Novak *et al. Critical Care* (2025) 29:330
https://doi.org/10.1186/s13054-025-05566-6

Critical Care

RESEARCH Open Access

Evaluation of the impact of artificial intelligence-assisted image interpretation on the diagnostic performance of clinicians in identifying endotracheal tube position on plain chest X-ray: a multi-case multi-reader study

Alex Novak[1*], Sarim Ather[1], Abdala T. Espinosa Morgado[1], Giles Maskell[2], Gordon W. Cowell[3], Douglas Black[4], Akshay Shah[5,6], James S. Bowness[7], Amied Shadmaan[8], Claire Bloomfield[9], Jason L. Oke[9], Hilal Johnson[9], Mark Beggs[9], Fergus Gleeson[9], Peter Aylward[10], Aqib Hafeez[11], Moustafa Elramlawy[12], Kin Lam[15], Benjamin Griffiths[13], Mirae Harford[14], Louise Aaron[12], Claire Seeley[14], Matthew Luney[12], James Kirkland[13], Louise Wing[13], Zahi Qamhawi[13], Indrajeet Mandal[13], Thomas Millard[13], Michelle Chimbani[13], Athirah Sharazi[13], Emma Bryant[13], Wendy Haithwaite[13] and Aurora Medonica[13]

Abstract

Background Incorrectly placed endotracheal tubes (ETTs) can lead to serious clinical harm. Studies have demonstrated the potential for artificial intelligence (AI)-led algorithms to detect ETT placement on chest X-Ray (CXR) images, however their effect on clinician accuracy remains unexplored. This study measured the impact of an AI-assisted ETT detection algorithm on the ability of clinical staff to correctly identify ETT misplacement on CXR images.

Methods Four hundred CXRs of intubated adult patients were retrospectively sourced from the John Radcliffe Hospital (Oxford) and two other UK NHS hospitals. Images were de-identified and selected from a range of clinical settings, including the intensive care unit (ICU) and emergency department (ED). Each image was independently reported by a panel of thoracic radiologists, whose consensus classification of ETT placement (correct, too low [distal], or too high [proximal]) served as the reference standard for the study. Correct ETT position was defined as the tip located 3–7 cm above the carina, in line with established guidelines. Eighteen clinical readers of varying seniority from six clinical specialties were recruited across four NHS hospitals. Readers viewed the dataset using an online platform and recorded a blinded classification of ETT position for each image. After a four-week washout period, this was

*Correspondence:
Alex Novak
alex.novak@ouh.nhs.uk

Full list of author information is available at the end of the article

repeated with assistance from an AI-assisted image interpretation tool. Reader accuracy, reported confidence, and timings were measured during each study phase.

Results 14,400 image interpretations were undertaken. Pooled accuracy for tube placement classification improved from 73.6 to 77.4% ($p = 0.002$). Accuracy for identification of critically misplaced tubes increased from 79.3 to 89.0% ($p = 0.001$). Reader confidence improved with AI assistance, with no change in mean interpretation time at 36 s per image.

Conclusion Use of assistive AI technology improved accuracy and confidence in interpreting ETT placement on CXR, especially for identification of critically misplaced tubes. AI assistance may potentially provide a useful adjunct to support clinicians in identifying misplaced ETTs on CXR.

Keywords Artificial intelligence, Chest X-ray, Endotracheal tube, Medical imaging, Radiology, Tube misplacement

Background

Endotracheal intubation remains the gold-standard method for securing a definitive airway in patients with reduced levels of consciousness, impending airway compromise, severe respiratory failure or to facilitate major surgery. Over 60% of critically ill patients require endotracheal intubation, with around 45% of patients experiencing adverse peri-intubation events [1]. Studies have reported endotracheal tube (ETT) misplacement in up to 17% of intubations, with risk factors such as emergency intubation in non-theatre settings commonly being identified [2–4]. Inaccurately positioned ETTs can lead to inadequate ventilation and oxygenation, resulting in impaired gas exchange, worsening respiratory failure, and diminished oxygen delivery to vital organs. Proximal misplacement can result in cuff leaks, and an increased risk of inadvertent extubation. Additionally, distal misplacement can contribute to complications such as inadvertent one-lung ventilation, pneumothorax, atelectasis, and pneumonia [5]. Prompt detection and correction of ETT misplacement are crucial to prevent adverse outcomes. Detecting misplacement through difficulty in ventilation, clinical examination, or by physiological derangement (e.g., continuous monitoring/capnography) is not without error and ETT position for patients requiring prolonged intubation is typically confirmed via chest x-ray (CXR) [6, 7]. Interpretation of the ETT position with respect to the carina on CXR can itself be challenging, depending on a number of factors such as the experience of the clinician, the quality of the image, characteristics of the patient or type of ETT [8].

CXR interpretation has been a particular focus for development and evaluation of AI-assisted image interpretation algorithms, with a large number of commercially available applications which are able to assist in the detection of a range of abnormalities such as cancer, pneumothorax, and pleural effusion [9–15]. A recent editorial in *Intensive Care Medicine* outlined the potential significance of AI-assisted CXR interpretation in the context of critical care [16]. However, while algorithms are already in clinical use worldwide, bodies such as the UK National Institute for Health and Care Excellence (NICE) have highlighted the relative lack of current evidence to justify their use [17]. A number of recent retrospective studies have reported on the accuracy of different dedicated AI algorithms to detect ETT position on CXR, with most of the literature comprised of comparisons between algorithm performance and a radiologist or clinician-derived reference standard, yet none of these algorithms are FDA-approved or CE-marked for independent autonomous detection, but instead are intended as assistive devices to aid clinical interpretation. NICE guidance and AI-specific reporting guidelines have emphasised the importance of conducting evaluations in the clinical context in which they are likely to be sited, including feedback on usability and confidence directly from the intended users [18–21]. To date however there has been limited research which directly evaluates the impact of these algorithms on the interpretation performance of the clinicians who routinely encounter such patients and are required to interpret ETT position CXRs in clinical practice [3, 22–28].

