast)	# scans
Wrist	A	COR	805
	B	TRA	805
MCP	C	COR	805
	D	TRA	805
Foot (MTP)	E	COR	792
	F	TRA	787

- Four time points (BL, 4, 12 and 24 months)
- Visually scored by the RAMRIS system, by clinical experts







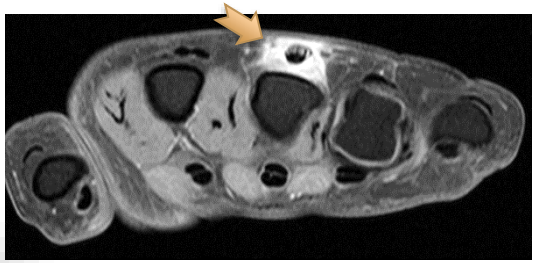
Inflammatory Lesions in Rheumatoid Arthritis (RA)

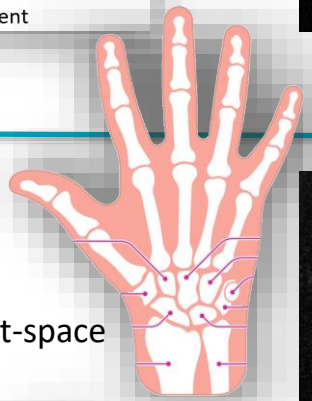
According to the RAMRIS system:

- Tenosynovitis

  Inflammation of tendon sheaths

0	No tenosynovitis
1	<1.5mm tenosynovial contrast enhancement
2	≥1.5mm but <3mm tenosynovial contrast enhancement
3	≥3mm tenosynovial contrast enhancement





- Synovitis

  Inflammation of synovial joint-space

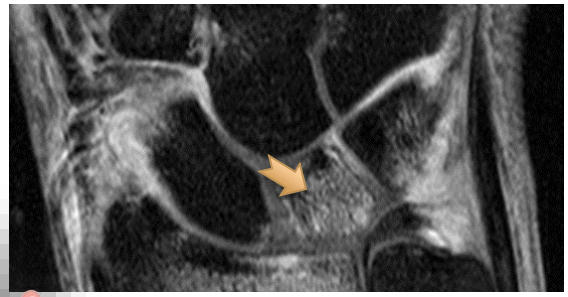
0	normal	
1	mild	(up to ⅓ of presumed maximum area)
2	moderate	(up to ⅔ of presumed maximum area)
3	severe	(>⅔ of presumed maximum)

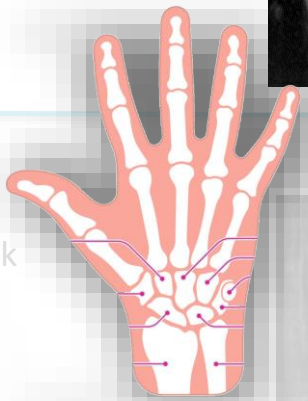
Inflammatory Lesions in RA

- Bone marrow edema (BME)

  A.k.a. osteitis (bone inflammation)

0	no BME
1	1 – 33% of bone with osteitis
2	34 – 66% of bone with osteitis
3	67 – 100% of bone with osteitis

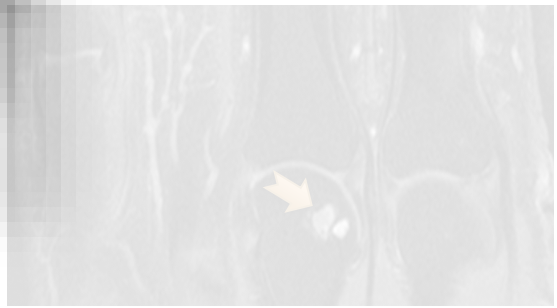




- Erosions

  Bone loss, with cortical break

0	no erosion
1	1 – 10% of bone volume eroded
2	11 – 20% of bone volume eroded
3	21 – 30% of bone volume eroded
…	…
10	91 – 100% of bone volume eroded



