Imaging AI/Radiomics decision support improves physicians' stratification of indeterminate nodules: An MRMC study

Anil Vachani[1], Pierre Massion[2], Reginald Munden[3], Fergus Gleeson[4], Christina Bellinger[5], Travis Dotson[5], Lutz Freitag[6], Noah Waterfield Price[6], Quentin Chometon[6], Vaclav Potesil[6], Timor Kadir[6]

1 University of Pennsylvania, Philadelphia, Pennsylvania
2 Vanderbilt University Medical Center, Nashville, Tennessee
3 Department of Radiology, Wake Forest Baptist Health, Winston Salem, North Carolina
4 Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
5 Department of Pulmonary/Critical Care, Wake Forest School of Medicine, Winston-Salem, North Carolina
6 Optellum Ltd., Oxford, United Kingdom

Lung Cancer Prediction (LCP) digital biomarker

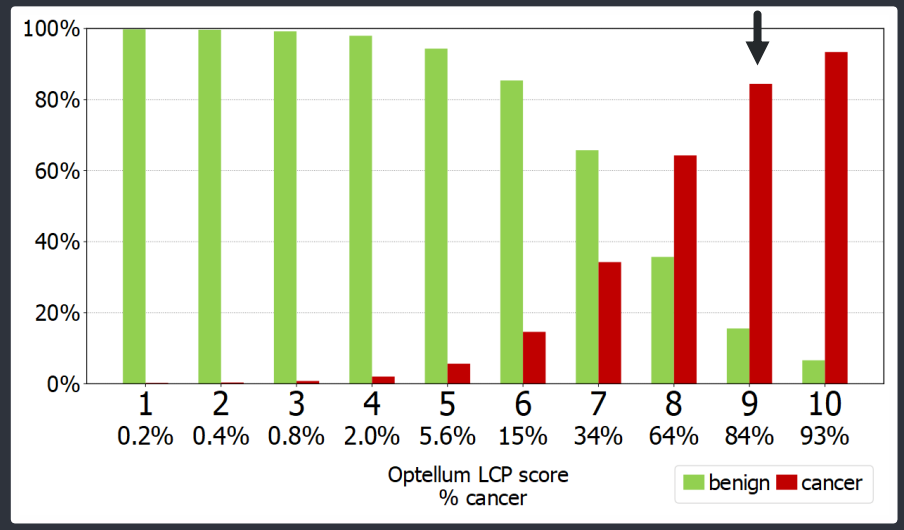
- Imaging AI/Radiomics-based biomarker – computed from CT images only.
- Software package developed by Optellum Ltd (Oxford, UK).
- Trained on more than 70,000 patient scans with known nodule diagnosis (cancer vs. benign).
- Model produces a score between 1 (benign) and 10 (malignant) for a nodule of interest.