Aims

We aimed to directly evaluate the impact of AI-assisted image interpretation on the ability of critical and acute care clinicians to accurately identify ETT misplacement on CXR. To achieve this we used the GE HealthCare (GEHC) Critical Care Suite, a collection of AI-based assisted image analysis algorithms which are FDA-approved and CE-marked, and already in clinical use in some centres [29]. This application includes an algorithm for detecting and localising ETT position on posteroanterior CXRs, with its output presented as an overlay indicating both the outline of the ETT and the position of the carina, with a corresponding measurement indicating their relative distance (Fig. 1).

Methods

We undertook a multicentre cohort, multi-case, multi-reader study which was conducted between January 2023 and April 2023. This closely followed the structure

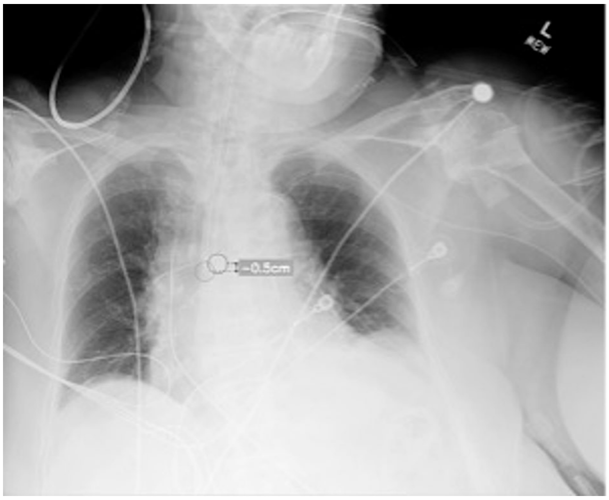

Fig. 1 Plain chest radiograph showing overlay from Critical Care Suite, with segmentation of ETT, identification of tube tip and carina, and calculation ofdistance between them

and approach of another recently published study conducted by our research group [30]. This manuscript follows the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guidelines for studies evaluating the performance of diagnostic tests [31, 32].

A total of 400 retrospectively collected and de-identified CXR images of patients requiring tracheal intubation, aged 18 years or older, were identified by searching the CRIS (Clinical Record Interactive Search) database in Oxford University Hospitals NHS Foundation Trust, Oxford, UK and were subsequently curated into the project via the NCIMI (National Consortium of Intelligent Medical Imaging) Databank.

The images were sourced from The John Radcliffe Hospital, a major tertiary referral centre and the principal hospital within the Trust. This 832-bedded hospital serves as the main accident and emergency service and the Major Trauma Centre for the Thames Valley region, providing acute medical, surgical, and intensive care services to a diverse population of approximately 655,000 people. The dataset includes images acquired from a range of clinical settings, notably the intensive care unit and emergency department, reflecting the hospital's broad patient base and high annual throughput, with around 180,000 emergency department visits and approximately 1,000 ICU admissions per year. Due to the de-identified nature of the dataset, further summary data on specific clinical characteristics or patient subgroups (e.g., surgical vs. medical) could not be obtained.

The initial dataset images were independently inferenced by the GE Critical Care Suite ETT algorithm to create a separate image dataset containing AI-assisted overlays for the second reader phase.

To define a reference standard, four senior consultant thoracic radiologists were recruited; one from Royal Cornwall Hospitals NHS Trust, two from NHS Greater Glasgow and Clyde, and one from Oxford University Hospitals NHS Foundation Trust. All CXR images were independently reviewed by three radiologists who marked ETT tip and carina, measured the distance between them using a web-based DICOM viewer (RAIQC), then classified them as correctly placed (when the tube tip is 30–70 mm above the carina), too high or proximal (when the tube tip is more than 70 mm above the carina), or too low or distal (when the tube tip is less than 30 mm above the carina). Critical misplacement was defined as tube tip position either less than 20 mm or greater than 80 mm above the carina. One radiologist reported the entire dataset and two radiologists each independently reported half of the dataset to provide a second read for all images. In cases of discordance in classification between the first and second reads, arbitration was undertaken by the fourth radiologist. Concordance between the algorithm output and the reference standard was independently measured and analysed, and has been published separately [33].

As per our previous study [30], all images were analysed by the radiologists for characteristics which would make interpretation challenging to human readers. This included categorising each image in terms of contributing factors for increasing the difficulty of determining ETT position– (i) patient specific factors (e.g., kyphosis or obesity); (ii) image quality (exposure factors and image penetration); (iii) presence and number of foreign objects on the image (nasogastric tube, another tube or (metal) line running parallel to or crossing the ETT, (iv) any structure obstructing tube tip or carina); and (v) type of ETT (standard/reinforced/double-lumen).

The reader panel was comprised of 18 clinicians with three levels of seniority (Consultant/Senior– more than seven years' experience; Middle Grade/Registrar– four to seven years; Junior– less than four years) equally derived from six different clinical specialties (Adult Intensive Care, Anaesthetics, Emergency Medicine, Anaesthetic Advanced Care Practitioners, Radiology and Radiography), who were working across four NCIMI hospital sites (one tertiary/major trauma centre, three busy district general hospitals) via the Thames Valley Emergency Medicine Research Network. Reads were undertaken remotely online using the Report and Image Quality Control (RAIQC) DICOM viewer which simulates a standard hospital PACS system. After enrolment, readers were given access to an online training module with a series of five test cases in order to familiarise them with the study design and online platform.