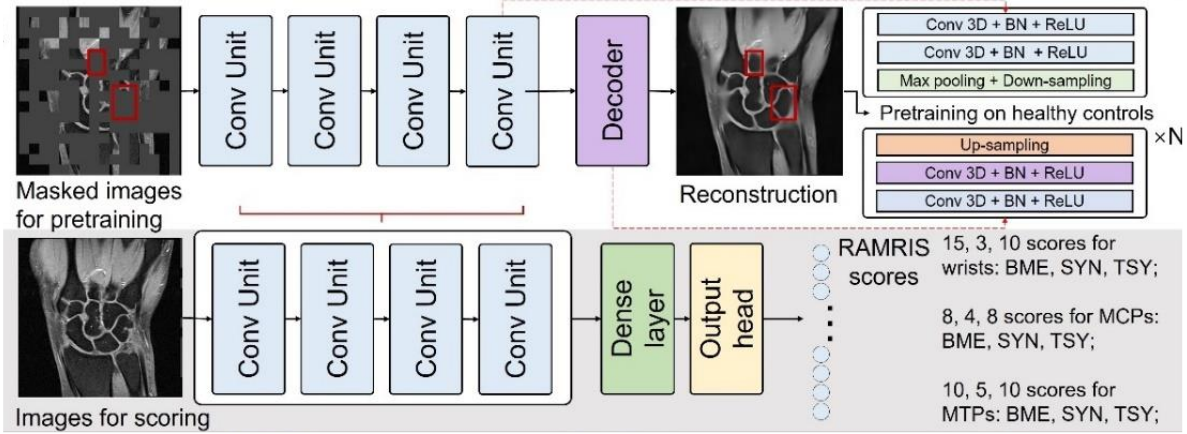
Can AI Help?

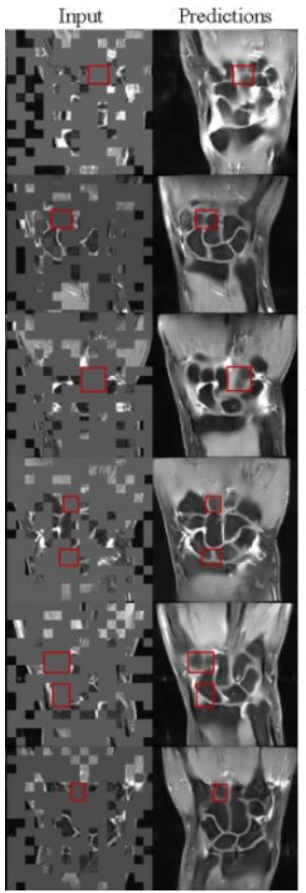
A. In assessing severity of arthritis?

- **By simulation visual scoring (RAMRIS)**

B.

C.





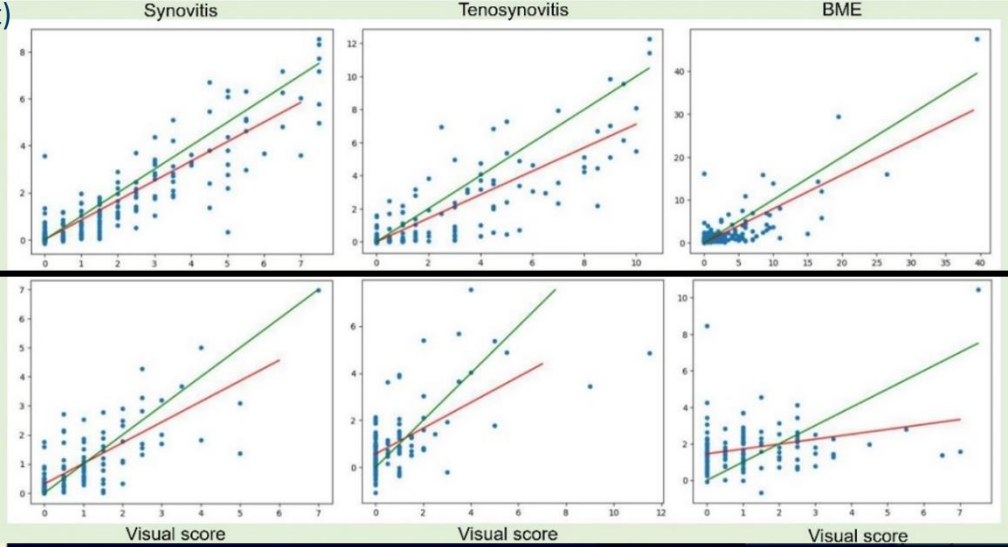
- Training set: 727 clinically suspect arthralgia (CSA); 1247 early onset arthritis; and 174 healthy controls
- Validation set: 127 CSA patients from the TREAT-EARLIER trial

