Substantial radiologist workload reduction can be achieved if artificial intelligence is used as a first-read filter at baseline in lung cancer screening

FPN 180P

H. L. Lancaster[1,2,3], B. Jiang[2,3], M. Silva[4], J.W. Gratema[5], D. Han[3], J.K. Field[6], G.H. De Bock[1,2], M.A. Heuvelmans[1,2,3], M. Oudkerk[2,3]

1. University Medical Center Groningen, Groningen, NL; 2. University of Groningen, Groningen, NL; 3. Institute for Diagnostic Accuracy, Groningen, NL; 4. Scienze Radiologiche, Department of Medicine and Surgery (DiMeC), University of Parma, Parma, Italy; 5. Department of Radiology, Gelre Hospital, Apeldoorn, NL; 6. University of Liverpool, Liverpool, UK

Objective

- To externally validate a **task-focused AI** lung cancer screening software, when used as a first-read filter, in a real-world baseline lung cancer screening dataset.

Conclusion and Implications

- AI negative predictive performance is better than **all** manual readers.
- If used as a first-read filter, radiologists would **only need to assess 35%** of cases with indeterminate-positive nodules.

Future research

- Comparison based on **gold standard histological results.**
- **Validation** at follow-up and as concurrent reader in the 4ITLR study.

Plain language summary

Why did we perform this research?

Artificial intelligence (AI) could help reduce the number of CT scans a radiologist needs to evaluate if it can be validated as an accurate first reader to rule out cases where no lung nodules or small lung nodules are present

How did we perform this research?

We compared the performance of AI and all manual human readers (when detecting, segmenting, and classifying lung nodules) to each other and an expert panel. Disagreements with the expert panel were called misclassifications

What were the results of this research?

AI performed better than all manual human readers, only missing 5% of lung nodules equal to or bigger than 100mm^3 compared to 16-19% by manual human readers.

What do these results mean for lung cancer screening?

If this AI was used as a first-reader in a comparable real-world lung cancer screening setting to rule out negative cases, radiologists would only need to evaluate 35% of cases.

What's next?

Comparison to gold standard lung cancer results in this UKLS dataset

Mobile-friendly poster

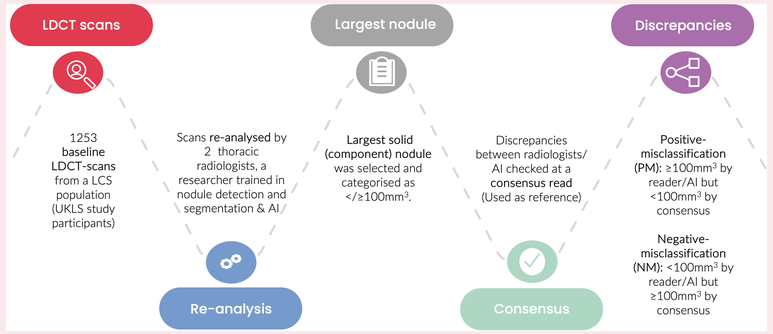
Please scan this QR code to access a copy of this poster on your mobile device.

Introduction

- Lung screening trials and programmes are being implemented to enable early detection and subsequently reduce lung cancer mortality worldwide.[1]
- There is a global shortage in radiologist workforce, and many suffer from stress and burnout due to an ever-increasing workload. [2]
- Lung cancer screening will inevitably lead to an increase in workload, adding to the existing workforce-workload mismatch.
- Task-focused artificial intelligence (AI) could help reduce the workload if used as a first-read filter to rule-out negative cases at baseline. [3]
- External validation of AI lung cancer screening software is desperately needed using high quality real-world datasets. [3]
- We present results from a real-world external validation study on AI lung nodule detection, segmentation, and classification in baseline lung cancer screening.

Methods

- LDCT scans from 1253 UKLS trial participants were used in this validation dataset. [4]
- All scans were read independently by 3 manual readers and the AI software. Results from the UKLS trial consensus read were also used as an additional independent reader.
- Per case, the nodule with the largest solid component was analysed.



- Nodules were classified as < or ≥ 100mm^3, the upper threshold for benign nodule growth.
- Discrepancies between any reader or AI were analysed by an expert panel forming the consensus reference read. The expert panel consisted of 2 thoracic radiologists with > 10 years of experience, who were blinded from the results of the initial reads.
- Based on the consensus read, results were classified as: correct positives (CP), correct negatives (CN), positive misclassifications (PM) or negative misclassifications (NM).

Results

- At consensus, 815 (65%) cases were deemed as having no nodules or nodules with a solid component <100mm^3.
- Results from individual readers can be seen in Table 1 and are graphically presented in Figure 4.

Table 1. Overview of results for each reader and AI when compared to the consensus reference standard.