The reader study was split into two phases. In the first phase, readers were asked to complete reads on the

RAIQC platform over a three-week period, during which they interpreted the entire image dataset. For each CXR they measured the distance between tip and carina on each image using the tools on the DICOM viewer and categorised the perceived position of the ETT on each image (too high/proximal; well positioned; too low/distal) and recorded their confidence on a four-point Likert scale (not confident; low confidence; high confidence; certain). Readers were blinded to the ground truth for each image. The order of images in each module was randomised for each reader.

This first phase was followed by a subsequent four-week 'washout' period before commencing the second reader phase, during which participants re-interpreted all the images in a randomised order, this time with access to the output of the AI algorithm. Qualitative surveys of user experience were completed by all participants before and after the two phases. Time taken to complete each read was recorded though an automated function on the DICOM viewer.

The primary outcome was a change in readers' diagnostic accuracy for the categorisation of ETT placement (too high/proximal; well positioned; too low/distal), with versus without AI assistance. Secondary outcomes included analysis in subgroups by degree of tube misplacement, medical profession/specialty, level of seniority, image difficulty, reader confidence and time taken to complete the reads. Statistical analysis was undertaken as per the methodology used in previous studies of this kind published by our research group [30, 34–36]. To account for correlated errors arising from readers interpreting the same cases with and without AI, the Obuchowski and Rockette, Dorfman-Berbaum-Metz (OR-DBM) procedure (a modality-by-reader random effects ANOVA model) was used for estimation. The "Multi-Reader Sample Size Program for Diagnostic Studies" was used to estimate power for the number of readers cases in our study. For 18 readers, reading 400 cases was calculated to yield 80% power to detect a difference in accuracy of 10% with a Type I error of 5% [37]. Statistical analyses were all performed using R software (v4·0·2; R Foundation for Statistical Computing, Vienna, Austria). The significance threshold was set at two-sided 5% ($p < 0{\cdot}05$) for all secondary analyses.

The study was conducted under an ethical framework held between the National Consortium of Intelligent Medical Imaging and Oxford University Hospitals NHS Foundation Trust, with data collection approved through a Data Protection Impact Assessment. Patient and Public Involvement (PPI) input was sought from the Academic Centre for Urgent and Emergency Care (ACUTECare) PPI Group during the design of the study and following initial data analysis.

Results

A total of 400 CXRs were initially incorporated into the dataset, of which 209 were classified as 'well positioned', with 56 'too high (proximal)', and 134 'too low (distal)'. Due to a single error in the initial case selection, one CXR was initially included in the dataset which did not include a tracheal tube. This case was subsequently removed from analysis. All CXR images were successfully inferenced using the GE CCS ETT algorithm. A summary of the make-up and characteristics of the dataset images is included in the supplementary materials, table S5.

Each reader interpreted each image both with and without AI assistance, with a resultant total of 14,400 image interpretations. Results for overall reader accuracy for the classification of tube placement with and without AI assistance are presented in Fig. 2. Accurate classification of tube placement across the whole group increased from 73.6% (95%CI: 68.7–78.6) without AI assistance to 77.4% (95%CI 72.2–82.5), $p = 0.002$ with AI assistance. In subgroup analyses accuracy for identification of critically misplaced tubes significantly increased from 79.3% (95%CI 72.9–85.7) to 89.0% (95%CI 85.2–92.8, $p = 0.001$), with differences in accuracy for proximal (too high) critically misplaced tubes increasing from 52.0% (95%CI 39.3–64.8) to 77.5% (95%CI 67.7–87.3, $p < 0.0001$), and distal (too low) critically misplaced tubes from 88.9% (95%CI 83.2–94.7) to 93.1% (95%CI 89.3–96.8, $p = 0.05$) (see Fig. 3, and Table S1 in supplementary material). A statistically significant increase was demonstrated in the middle grade seniority subgroup (75.4% [95%CI 65.4–85.3] to 79.8% [95%CI 69.7–89.9], $p = 0.004$), with no statistically significant changes demonstrated in other seniority subgroups, including junior (75.0% [95%CI 65.9–84.0] to 79.7% [95%CI 71.6–87.7], $p = 0.23$), senior (74.0% [95%CI 66.7–81.3] to 76.5% [95%CI 69.7–89.9], $p = 0.25$) and Advanced Nurse Practitioner (67.9% [95%CI 22.5–100.0] to 70.9% [95%CI 22.4–100.0], $p = 0.13$). The ICU subgroup showed a statistically significant improvement in correct classification of 6.9% (95%CI 4.3–9.6, $p < 0.001$) (See Figs. 4 and 5, and Table S2 in supplementary material). Individual reader accuracies with and without AI are presented in supplementary material; table S6 and figures S1 to S2. Differences in reader accuracy for different image characteristic subgroups are summarised in Table S3 and Figures S3 to S9 in Supplementary Material.

The impact of the use of the CCS PTX AI tool on time taken for reader to interpret images is shown in Table S4 in Supplementary Material. In the absence of the AI tool, mean reporting time by all 18 readers was 35.7 s per image. The use of the CCS PTX AI tool had no statistically significant effect on reporting time with a mean of 36.0 s per image ($p = 0.47$ for effect of AI tool).

