

Citation for published version:

Graby, J, Harris, M, Jones, C, Waring, H, Lyen, S, Hudson, BJ & Rodrigues, JCL 2023, 'Assessing the role of an artificial intelligence assessment tool for thoracic aorta diameter on routine chest CT', *The British journal of radiology*, vol. 96, no. 1151, 20220853. <https://doi.org/10.1259/bjr.20220853>

DOI:

[10.1259/bjr.20220853](https://doi.org/10.1259/bjr.20220853)

Publication date:

2023

Document Version

Peer reviewed version

[Link to publication](#)

© 2023 The Authors. Published by the British Institute of Radiology.

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Title: Assessing the role of an artificial intelligence assessment tool for thoracic aorta diameter on routine chest CT

Short title: The role of an AI tool for thoracic aorta measurement on routine CT

Abstract

Objectives

To assess the diagnostic accuracy and clinical impact of automated artificial intelligence (AI) measurement of thoracic aorta diameter on routine chest CT.

Methods

A single-centre retrospective study involving three cohorts. 210 consecutive ECG-gated CT aorta scans (mean age 75 ± 13) underwent automated analysis (AI-Rad Companion Chest CT, Siemens) and were compared to a reference standard of specialist cardiothoracic radiologists for accuracy measuring aortic diameter. A repeated measures analysis tested reporting consistency in a second cohort (29 patients, mean age 61 ± 17) of immediate sequential pre-contrast and contrast CT aorta acquisitions. Potential clinical impact was assessed in a third cohort of 197 routine CT chests (mean age 66 ± 15) to document potential clinical impact.

Results

AI analysis produced a full report in 387/436 (89%) and a partial report in 421/436 (97%). Manual vs AI agreement was good to excellent (ICC 0.76-0.92). Repeated measures analysis of expert and AI reports for the ascending aorta were moderate to good (ICC 0.57-0.88). AI diagnostic performance crossed the threshold for maximally accepted limits of agreement ($>5\text{mm}$) at the aortic root on ECG-gated CTs. AI newly identified aortic dilatation in 27% of patients on routine thoracic imaging with a specificity of 99% and sensitivity of 77%.

Conclusion

AI has good agreement with expert readers at the mid ascending aorta and has high specificity, but low sensitivity, at detecting dilated aortas on non-dedicated chest CTs.

Advances in knowledge

An AI tool may improve the detection of previously unknown thoracic aorta dilatation on chest CTs versus current routine reporting.

Manuscript

Introduction

Aortic aneurysm is the pathological dilatation of the aorta and is defined as $>50\%$ of the normal expected width for a given aortic segment with a tendency to expand further[1]. Normal aortic root measurements vary with sex, age and anthropometrics. However, thoracic aorta dilatation has been defined previously as $\geq 40\text{mm}$ in the ascending aorta and $\geq 30\text{mm}$ in the descending aorta, including in European guidelines[2–4].

Aortic aneurysms enlarge approximately 0.07cm/year , though this can vary based on location, demographics and aetiology[5]. Risk of fatal complications increases with size[6]. There is an incidence of 5.3 cases per 100,000 population per year, though the accuracy of this is debated given the indolent nature of the disease[7].

Most thoracic aortic aneurysms will remain asymptomatic until presentation with life-threatening complications. In patients with a diameter greater than 6cm there is a 14.1% annual risk of complication, which includes rupture and dissection[8]. In-hospital mortality for type A dissection is reported at 57% without surgery and remains approximately 23% with emergency surgery[9]. There is a 50% mortality at 30 days from dissection[10]. Additionally, patients with thoracic aortic aneurysms are at increased risk of other major adverse cardiovascular events (MACE), thought to be related to shared risk factors (e.g. smoking, hypertension)[2].

There are no known preventative treatments against developing thoracic aneurysm and so early diagnosis and management are key to reducing mortality[11]. As a result of their indolent nature and slow growth, early diagnosis is often through an incidental finding or screening where there is a strong family history[11, 12].