Nodule of interest marked in CT image

LCP

AI-based digital biomarker

LCP Score 9

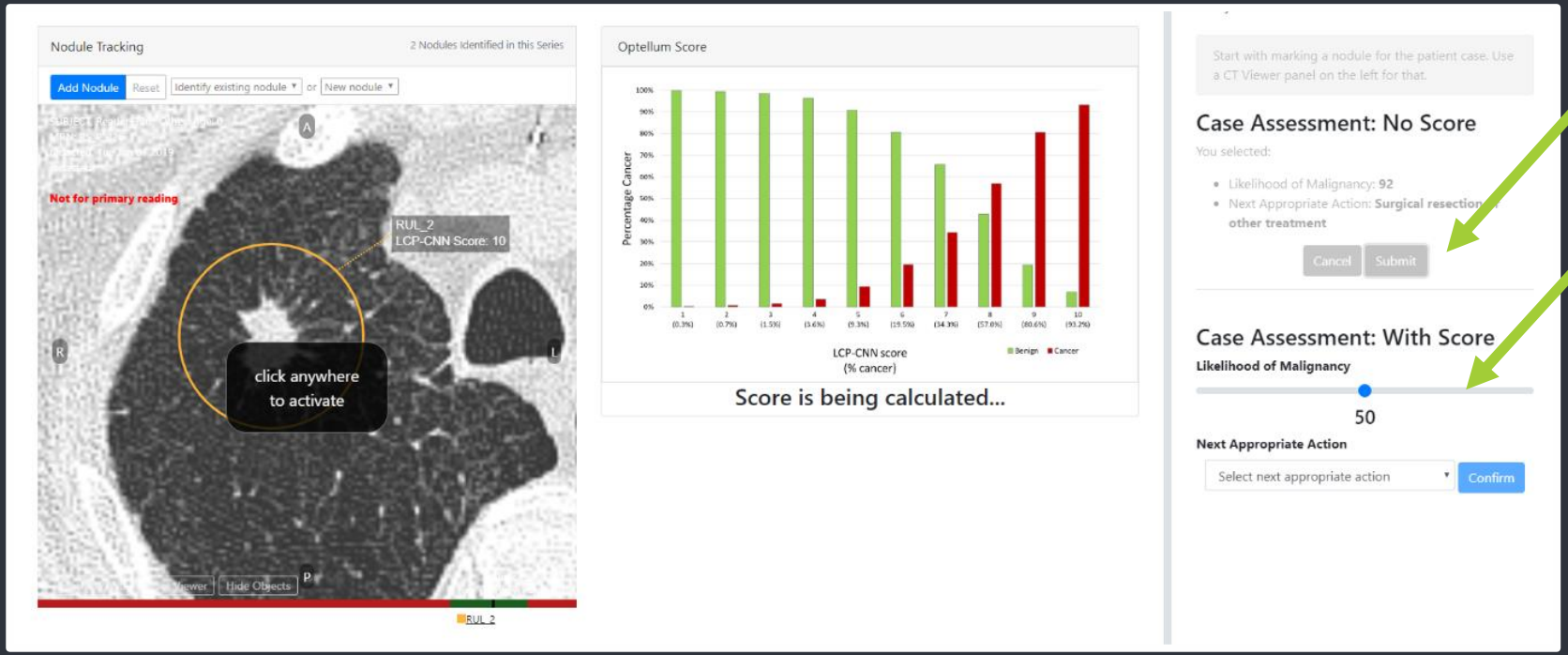


Likely benign

Likely malignant

Study design: Multi-case Multi-reader study

- 300 cases: Solid and semi-solid nodules, 150 cancer and 150 benign, 5-30mm in diameter.
- 12 readers: 6 radiologists and 6 pulmonologists.
- Two sequential reads of each case: blinded and unblinded to the AI score.
- For each case, **likelihood of malignancy** and **clinical recommendation (**blinded and unblinded).



Likelihood of malignancy:
1 to 100

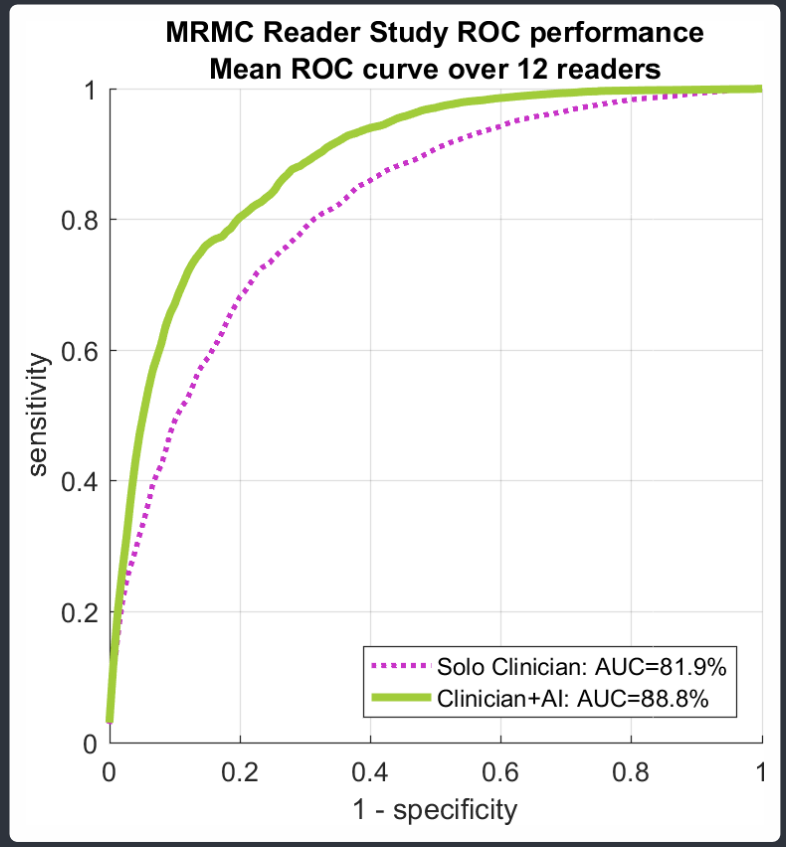
Clinical recommendation:

- No action
- Long term CT follow-up
- Short term CT follow-up
- PET-CT/Biopsy
- Surgical resection/other treatment