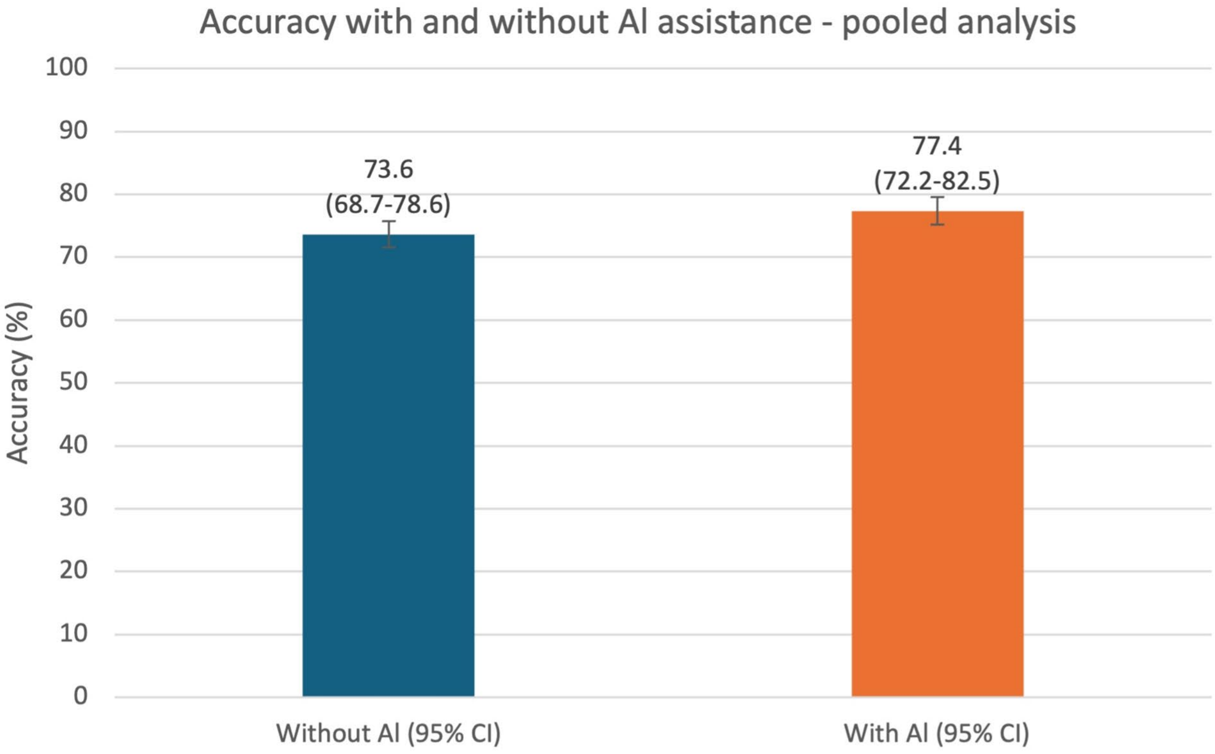

Fig. 2 Bar chart showing pooled accuracy with and without use of AI assistance

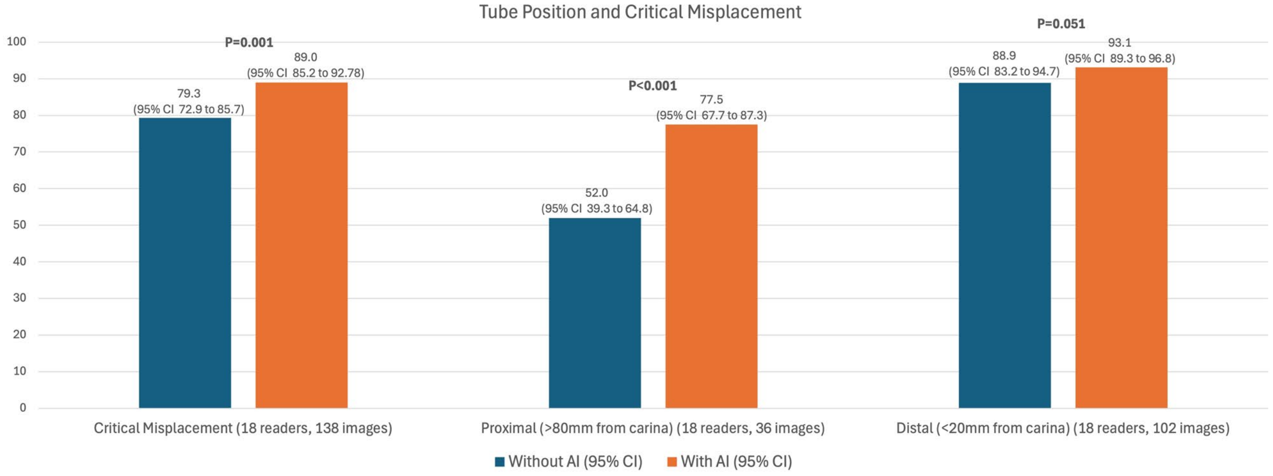

Fig. 3 Bar chart showing the accuracy for identifying critically incorrectly placed tubes, comparison between without AI and with AI

Overall, confidence in correctly interpreted images generally increased in the aided reader phase compared with unaided reader phase. The proportion of ‘certain’ and ‘high’ confidence interpretations in the correct interpretation category (i.e., true positives/true negatives) increased in the aided reader phase (Supplementary materials; Table S7, and Figure S10). However, small increases in reported ‘certainty’ and ‘high confidence’ were observed for incorrect classifications of well-positioned tubes as misplaced with AI assistance.