Simulating Visual Scoring – Results

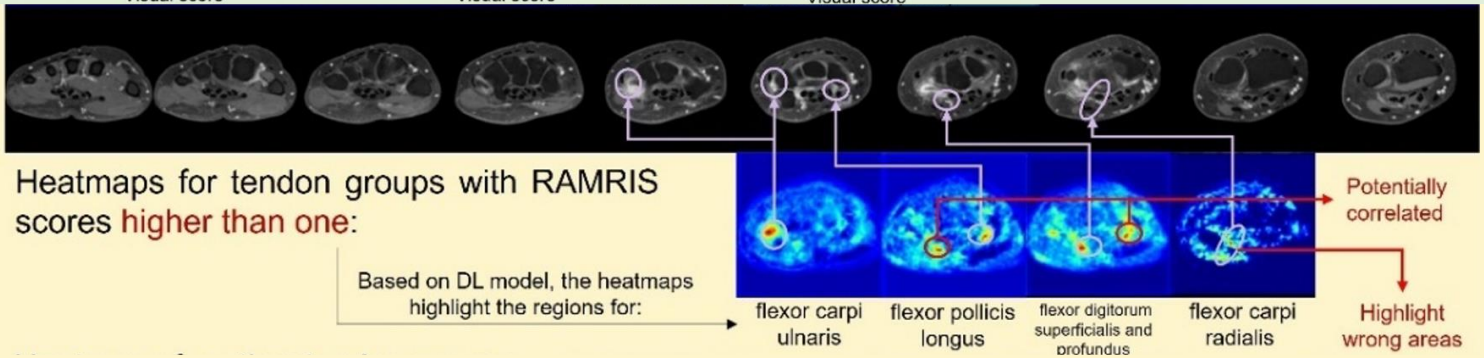
(wrist)

Training

Validation



Score (from wrist, MCP and MTP)	Corr. (SD)	
Total inflammation	**0.88** (±0.03)	Training
(rough reference, 2 observers)	**0.86**	
Synovitis	**0.86** (±0.05)	
Tenosynovitis	**0.83** (±0.06)	
BME	**0.65** (±0.11)	
Total inflammation	**0.73** (±0.03)	Validation
(rough reference, 2 observers)	**0.83**	
Synovitis	**0.74** (±0.02)	
Tenosynovitis	**0.61** (±0.06)	
BME	*0.12* (±0.04)	



Li, Y et al., Ann Rheum Dis, POS1376, 2024

Can AI Help?

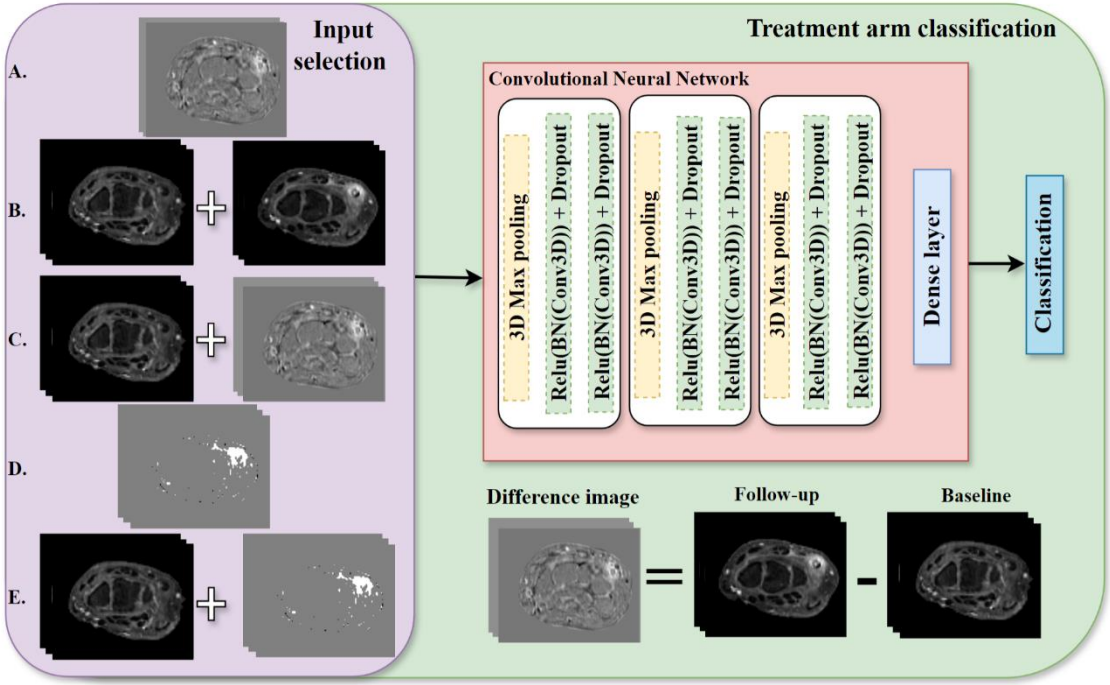
A. In assessing severity of arthritis?
   - By simulation visual scoring (RAMRIS)

