Decoding pulmonary nodules: can machine learning enhance malignancy risk stratification?

Colin Jacobs

Randomised controlled trials, with the National Lung Screening Trial and Dutch-Belgian NELSON trial being the two largest, have demonstrated that lung cancer screening of high-risk individuals using low-dose CT reduces lung cancer mortality compared with no screening or screening with chest X-ray. Fuelled by the positive results of these landmark trials, low-dose CT-based lung cancer screening of high-risk individuals is being implemented at national or regional scale in an increasing number of countries worldwide. A comprehensive overview of the current status of implementation of lung cancer screening worldwide can be observed in the interactive lung cancer screening overview maintained by the Lung Cancer Policy Network.[1]

One of the challenges in the radiological interpretation of screening CT scans is the management of screen-detected pulmonary nodules. The prevalence of nodules in screening CT scans of the eligible population is high and ranges from 22% to 74%, depending on the inclusion criteria, CT parameters and minimal size cut-off. The vast majority of these pulmonary nodules are benign. This illustrates the importance of accurate risk estimators and management guidelines for screen-detected nodules; they will be crucial to maintain a good sensitivity for malignant nodules, but at the same time keep the false positive rate as low as possible.

In the recent study by Warkentin *et al*,[2] the authors, part of the INTEGRAL consortium, conducted a large study with data originating from four international lung cancer screening studies. With a total dataset of 16 797 nodules, of which 513 malignant, the authors improved over previous radiomics studies for nodule risk estimation that used smaller datasets or datasets with a reference standard set by radiologist consensus opinion. The authors developed and rigorously tested a machine learning model for malignancy risk estimation of indeterminate pulmonary nodules detected at baseline CT. For validation, 20% of the full dataset from the four international screening studies was set aside, leaving a total of 107 malignant nodules and 3256 benign nodules in the test cohort.

The results of the final machine learning model showed excellent discrimination and calibration performance and were compared with the Brock model, the most well-known risk calculator for pulmonary nodules. The comparison showed that the final machine learning model, which used both imaging and epidemiological parameters, was able to outperform the Brock model on the test set. This again demonstrates that detailed computerised analysis of the imaging data can extract additional imaging information that is not encoded in the imaging-related parameters included in the Brock model.

This is an important scientific observation. When nodule management in lung cancer screening will be guided by accurate nodule risk estimators, these improvements are important to optimise the effectiveness of screening programmes. The International Lung Screening Trial is an important ongoing study that is prospectively investigating whether nodule management based on a nodule risk calculator, the Brock model in this study, can perform better than the Lung-RADS (Lung CT Screening Reporting & Data System designed by American College of Radiology) management guidelines in terms of sensitivity and false positive rates.[3]

Previous studies using deep learning approaches have also demonstrated to be able to perform better or on par with the Brock model.[4 5] It is important to stress that these deep learning models did not encode the epidemiological covariates, nine in total, that were included in the machine learning model in this study. Head-to-head comparisons on representative datasets, as being conducted more and more in the scientific literature,[6–8] will need to demonstrate whether there are significant differences between the various published machine learning and deep learning models.

It will be important to assess in future studies how the presented nodule level results translate to sensitivity and false positive rates at screening participant level. When these analyses are performed, the effect on the sensitivity and false positive rate at participant level can be determined, which are important metrics to monitor the effectiveness of lung cancer screening programmes. The results presented by Warkentin *et al* in table 3 from their publication already provide information in this direction, but the presented sensitivity and specificity metrics are still measured at nodule level.

After the extensive study by Warkentin *et al*, important unexplored ground remains. How can accurate risk estimators like the presented machine learning model be best integrated in nodule management guidelines? What risk thresholds on the predicted malignancy risk can be used in nodule management guidelines? Do models need to be recalibrated when applied to new populations to make sure that risk thresholds still apply? Do different risk thresholds need to be applied for subsolid nodules in nodule management guidelines?

In addition, when organised annual lung cancer screening is in place, most nodules being detected in participants will already have one or multiple prior screening CT scans available. In the evaluation of these nodules, it is sensible that machine learning models include the information from prior CT scans, mirroring how radiologists interpret the CT scans in practice. It is therefore apparent to include information from the prior CT, indicating whether or not a nodule was already visible and with what appearance, in future machine learning models.[9]

The study by Warkentin *et al* is an important next step for optimisation of the management of indeterminate lung nodules in lung cancer screening and an important strength of the study stems from the international nature of the dataset. These collaborative research studies with data sources from different populations and geographical regions will be essential to test the generalisability and robustness of the new-generation risk estimation models.

Correspondence to Dr Colin Jacobs, Medical Imaging, Radboudumc, Nijmegen, The Netherlands; colin.jacobs@radboudumc.nl

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Not applicable.

Provenance and peer review Commissioned; internally peer reviewed.

To cite Jacobs C. *Thorax* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/thorax-2023-221300

Accepted 18 January 2024

► http://dx.doi.org/10.1136/thorax-2023-220226

Thorax 2024;**0**:1–2. doi:10.1136/thorax-2023-221300

ORCID iD

Colin Jacobs http://orcid.org/0000-0003-1180-3805

REFERENCES

1 Available: https://www.lungcancerpolicynetwork.com/interactive-map-of-lung-cancer-screening [Accessed 12 Jan 2024].
2 Warkentin MT, Al-Sawaihey H, Lam S, *et al*. Radiomics analysis to predict pulmonary nodule malignancy using machine learning approaches. *Thorax* 2024.
3 Lim KP, Marshall H, Tammemägi M, *et al*. ILST (International Lung Screening Trial). *Investigator Consortium Protocol and Rationale for the International Lung Screening Trial Ann Am Thorac Soc* 2020;17:503–12.
4 Venkadesh KV, Setio AAA, Schreuder A, *et al*. Deep Learning for Malignancy Risk Estimation of Pulmonary Nodules Detected at Low-Dose Screening CT. *Radiology* 2021;300:438–47.
5 Baldwin DR, Gustafson J, Pickup L, *et al*. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax* 2020;75:306–12.
6 González Maldonado S, Delorme S, Hüsing A, *et al*. Evaluation of Prediction Models for Identifying Malignancy in Pulmonary Nodules Detected via Low-Dose Computed Tomography. *JAMA Netw Open* 2020;3:e1921221.
7 Jacobs C, Setio AAA, Scholten ET, *et al*. Deep Learning for Lung Cancer Detection on Screening CT Scans: Results of a Large-Scale Public Competition and an Observer Study with 11 Radiologists. *Radiol Artif Intell* 2021;3:e210027.
8 van Leeuwen KG, Schalekamp S, Rutten MJCM, *et al*. Comparison of Commercial AI Software Performance for Radiograph Lung Nodule Detection and Bone Age Prediction. *Radiology* 2024;310:e230981.
9 Venkadesh KV, Aleef TA, Scholten ET, *et al*. Prior CT Improves Deep Learning for Malignancy Risk Estimation of Screening-detected Pulmonary Nodules. *Radiology* 2023;308:e223308.