In the UK, around half a million CT chest and/or abdomen studies were performed in the year 2020-21[13]. In routine clinical practice, manual segmentation measurements of the aorta are very rarely performed on non-dedicated aorta CT scans, and double oblique measurements even less frequently. This is particularly true when reported by a general on-call radiologist,

with both requiring a degree of training and can be time consuming. An artificial intelligence (AI) analysis could help address this. AI tools now exist that provide automated segmentation of the aorta on a variety of CT chest acquisitions, providing a measure of vessel diameter at recognised anatomical levels[12, 14, 15]. Whilst this tool is available for commercial purchase, the authors felt a local validation was warranted prior to a final judgement.

The aims of this validation study were to: (1) compare aortic diameter measurements on CT between an automated AI assessment and that of specialist cardiothoracic radiology consultants; (2) determine the consistency of the automated measurements, and (3) assess the potential clinical impact of automated aortic segmentation on routine CT chest studies.

Materials and Methods

Patient Population

This was a retrospective study of CT chest imaging at our institution, an NHS trust covering a population of 500,000. Three separate patient cohorts were included in the study to address the separate aims of measuring diagnostic accuracy, consistency, and clinical impact. Recorded demographic data included patient age and sex.

Inclusion criteria for the first cohort, evaluating the diagnostic accuracy of an AI tool, were consecutive ECG-gated CT Aorta studies reported by an expert reader between February 2018 and August 2021. The study period selected was to ensure contemporaneous imaging protocols and make use of existing expert reports. Exclusion criteria were: (1) the expert reader reported artefact significantly impairing accurate measurement (excluded at the relevant anatomical level if focal, or completely if severe), (2) aortic valve replacement (AVR) or transcatheter aortic valve implantation (TAVI), and (3) a dissection was reported. Both AVR/TAVI and dissections were excluded at the anatomical level of the abnormality rather than excluding the entire study.

For the second cohort, evaluating the reliability of the AI tool, inclusion criteria were non-ECG-gated non-contrast and ECG-gated contrast phases of the same acquisition CT angiogram aorta

undertaken between February 2019 and July 2021. Exclusion criteria consisted of: (1) significant artefact affecting accurate measurements (excluded at the relevant anatomical level), and (2) failure of the AI analysis affecting at least one phase of the acquisition. To test the AI tool's performance on the non-ECG-gated non-contrast imaging, additional analysis was undertaken comparing its accuracy with an expert reader on the pre-contrast acquisition of these studies.

The final cohort assessed the potential clinical impact of the AI tool. Inclusion criteria consisted of sequential thoracic CT studies (undertaken April-May 2019) until ≥ 48 studies were included for each of CT thorax, CT pulmonary angiogram, high-resolution CT (HRCT) and acute ECG-gated CT Aorta. These studies were selected as the most notable potential clinical utility of the AI tool may be in scans not routinely reported by expert cardiothoracic readers. Studies failing AI analysis were excluded.

CT acquisition

All CT scans were obtained using routine acquisition parameters, automated tube voltage and current settings on either Siemens Definition Edge or Drive CT machines (Siemens Healthineers, Erlangen, Germany) with suspended respiration from lung apices to bases. Standard CT protocols were followed and are outlined in the supplementary materials.

AI Tool

The thinnest available axial reconstructed images were sent for fully automated analysis by an AI-derived post-processing software (AI-Rad Companion Chest CT, Siemens).

Training and validation of the software was conducted by Siemens and the following details of the software are external to our study. The Siemens software was trained on over 1000 chest CT scans including contrast and non-contrast scans and was further validated with 193 different chest CT data-sets[16]. All data-sets used for this were external to our institution.

Deep Reinforcement Learning is used in combination with scale-space analysis to detect the aortic landmarks automatically[17]. The aortic root is defined as the region of interest (ROI) for the segmentation algorithm[16]. An adversarial Deep Image-to-Image Network is used to perform segmentation in a symmetric convolutional encoder-decoder architecture from the

ROI[18] . The front part is a convolutional encoder-decoder network, and the backend a deep supervision network. Blocks inside consist of convolutional and upscaling layers[16]. A centreline model generates the aortic centreline, which is used in combination with aortic landmarks to identify measurement planes at each anatomical location. In each measurement plane, multiple diameters are computed and the maximum in-plane diameter provided[16]. As an output, an image of the aorta accompanies the measurements for ease of reviewing results. Measurements are colour coded in a traffic-light fashion to highlight abnormalities to the interpreter in a table accompanying the figure, or left out at anatomical levels where measurement failed. Where the AI tool is unable to analyse a scan at any anatomical level a report is returned stating “aortic landmarks not found”.