Results: Overall

Primary endpoint: Change in AUC

- Mean AUCs averaged over the 12 readers:
  - Pre-AI: 81.9% (95%CI 80.5-83.3%)
  - Post-AI: 88.8% (95%CI 87.7-89.8%)
- Mean effect size: 6.85 AUC points (95%CI 4.29-9.41%, p<.001)



Secondary endpoints:

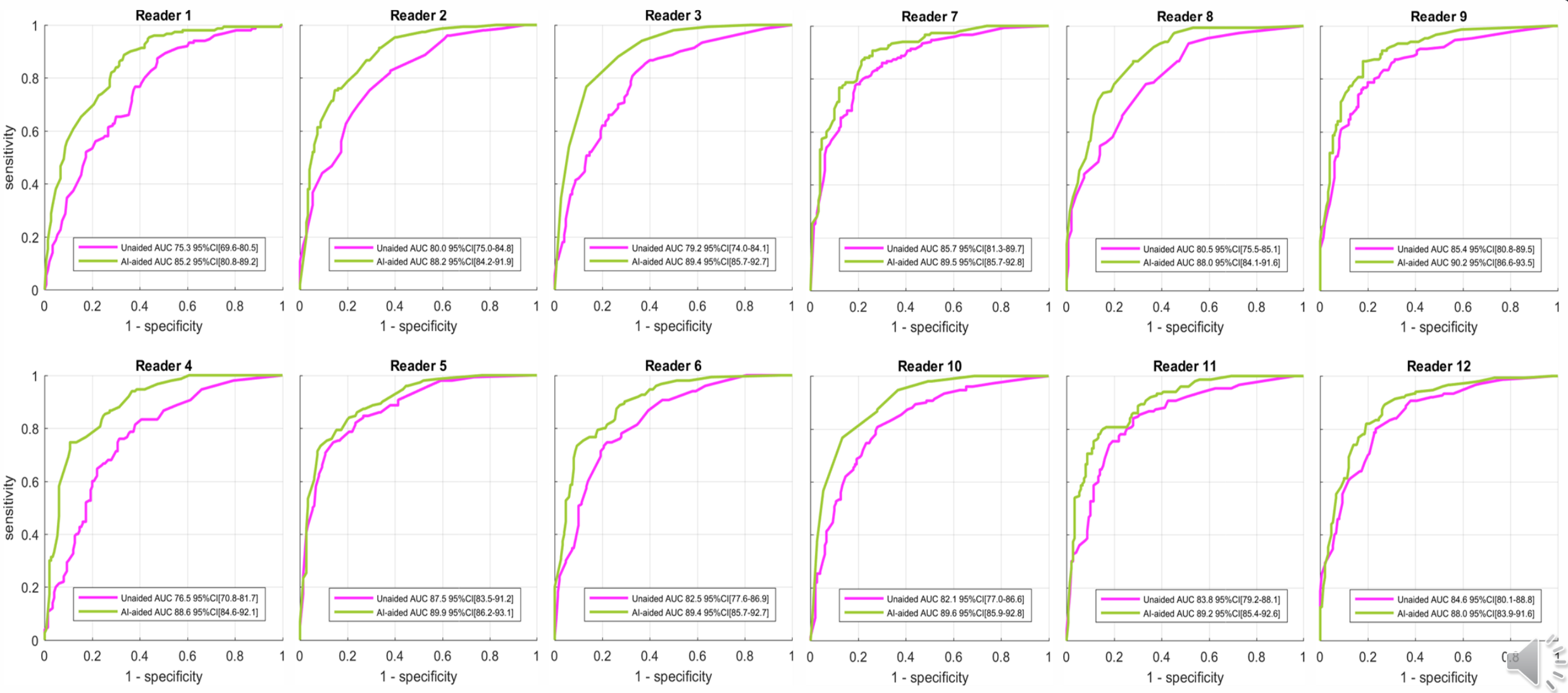
- **Clinical recommendation**: Overall 26% of decisions improved when assisted by the AI.
- **Inter-reader consistency**: Variation in assessment reduced from 16.65 pp to 11.07 pp when assisted by the AI.

LoM Threshold	Sensitivity			Specificity		
	Unassisted	Assisted	P	Unassisted	Assisted	P
5%	94.06	97.89	< .001	37.44	42.28	.001
65%	55.67	65.00	< .001	86.50	89.28	.001

Results: Individual reader ROC performance

- The individual assessments of all readers were improved when assisted by the AI.

Without AI
With AI



Conclusions

- The LCP digital biomarker helped all participant radiologists and pulmonologists improve their performance in categorising IPNs as benign or malignant.
- Overall AUC, sensitivity, and specificity were significantly improved by using the LCP score.
- The LCP digital biomarker improved the clinical decision-making of all readers.
- The LCP digital biomarker improved readers' consistency.