Discussion

The key findings of the study were: (i) a small but statistically significant improvement in pooled reader accuracy for correct classification of ETT placement; (ii) a marked improvement in pooled reader accuracy for detecting

Fig. 4 Bar chart showing a comparative analysis of reader performance with and without AI according to the reader's specialty

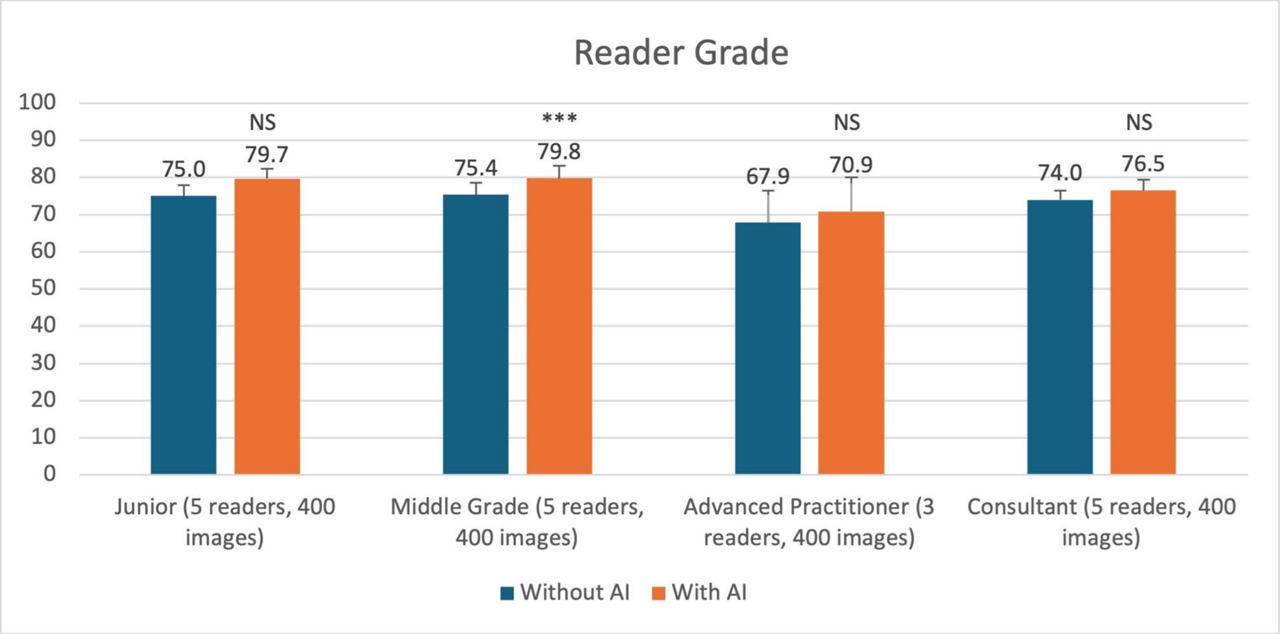

Fig. 5 Bar chart showing a comparative analysis of reader performance with and without AI according to the reader's level of experience

critically misplaced tubes, with no increase in clinician reporting time associated with the use of AI assistance; and (iii) an overall improvement in self-reported confidence in an accurate classification.

Other studies have explored the use of machine learning algorithms in the detection of endotracheal tube placement on chest radiography. Whilst these studies provide strong evidence for the potential use of AI to detect ETT position, they do not evaluate the impact of these algorithms in their legal and intended clinical use (i.e., as decision aids rather than replacements or competitors for human readers), hence this study represents important addition to the evidence base regarding the impact of AI-assisted image interpretation in this use case. Primary interpretation of ETT placement and clinical action based on radiograph findings is often undertaken by middle grade clinicians and our study offers new evidence to support the potential clinical efficacy of AI assistance within this cohort. Other studies have demonstrated similar findings in terms of maintained productivity with the use of AI with similar interpretation times both with and without use of the algorithm [14, 38].

The significant improvement seen in accurate detection of critically misplaced tubes is an important finding, as this arguably represents a key application for the algorithm in clinical practice. Subgroup analysis suggested a greater effect from use of the AI algorithms seen in less experienced (middle grade) readers, though this effect did not reach statistical significance in the most junior group. No significant change in the reader accuracy with and without AI assistance was seen when categorized in specialty subgroups (apart from intensive care doctors), however the relatively small size of these subgroups may have limited the ability to demonstrate a statistically significant effect. Whilst ICU clinicians are arguably the most directly relevant professional group evaluated in this context, the small sample size ($n = 3$) limits the generalisability of any findings specifically related to the specialty.

Taken together, these findings have the potential to translate into better clinical outcomes by lowering morbidity associated with misplaced ETTs and improving clinician confidence. An improvement in accuracy was demonstrated across most image characteristic subcategories, with those not demonstrating statistically significant improvements tending to have smaller sample size which would require a larger effect to show significance. There was no clear trend in accuracy related to image difficulty characteristics, i.e. images characteristics judged to increase the difficulty of interpretation were not associated with greater or lesser changes in accuracy with AI. Productivity was unaffected by the algorithm, with the time taken by clinicians to interpret each image being unaffected.

This study employed a large, carefully curated dataset derived from real-world cases with varying degrees of complexity, thus encompassing a broad range of image characteristics which would be expected to occur in routine clinical practice. Reads were undertaken by a broad range of clinicians and radiographers of varying seniority from multiple hospital sites. As such, it is the first known study to explore the direct impact of AI on the reporting performance of frontline clinical practitioners in detecting ETT placement on plain CXR, and one of the first AI multi-case, multi-reader studies to be conducted in the NHS.