	Reader 1		Reader 2		Reader 3		Reader 4		AI	
n=1253	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Correct positive: ≥100mm³	232	19	238	19	202	16	218	17	370	30
Correct negative: <100mm³	768	61	808	64	813	65	802	64	757	60
Positive misclassifications (PM)	48	4	8	1	3	0	14	1	59	5
Negative misclassifications (NM)	205	16	199	16	235	19	219	17	67	5
Total discrepancies	253	20	207	17	238	19	233	19	126	10
	%	**95%CI**	**%**	**95%CI**	**%**	**95%CI**	**%**	**95%CI**	**%**	**95%CI**
Sensitivity	65.9	61.3-70.3	54.5	49.7-59.2	46.2	41.5-51.0	49.9	45.1-54.7	84.7	81.0-88.0
Specificity	91.5	89.4-93.4	99.0	98.1-99.6	99.6	98.9-99.9	98.3	97.1-99.0	92.8	90.8-94.5
Positive predictive value (PPV)	80.7	76.7-84.1	96.8	93.7-98.4	98.5	95.6-99.5	94.0	90.2-96.4	86.3	83.0-88.9
Negative predictive value (NPV)	83.4	81.5-85.1	80.2	78.6-81.8	77.6	76.0-79.0	78.6	77.0-80.0	91.9	90.1-93.4

n: absolute number, positive misclassifications: reader/AI classified nodule as ≥100mm^3 whereas consensus <100mm^3, negative misclassifications: reader/AI classified nodule as <100mm^3 whereas consensus ≥100mm^3, %: percentage of total cases n=1253, 95%CI: 95% confidence interval

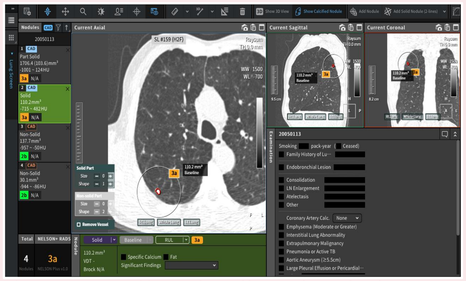


Figure. 1 AI correct positive. Not detected by several manual readers

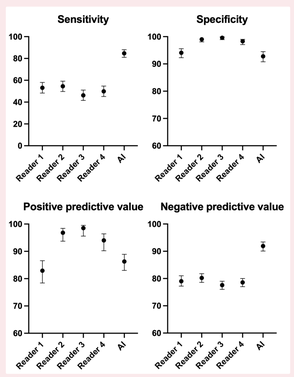


Figure 4.
Overview of sensitivity, specificity, positive predictive value and negative predictive value per reader/AI

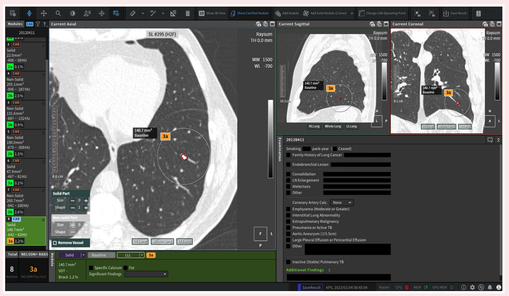


Figure. 2 Positive misclassification by 3 out of 4 manual readers and AI. Predominantly vessel segmented.

Figure. 3 AI negative misclassification. Classified nodule as part-solid with solid component <100mm^3

Limitations

- These results are based on an expert panel read and not gold standard histology outcomes. Subsequent research will compare the results of the manual readers and AI to the true lung cancer diagnoses.
- In this study extra parenchymal lung nodules such as endobronchial were excluded.

References

1. Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JK. Lung cancer LDCT screening and mortality reduction—evidence, pitfalls and future perspectives. Nat Rev Clin Oncol. 2021;18:135–51.
2. RSNA (2022) 'RSNA News', 10 May. Available at: https://www.rsna.org/news/2022/may/global-radiologist-shortage.
3. Lam S, Bai C, Baldwin DR, Chen Y, Connolly C, de Koning H, Heuvelmans MA, Hu P, Kazerooni EA, Lancaster HL, Langs G, McWilliams A, Osarogiagbon RU, Oudkerk M, Peters M, Robbins HA, Sahar L, Smith RA, Triphuridet N, Field J. Current and Future Perspectives on Computed Tomography Screening for Lung Cancer: A Roadmap From 2023 to 2027 From the International Association for the Study of Lung Cancer. J Thorac Oncol. 2024 Jan;19(1):36-51. doi: 10.1016/j.jtho.2023.07.019. Epub 2023 Jul 23.
4. Field JK, Vulkan D, Davies MPA, Baldwin DR, Brain KE, Devaraj A, et al. Lung cancer mortality reduction by LDCT screening: UKLS randomised trial results and international meta-analysis. Lancet Reg Health Eur. 2021 Sep 11;10:100179. doi: 10.1016/j.lanepe.2021.100179.

Acknowledgements

We thank participants of the UK Lung Cancer Screening trial who made this study possible, alongside all those who were involved in the recruitment, scanning, data collection and cleaning, and follow-up of trial participants.

Contact information: Harriet Louise Lancaster, MD-PhD candidate, UMCG, iDNA, The Netherlands; h.l.lancaster@umcg.nl

https://www.linkedin.com/in/harriet-louise-lancaster

Poster presented at Early Lung Cancer Congress (ELCC) 2024. Prague, 20th -23rd March 2024

No conflicts of interest to disclose