B. Localizing treatment effects?

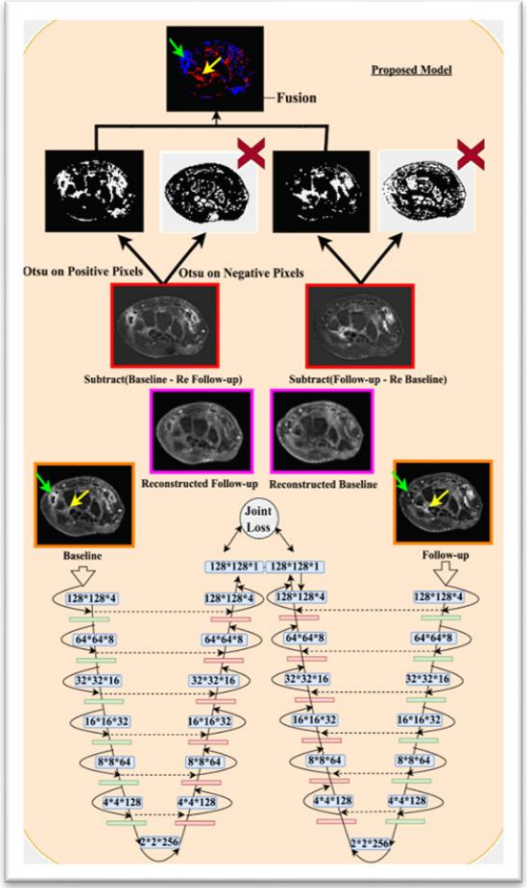
C. Predicting treatment response?

- By classifying MRI scans into treatment arm: active or placebo
- The learned features then contain the treatment effects, to be displayed in saliency maps

10-fold cross-validation
AUC: 0.80

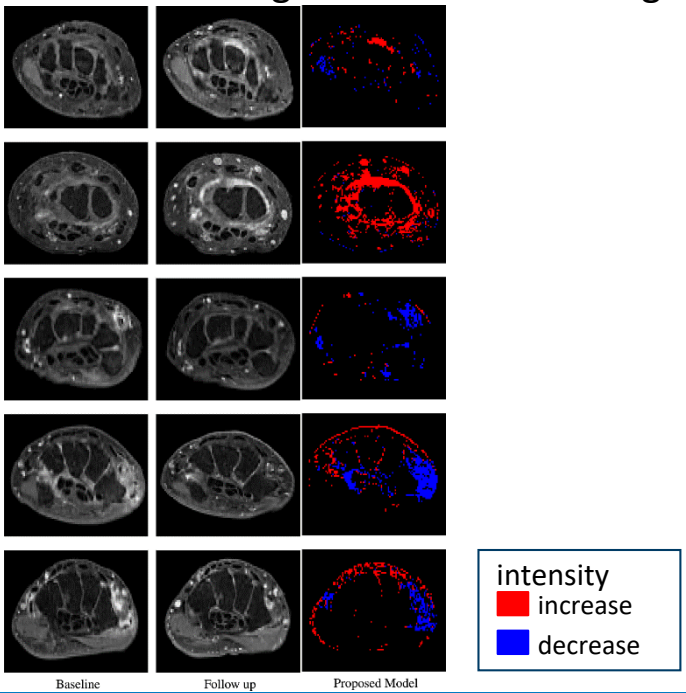


Localizing Possible Treatment Effects: Change Maps



Assumptions:

- Predictable changes are coming from the (non-calibrated) MRI scanner
- Unpredictable changes are relevant changes

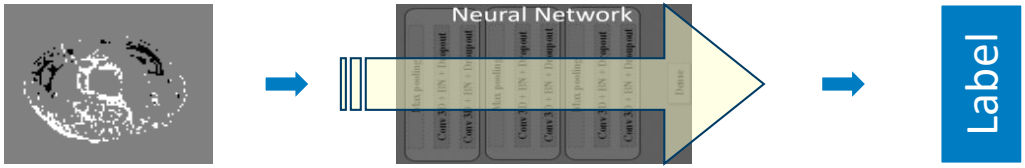


HassanZadeh, T et al. Biomedical Signal Processing and Control 88, 2024

Classification, Explanation Methods

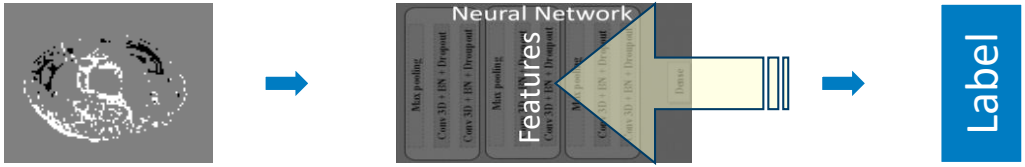
Focused on change maps of the wrist (since segmentations are available)