Our study has limitations. To achieve a statistical balance this study used an artificially high prevalence of misplaced ETTs which would not normally be encountered in clinical practice. Equally, readers undertook multiple sequential CXR interpretations which does not reflect a normal clinical workflow. The algorithm output (i.e. the identification of ETT position) did not directly correlate to the primary outcome measure used in this study (i.e., the interpretation of that position as misplaced or well positioned), which depended on reader interpretation. Whilst the reader group was large compared to other studies of its kind, only one representation of each reader specialty was included at a given level of seniority. However, this is fully commensurate with other reader studies of its kind [30, 38], and the results from pooled analyses are still more likely to be generalizable to the wider clinical population as compared to previous similar studies. It is possible that some of the improvement in classification seen in the study may have been due partly to the increase in experience from interpreting large numbers of chest radiographs containing ETTs, however a sequential analysis of pooled reader performance does not indicate such a trend. Critically, readers in this study interpreted images free from the distractions associated with a busy clinical environment– this may have improved their baseline accuracy, potentially reducing the perceived impact of the intervention, and real-world evaluations in clinical practice would be required to fully evaluate this effect.

The findings of this study suggest that the use of AI-assisted image interpretation may improve the overall diagnostic performance and confidence of clinicians and radiographers when identifying ETT misplacement on CXR. A small but statistically significant improvement in ETT placement accuracy could enhance patient safety by reducing complications and facilitating quicker interventions potentially leading to fewer adverse events and improved outcomes, particularly in challenging cases or critical care settings. The ease with which this technology can be integrated with existing workflows and applied across a large population of intubated patients in multiple centres means that even a modest increase in accuracy may be beneficial in airway management and in reducing rare but serious adverse outcomes. Of note, the largest benefits were seen for the accurate identification of critically misplaced tubes, and for the accuracy of middle grade clinicians, which represent key clinical applications for this technology. Whilst this retrospective in silico study found a significant improvement in the AI-assisted detection accuracy of readers for misplaced tubes, further work must be done in the real life setting however to fully evaluate the clinical impact or health economic benefit. Furthermore, future studies could include the radiographs of critically ill children which represent another important potential use case for this technology [39–41].

Conclusions

In summary, use of AI-assisted interpretation demonstrated a statistically significant improvement in the ability of clinicians to identify misplaced endotracheal tubes on CXR, with most significant benefits seen for critically misplaced tubes. AI assistance could potentially be useful adjunct to support clinicians in identifying misplaced endotracheal tubes on chest X-Ray, especially for less experienced practitioners. Future prospective studies are required to generate definitive evidence in this regard and to determine the magnitude of any potential clinical and health economic impact.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13054-025-05566-6.

Supplementary Material 1.

Acknowledgements

All datasets and documents related to this study currently reside securely in Oxford University Hospitals NHS Foundation Trust, and will be made available upon reasonable request to the corresponding author. The AI algorithm used in this research is a commercially available third-party product and as such the authors do not have sharing rights– enquires can be made via the GE Healthcare website (https://www.gehealthcare.com). This work was funded and supported by the National Consortium of Intelligent Medical Imaging through the Industrial Strategy Challenge Fund (Innovate UK Grant 104688). GEHC provided inferencing of the CXR images, otherwise research activity and data analysis were conducted independently of the funder. We gratefully acknowledge Amied Shadmaan, Sharon Ghelman, Poonam Dalal and Alex Baenen from GE Healthcare for their invaluable contributions to this study. Their expertise and support were instrumental in the successful completion of this research. PA, SA and FG are employees of Report and Image Quality Control (RAIQC http://www.raiqc.com, a spin-out company from Oxford University Hospitals NHS Foundation Trust. ASh is an Editor of Anaesthesia. JSB is a Senior Clinical Advisor to Intelligent Ultrasound, a company making AI products for medical ultrasound.

Co-authors - the critical care suite endotracheal tube reader study group:

Peter Aylward[3], Aqib Hafeez[2], Mustafa Elramlawy[9], Kim Lam[7], Benjamin Griffiths[2], Mirae Harford[8], Louise Aaron[9], Claire Seeley[8], Mathew Luney[9], James Kirkland[2], Louise Wing[2], Zahi Qamhawi[2], Indrajeet Mandal[2], Thomas Millard[2], Michelle Chimbani[2], Athirah Sharazi[2], Emma Bryant[2], Wendy Haithwaite[2], Aurora Medonica[2]. The Corresponding Author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to (i) publish, reproduce, distribute, display and store the Contribution, (ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, (iii) create any other derivative work(s) based on the Contribution, (iv) to exploit all subsidiary rights in the Contribution, (v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, (vi) licence any third party to do any or all of the above.

Author contributions

AN was responsible for the overall conduct of the project including analysis of results, independent write up, and publication. Independent statistical analysis was performed by AN and JO. GM, GC, DB and FG conducted ground truth reads, whilst the clinical reader group comprised AH, ME, KL, BG, MH, LA, CS, ML, JK, LW, ZQ, IM, TM, MC, AS, EB, WH, AM. ATEM, ASh and JB helped prepare the manuscript, and MB was involved in study design, contract completion data analysis, report generation and steering group inputs. FG was involved in study design and ground truth assessment.

Data availability

All datasets and documents related to this study currently reside securely in Oxford University Hospitals NHS Foundation Trust, and will be made available upon reasonable request to the corresponding author. The AI algorithm used in this research is a commercially available third-party product and as such the authors do not have sharing rights– enquires can be made via the GE Healthcare website. https://www.gehealthcare.com

Declarations

Ethics approval and consent to participate

All participants provided written informed consent before enrolment in the study. The consent process included a detailed explanation of the study's purpose, procedures, potential risks and benefits, and the voluntary nature of participation. Participants were informed about the two phases of the study - interpreting chest X-rays without and then with AI assistance after a washout period. They were made aware that their interpretations would be compared to expert radiologist readings. Participants were assured of their right to withdraw from the study at any time without any negative consequences. Participants were also informed that their anonymised data would be analysed and potentially published in aggregate form. The consent process emphasized that participation was voluntary and would not affect their employment or standing at their institution.

