AI as an adjunct in recommending lung nodule follow up

M. Din, R. Rachmat, H. Zahid, C. Horst, G. Benedetti, S. Syed

NHS Guy's and St Thomas' NHS Foundation Trust

Aims

- Accurate follow-up of lung nodules is crucial for early malignancy detection, yet adherence to BSTI guidelines varies, particularly among non-thoracic radiologists.
- This study evaluates whether an AI-based tool utilising a large language model (LLM) can enhance guideline adherence compared to recommendations made by non-thoracic radiologists.

Methods

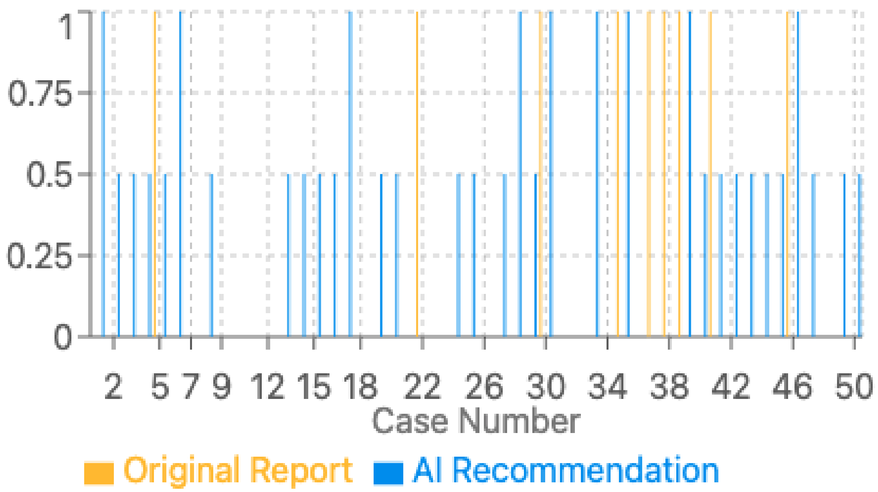
We conducted a retrospective analysis of cases (n=51) featuring lung nodules requiring follow-up. Each case compared a single non-thoracic radiologist's report against recommendations by a LLM based on identical report data.

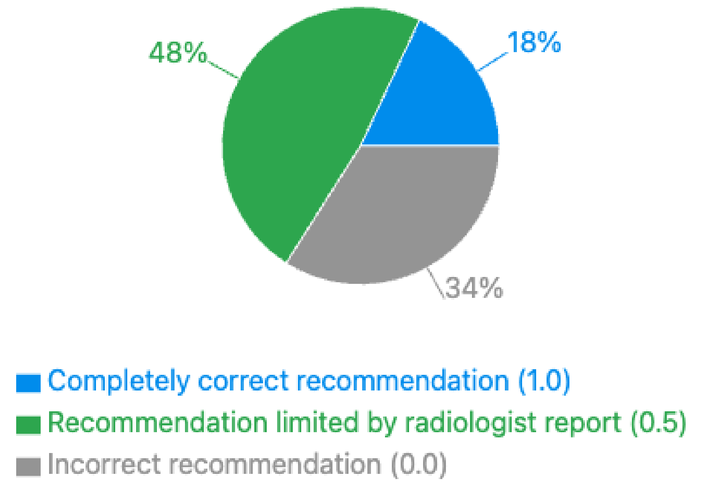
The reference standard was established by lung nodule MDM outcomes, where multiple imaging studies were reviewed by consultant cardiothoracic-radiologists and respiratory physicians.

Statistical analysis was conducted using paired t-test.

Results

- The LLM provided fully accurate follow-up recommendations in 9 cases.
- In 25 cases, it correctly stated the appropriate BSTI guideline (for subsolid vs solid nodule) but failed to state the exact next step, earning partial scores (0.5) for these.
- For example, the LLM output in one such case was *"Based on BTS guidelines, a follow-up CT scan at 3 months is recommended for the identified solid nodule in the right middle lobe".* This is because the original report only mentioned a nodule without acknowledgement of previous scans/reports showing that this nodule was already shown to be stable for 3 months.
- Guideline recommendation improved with LLM-assisted reporting compared to general radiologists who had only 9/51 correct recommendations (p =<0.001).





Conclusion

- AI demonstrates superior performance in recommending lung nodule follow-up compared to conventional (non-subspeciality) radiologist reporting. In certain cases, it was able to find discrepancies in the report body vs report conclusion.
- However, LLM effectiveness is significantly limited by the available clinical documentation/reporting.
- A follow up study which makes multiple previous text reports available to the LLM with up-to-date guidelines would be of benefit to assess performance.
- This study highlights a potential use for LLM based AI in radiology report augmentation while also demonstrating its pitfalls.