1. By Input Blocking
   - At a **voxel** level: by square masks at each position in the image

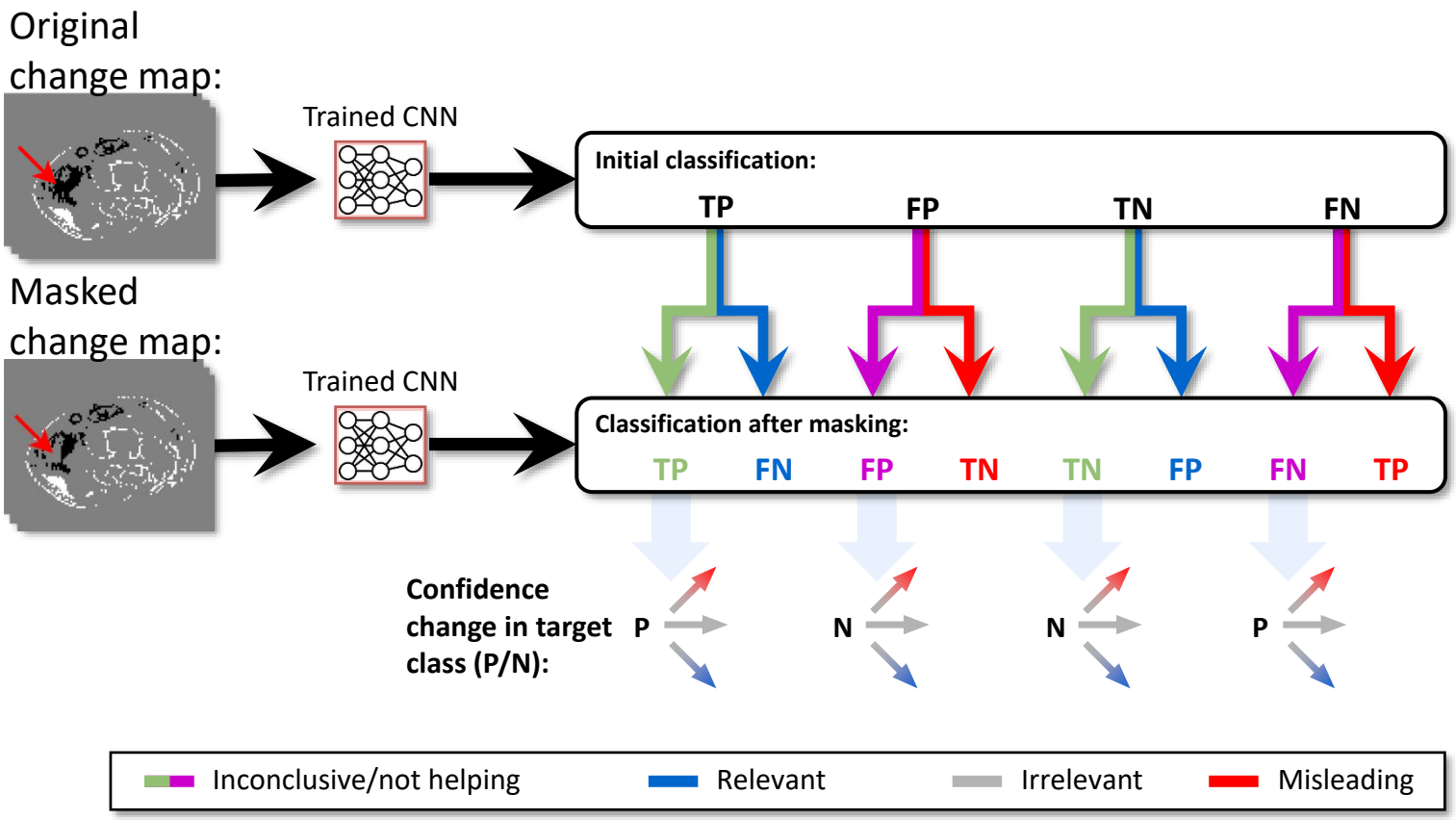


   - At **anatomical** level (tenosynovium, synovium, bones, tendons, vessels, skin, remaining tissue)

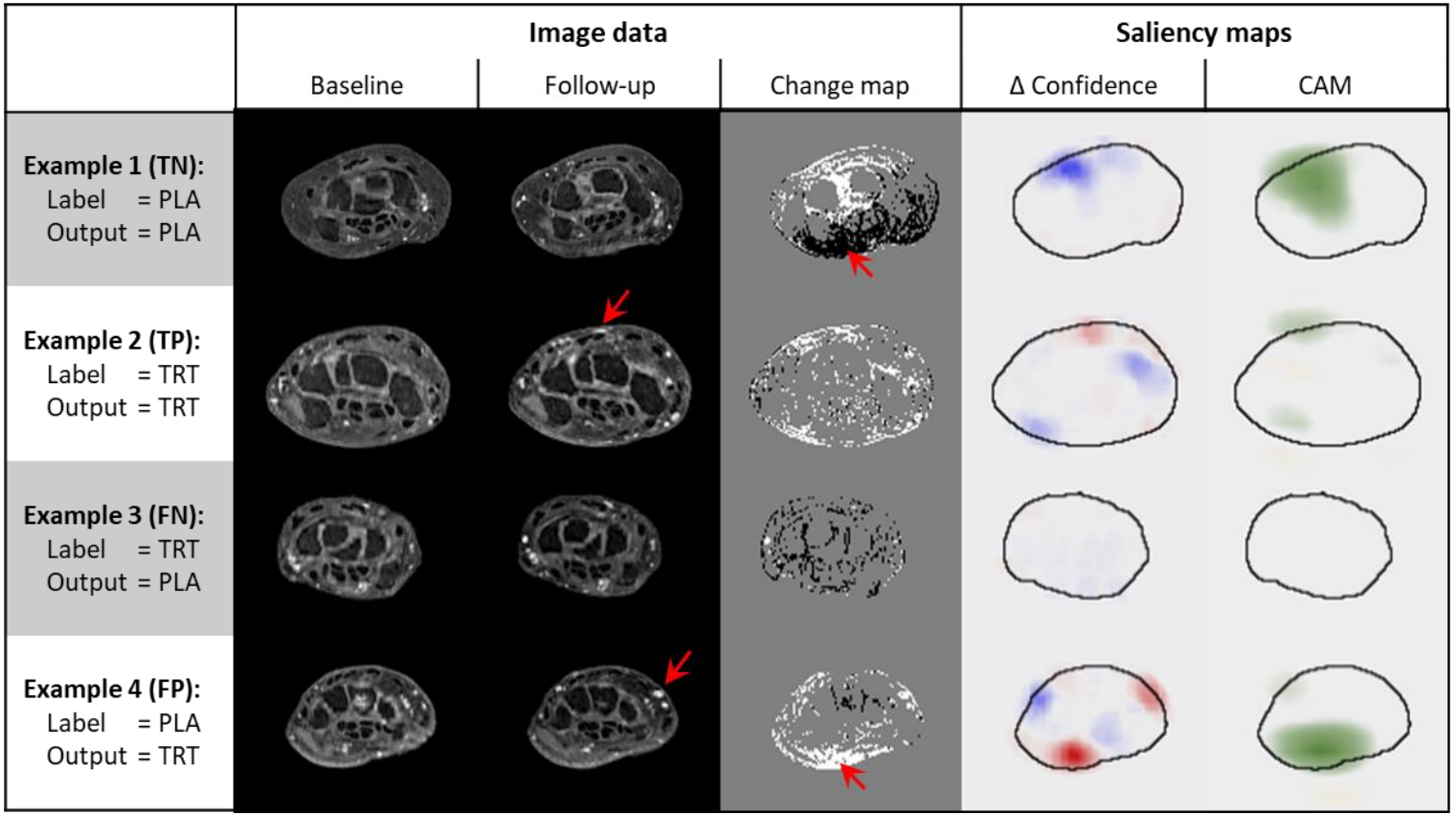
2. By Class Activation Maps (CAMs)