Consent for publication

Not Applicable. This manuscript does not contain any individual person's data in any form, including identifying images or personal or clinical details that could compromise anonymity. As such, no specific consent for publication was required from any individuals. All data presented in this study are aggregated and anonymized, ensuring the privacy and confidentiality of all participants involved in the research. We have adhered strictly to ethical guidelines and data protection regulations throughout the study and in the preparation of this manuscript.

Competing interests

AS, SG, AB are employed by GE HealthCare, a key NCIMI stakeholder. AN and CB have undertaken paid consultancy work for GEHC. PA, SA and FG are employees of Report and Image Quality Control (http://www.raiqc.com RAIQC), a spin-out company from Oxford University Hospitals NHS Foundation Trust. A.Shah is an Editor of Anaesthesia and has received honorariums from Pharmacosmos UK, outside of the submitted work. JSB is a Senior Clinical Advisor to Intelligent Ultrasound, a company making AI products for medical ultrasound. There are no other relationships or activities that could appear to have influenced the submitted work.

Author details

[1]Oxford Clinical Artificial Intelligence Research (OxCAIR), Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[2]Royal Cornwall Hospitals NHS Trust, Cornwall, UK
[3]Department of Imaging, Queen Elizabeth University Hospital, Glasgow, UK
[4]NHS Greater Glasgow and Clyde, Glasgow, UK
[5]Department of Anaesthesia, Hammersmith Hospital, Imperial College Healthcare NHS Trust, London, UK
[6]Department of Anaesthesia, University College London Hospitals NHS Foundation Trust, London, UK
[7]Department of Targeted Intervention, University College London, London, UK
[8]GE Healthcare (GEHC), Chalfont St. Giles, UK
[9]The University of Oxford, Oxford, UK
[10]Reporting and Image Quality Control Ltd, Oxford, UK
[11]Emergency Medicine Research Oxford (EMROx), Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[12]Buckinghamshire Healthcare NHS Trust, Aylesbury, UK
[13]Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[14]Royal Berkshire NHS Foundation Trust, Reading, UK
[15]NHS Frimley Health Foundation Trust, Frimley, UK

