# External validation of a commercially available AI tool for nasogastric tube (NGT) position decision support in the NHS

Bartsch A-M[1], Packer J[2], Storey M[1], Togher D[3], Dean G[2], Shelmerdine S[4]

1. *Department of Radiology, St George's University Hospitals NHS Foundation Trust*
2. *Department of Radiology, Epsom & St Helier University Hospitals NHS Trust*
3. *Department of Radiology, University Hospitals of Leicester NHS Trust*
4. *Department of Radiology, Great Ormond Street Hospital for Children NHS Foundation Trust*

Epsom and St Helier University Hospitals NHS Trust

## Introduction & Background

Nasogastric tubes (NGTs) are commonly used in clinical practice for hydration, nutrition, and medication delivery. Ensuring correct NGT placement is critical, as use of a misplaced NGT can be fatal and is considered a serious 'Never Event' by National Health Service (NHS) standards.

Chest radiographs are used to assess NGT position in cases where pH evaluation is inconclusive. However, timely reporting is variable due to a shortage of the radiology workforce, which can lead to delays in patient management and potential harm despite the national safety measures in place.

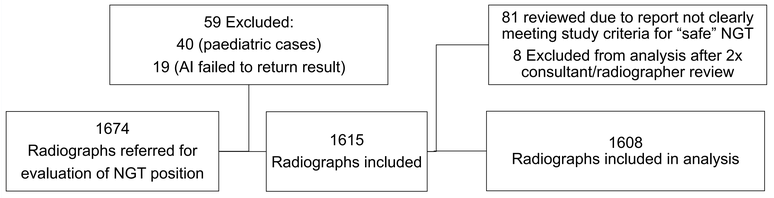
We evaluated the performance of an AI tool (**Annalise CXR v2.3**) in the identification of unsafe NGT placement on chest radiographs, to assess the feasibility of automated reporting of NGT examinations in the future.

## Methods

Adult NGT chest radiographs performed between 16/01/2023 to 31/01/2024 were retrospectively analysed. All radiographs were formally reported by a consultant radiologist or reporting radiographer, as well as being screened by the Annalise software. The authorised reports were used as the reference standard and compared to the AI output. NGT placement was classified as either "safe" or "unsafe". In instances where the report was not clear or there was a discrepancy between the AI output and the authorised report, a reporting radiographer and consultant radiologist acted as the arbitrators by retrieving the images and determining the safety of NGT position.

A NGT was classified as "unsafe" if the tip was in the airway, oesophagus or at the gastro-oesophageal junction (GOJ) but not beyond it. A "safe" NGT was assigned to any report that described the NGT tip as being projected over the stomach or passing distal to the GOJ.

NGT placement that was considered safe by both the authorised imaging report and the AI product was considered a true negative. Those flagged as unsafe by the report, and suboptimal by the AI product were considered true positives.



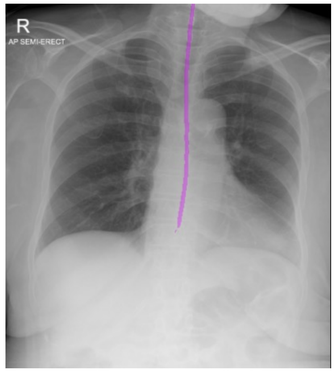




**Fig. 1**
*Example of True Positive – "unsafe" NGT identified correctly by AI*

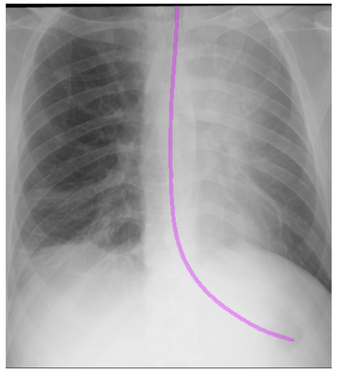


**Fig. 2**
*Example of True Negative – "safe" NGT identified correctly by AI*

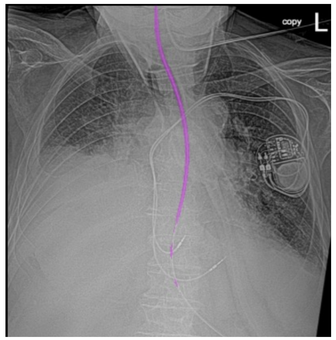



**Fig. 3**
*Example of False Positive – "safe" NGT identified incorrectly as "unsafe" by AI*

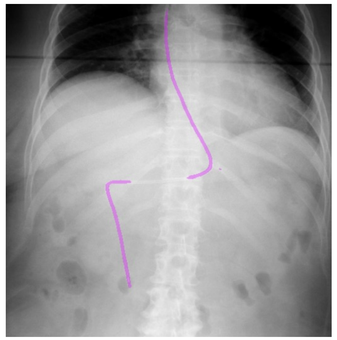


**Fig. 4**
*Example of False Positive – "safe" NGT identified incorrectly as "unsafe" by AI*

## Results

1608 NGT examinations were analysed, and 559 discrepancies occurred where the authorised report did not agree with the AI output. The sensitivity, specificity, positive and negative predictive values (PPV/NPV) of the AI decision tool are outlined below with 95% confidence intervals.

| AI Tool | Human (Gold Standard) | | |
|---|---|---|---|
| | Positive (unsafe) | Negative (safe) | |
| Positive (unsafe) | 108 | 542 | PPV 0.17 (0.14-0.20) |
| Negative (safe) | 17 | 941 | NPV 0.98 (0.97-0.99) |
| | Sensitivity 0.86 (0.79-0.91) | Specificity 0.63 (0.61-0.66) | |

## Conclusion

The AI model has a high negative predictive value (98%), indicating that the tool may be of benefit in reassuring clinicians of safe NGT placement. However, it is more limited in identifying incorrectly sited NGTs due to a high false positive rate and a sensitivity of 0.86 (missing 14% of unsafe NGTs).

Given that the few false negatives may lead to 'Never Events' such as feeding through an unsafe NGT, a 2% false negative rate (equating to 17 patients in this study) is still considered unacceptably high in this context. Therefore further refining is needed to avoid 'Never Events' as a priority, and to reduce the high number of false positives to further benefit the service.

**Why was there such a high false positive rate?**

From the discrepancy cases that were reviewed, the AI tool commonly failed to identify the full course of the NGT, leading to many being misclassified as unsafe (**Fig. 3-4**).

Discrepancies between the parameters by which the AI tool and human reporter classifies an NGT as safe or unsafe – including distance of the NGT tip from the GOJ and clear visualisation of the tip – may also contribute to misclassification.

## References

Seah JCY, Tang CHM, Buchlak QD, et al.. Effect of a comprehensive deep-learning model on the accuracy of chest X-ray interpretation by radiologists: a retrospective, multireader multicase study. Lancet Digit Health 2021;3:e496–506. 10.1016/S2589-7500(21)00106-0

Jones CM, Danaher L, Milne MR, Tang C, Seah J, Oakden-Rayner L, Johnson A, Buchlak QD, Esmaili N. Assessment of the effect of a comprehensive chest radiograph deep learning model on radiologist reports and patient outcomes: a real-world observational study. BMJ Open. 2021 Dec 20;11(12):e052902. doi: 10.1136/bmjopen-2021-052902. PMID: 34930738; PMCID: PMC8689166.