Interpretation of Blocking Effects



Results: Voxel-based Blocking and CAMs



Misleading
Irrelevant
Relevant

Anatomical Level Masking

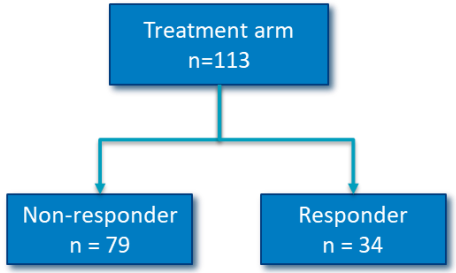
(%)	Synovium	Tenosynovium	Bone	Skin	Vessel	Tendon
Irrelevant	74.05	63.24	77.84	72.97	75.14	80.00
Relevant	6.49	**17.30**	2.70	7.57	5.41	0.54
Misleading	1.08	1.08	0.00	2.16	0.54	0.00
Not helping	18.38	18.38	19.46	17.30	18.92	19.46

Can AI Help?

A. In assessing severity of arthritis?
   - By simulation visual scoring (RAMRIS)

B. Localizing treatment effects?

C. Predicting treatment response?
   - Facilitate patient selection
   - Prevent overtreatment

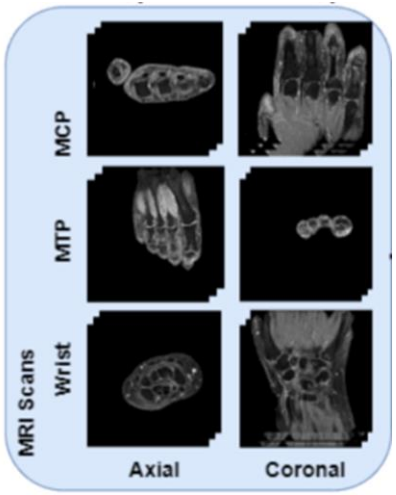


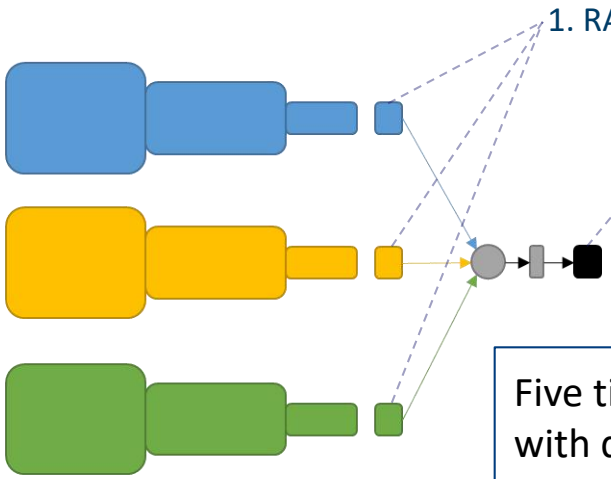
Definition of treatment response:

Decrease in visual score of inflammation > 3 points (RAMRIS)

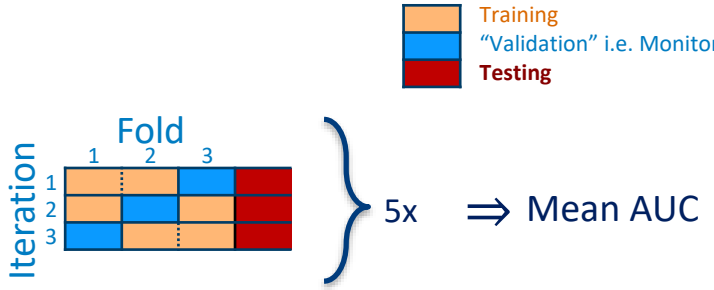
≡ smallest detectable change in visual score, by two observers

Treatment Response Prediction: Two-step Approach





Five times 3-fold cross-validation
with different splits, with hold-out test set



[a] Li, Y et al., Ann Rheum Dis, POS1376, 2024

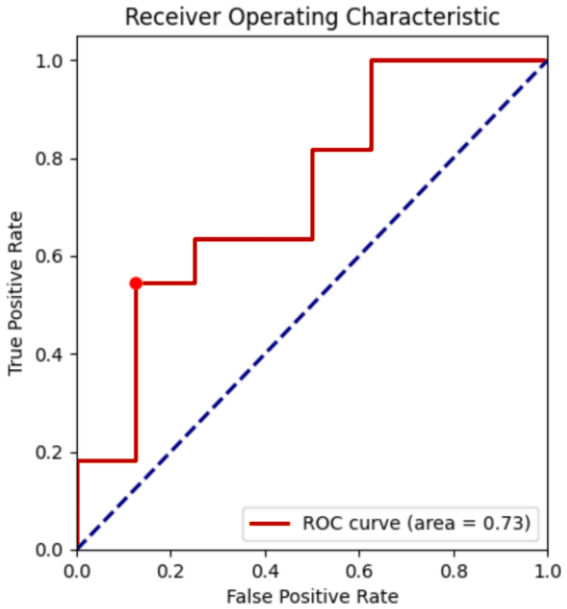
Treatment Response Prediction: Results

AUC: 72% ± 7.83

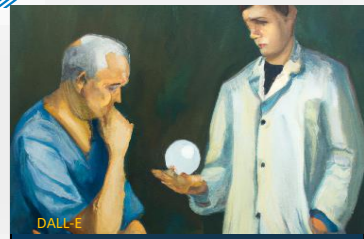
Test characteristics (n=27):

- Positive predictive value: 65 %
- Negative predictive value: 79 %
- Sensitivity: 53 %
- Specificity: 86 %

5th Experiment:



Conclusions on the Role of AI



AI can help in

- Assessing severity of arthritis
- Localizing treatment effects
- Predicting treatment response, to facilitate patient selection

Classical imaging biomarker	Deep learning based biomarker
Easily explainable	Explainable AI: work in progress
Manual correction possible	Work in progress
Quality control is relatively simple	Complex
Imaging biomarkers are interpretable	Interpretation through saliency maps
Restricted to chosen features	Possibility to discover new features
Based on prior knowledge	Hypothesis generation
Both need stringent quality control of image acquisition	

Leiden University
Medical Center

$e^{i\pi}+1$

Division of Image Processing:

- Yanli Li, MSc
- Tahereh HassanZadeh, PhD
- Jingnan Jia, PhD
- Denis Shamonin, MSc
- Els Bakker, PhD
- Marius Staring, PhD

Rheumatology:

- Annette van der Helm-van Mil, MD, PhD
- Dennis Ton, MD

Pulmonology:

- Jan Stolk, MD, PhD

Radiology:

- Monique Reijnierse, MD, PhD
- Lucia Kroft, MD, PhD