Received: 6 March 2025 / Accepted: 12 July 2025

References

1. Smischney N, Kashyap R, Seisa M, Schroeder D, Diedrich D. Endotracheal intubation among the critically ill: protocol for a multicenter, observational, prospective study. JMIR Res Protoc. 2018;7(12):e11101.
2. Özkurtul O, Struck MF, Fakler J, Bernhard M, Seinen S, Wrigge H, et al. Physician-based on-scene airway management in severely injured patients and in-hospital consequences: is the misplaced intubation an underestimated danger in trauma management? Trauma Surg Acute Care Open. 2019;4(1):e000271.
3. Jemmett ME, Kendal KM, Fourre MW, Burton JH. Unrecognized misplacement of endotracheal tubes in a mixed urban to rural emergency medical services setting. Acad Emerg Med. 2003;10(9):961–5.
4. Maybauer D, Maybauer M, Wolff H, Pfenninger E, Geisser W. Radiological validation of endotracheal tube insertion depth in prehospital emergency patients. Crit Care. 2009;13(Suppl 1):P3.
5. Wang HE, Cook LJ, Chang CCH, Yealy DM, Lave JR. Outcomes after out-of-hospital endotracheal intubation errors. Resuscitation. 2009;80(1):50–5.
6. Hossein-Nejad H, Payandemehr P, Bashiri SA, Nedai HHN. Chest radiography after endotracheal tube placement: is it necessary or not? Am J Emerg Med. 2013;31(8):1181–2.
7. Silvestri S, Ladde JG, Brown JF, Roa JV, Hunter C, Ralls GA, et al. Endotracheal tube placement confirmation: 100% sensitivity and specificity with sustained four-phase capnographic waveforms in a cadaveric experimental model. Resuscitation. 2017;115:192–8.
8. Kamel T, Sauvage B, Lakhal K, Ottavy G, Janssen-Langenstein R, Jacquier M, et al. The accuracy of intensive care nurses' interpretation of chest radiographs to recognise misplacement of endotracheal and nasogastric tubes after a single training session and comparison with residents' interpretation. Aust Crit Care. 2023;36(6):948–54.
9. Van Leeuwen KG, Schalekamp S, Rutten MJCM, Van Ginneken B, De Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol. 2021;31(6):3797–804.
10. Seah JCY, Tang CHM, Buchlak QD, Holt XG, Wardman JB, Aimoldin A, 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(8):e496–506.
11. Yoo H, Kim EY, Kim H, Choi YR, Kim MY, Hwang SH, et al. Artificial Intelligence-Based identification of normal chest radiographs: A simulation study in a multicenter health screening cohort. Korean J Radiol. 2022;23(10):1009.
12. Jin KN, Kim EY, Kim YJ, Lee GP, Kim H, Oh S, et al. Diagnostic effect of artificial intelligence solution for referable thoracic abnormalities on chest radiography: a multicenter respiratory outpatient diagnostic cohort study. Eur Radiol. 2022;32(5):3469–79.
13. Van Beek EJR, Ahn JS, Kim MJ, Murchison JT. Validation study of machine-learning chest radiograph software in primary and emergency medicine. Clin Radiol. 2023;78(1):1–7.
14. Ahn JS, Ebrahimian S, McDermott S, Lee S, Naccarato L, Di Capua JF, et al. Association of artificial Intelligence–Aided chest radiograph interpretation with reader performance and efficiency. JAMA Netw Open. 2022;5(8):e2229289.
15. Homayounieh F, Digumarthy S, Ebrahimian S, Rueckel J, Hoppe BF, Sabel BO, et al. An artificial Intelligence–Based chest X-ray model on human nodule detection accuracy from a multicenter study. JAMA Netw Open. 2021;4(12):e2141096.
16. Joskowicz L, Beil M, Sviri S. Artificial intelligence interpretation of chest radiographs in intensive care. Ready for prime time? Intensive Care Med. 2025;51:154–156.
17. Overview NICE. 2023 [cited 2025 Jan 27]. Available from: https://www.nice.org.uk/guidance/hte12
18. Unsworth H, Wolfram V, Dillon B, Salmon M, Greaves F, Liu X, et al. Building an evidence standards framework for artificial intelligence-enabled digital health technologies. Lancet Digit Health. 2022;4(4):e216–7.
19. Ibrahim H, Liu X, Rivera SC, Moher D, Chan AW, Sydes MR, et al. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. Trials. 2021;22(1):11.
20. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924–33.
21. RCEM_Position_Statement_-Artificial_Intelligence_Dec_2022. pdf [Internet]. [cited 2025 Jan 27]. Available from: https://rcem.ac.uk/wp-content/uploads/2022/12/RCEM_Position_Statement_-Artificial_Intelligence_Dec_2022.pdf
22. Lakhani P, Flanders A, Gorniak R. Endotracheal tube position assessment on chest radiographs using deep learning. Radiol Artif Intell. 2021;3(1):e200026.
23. Schultheis WG, Lakhani P. Using deep learning segmentation for endotracheal tube position assessment. J Thorac Imaging. 2022;37(2):125–31.
24. Huang M, Chen CY, Horng MH, Li CI, Hsu IL, Su CM, et al. Validation of a deep Learning–based automatic detection algorithm for measurement of endotracheal Tube–to–Carina distance on chest radiographs. Anesthesiology. 2022;137(6):704–15.
25. Hsu CC, Ameri R, Lin CW, He JS, Biyari M, Yarahmadi A, et al. A robust approach for endotracheal tube localization in chest radiographs. Front Artif Intell. 2023;6:1181812.
26. Wang CH, Hwang T, Huang YS, Tay J, Wu CY, Wu MC, et al. Deep Learning-Based localization and detection of malpositioned endotracheal tube on portable supine chest radiographs in intensive and emergency medicine: A multicenter retrospective study**. Crit Care Med. 2024;52(2):237–47.
27. Tang CHM, Seah JCY, Ahmad HK, Milne MR, Wardman JB, Buchlak QD, et al. Analysis of line and tube detection performance of a chest X-ray deep learning model to evaluate hidden stratification. Diagnostics. 2023;13(14):2317.
28. Jones CM, Danaher L, Milne MR, Tang C, Seah J, Oakden-Rayner L, et al. 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;11(12):e052902.
29. Critical Care Suite. 2.0 [Internet]. [cited 2025 Feb 12]. Available from: https://www.gehealthcare.co.uk/products/radiography/critical-care-suite
30. Novak A, Ather S, Gill A, Aylward P, Maskell G, Cowell GW, et al. Evaluation of the impact of artificial intelligence-assisted image interpretation on the diagnostic performance of clinicians in identifying Pneumothoraces on plain chest X-ray: a multi-case multi-reader study. Emerg Med J. 2024;41(10):602–9.
31. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.
32. Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open. 2021;11(6):e047709.
33. Novak A, Ather S, Espinosa Morgado AT, Shah A, Bowness JS, Maskell G, Cowell GW, Black D, Oke JL, Johnson H, Bloomfield C, Beggs M, Gleeson F. Performance analysis of an artificial intelligence algorithm for detection and localisation of tracheal tube position by chest radiography. Br J Anaesth. 2025;134(5):1554–1557.
34. Fu H, Novak A, Robert D, Kumar S, Tanamala S, Oke J, et al. AI assisted reader evaluation in acute CT head interpretation (AI-REACT): protocol for a multireader multicase study. BMJ Open. 2024;14(2):e079824.
35. Khan F, Das I, Kotnik M, Wing L, Van Beek E, Murchison J, et al. AI-assisted detection for chest X-rays (AID-CXR): a multi-reader multi-case study protocol. BMJ Open. 2024;14(12):e080554.
36. Novak A, Hollowday M, Espinosa Morgado AT, Oke J, Shelmerdine S, Woznitza N, et al. Evaluating the impact of artificial intelligence-assisted image analysis on the diagnostic accuracy of front-line clinicians in detecting fractures on plain X-rays (FRACT-AI): protocol for a prospective observational study. BMJ Open. 2024;14(9):e086061.
37. Hillis SL, Schartz KM. Multireader sample size program for diagnostic studies: demonstration and methodology. J Med Imaging. 2018;5(04):1.
38. Guermazi A, Tannoury C, Kompel AJ, Murakami AM, Ducarouge A, Gillibert A, et al. Improving radiographic fracture recognition performance and efficiency using artificial intelligence. Radiology. 2022;302(3):627–36.
39. Miller KA, Kimia A, Monuteaux MC, Nagler J. Factors associated with misplaced endotracheal tubes during intubation in pediatric patients. J Emerg Med. 2016;51(1):9–18.
40. Rost F, Donaubauer B, Kirsten H, Schwarz T, Zimmermann P, Siekmeyer M, et al. Tracheal tube misplacement after emergency intubation in pediatric trauma patients: A retrospective, exploratory study. Children. 2022;9(2):289.

41. Shim JG, Lee EK, Oh EJ, Cho EA, Park J, Lee JH, et al. Predicting the risk of inappropriate depth of endotracheal intubation in pediatric patients using machine learning approaches. Sci Rep. 2023;13(1):5156.

Publisher's note