The frequency that AI analysis failed completely, processed the analysis but created a random output and processed the analysis with inaccurate segmentation were recorded. The frequency of partial (segmentation considered visually appropriate and >1 aortic measurement provided) and full reports were recorded.

Diagnostic accuracy

Dedicated ECG-gated CT aorta acquisitions reported by three expert cardiothoracic radiologists were considered the most appropriate reference standard to compare AI segmentation of the aorta. The expert read was considered the ground truth. Expert cardiothoracic radiologists (a Cardiothoracic fellowship Consultant Radiologist with >10 years’ experience) reported these scans from multiplanar reconstructions in a double oblique view generated using Syngo.via (Siemens Healthineers, Germany). Widely reported anatomical levels of the thoracic aorta were measured [3]: the sinus of Valsalva (SoV, measured cusp-to-cusp), sino-tubular junction (STJ), mid ascending aorta, mid arch, isthmus, mid descending aorta and hiatus (level of the diaphragm).

Measurements were compared on a per anatomical landmark basis between expert reader (using the largest diameter reported) and AI analysis for all included studies. Interclass correlation was used to measure the reliability between expert and AI readings. Single rater,

two-way mixed effects model was used to define the absolute agreement between AI and manual reporters.

The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for the AI tool's ability to detect a dilated aorta was measured versus the gold standard expert reader. This was measured both in the ascending and descending aorta separately (defined as $\geq 40\text{mm}$ ascending aorta / $\geq 30\text{mm}$ descending aorta[2–4]) and across the thoracic aorta as a whole on a per-patient basis.

Repeated measures

Acute CT aorta studies reported by expert readers provided an opportunity to test for a repeated measures assessment across separate phases of the same acquisition protocol whilst using routine clinically acquired imaging. The non-ECG-gated non-contrast and contrast ECG-gated phases of the same CT angiogram acquisition for 29 patient's CTs were sent for AI analysis. Measurements were also taken on both phases by the same expert cardiothoracic reader on two separate occasions >4 weeks apart. Repeated measures performance of the AI tool and expert cardiothoracic reader were compared. The diagnostic accuracy of the AI tool in non-ECG-gated non-contrast imaging was also assessed against the expert read, with analysis matching that outlined in the diagnostic accuracy section above.

Clinical Impact

Consecutive chest CTs reported by unselected general consultant radiologists at the time of the scan (April–May 2021) were reviewed and sent for AI analysis. Studies were selected to represent a real-life cohort with contemporary imaging protocols to reflect the modern practice of non-expert thoracic readers. This included 50 CT thorax, 50 CT pulmonary angiograms (CTPA), 49 high-resolution CT chests (HRCT) and 48 ECG-gated CT aortogram studies (acute scans reported by non-cardiothoracic radiologists).

The manual general radiologists report and electronic clinical records for each patient were reviewed for pre-existing documented aortic dilatation. The proportion of patients where AI analysis would have newly identified a dilated thoracic aorta was recorded.

Statistical analysis

Statistical analysis was performed using SPSS v.21 (IBM Corp Armonk, NY, USA). Normality was assessed through visual analysis of histograms and the Shapiro-Wilk test. Categorical data are presented as frequency and percentage, and continuous data as mean \pm standard deviation or median with interquartile range (IQR). Comparative testing was performed using paired independent t-tests, where appropriate. Reliability was assessed with Bland-Altman plots and via the intra-class correlation coefficient (ICC), with <0.5 indicating poor reliability, $0.5-0.75$ moderate, $0.75-0.9$ good, and >0.9 excellent reliability[19]. ICC estimates and their 95% confidence intervals were calculated based on a single measures rating, absolute agreement and two-way mixed effects model. For Bland-Altman analysis, the maximum acceptable limits of agreement were set at ± 5 mm based on prior assessment of expert reader inter-observer variability[23]. Clinical impact was assessed as change in diagnosis of aortic dilatation with a McNemar test. Statistical significance was defined as a two-tailed p value ≤ 0.05 .

Ethics

This is a retrospective service evaluation approved by our institution's Trust Audit Committee, waiving the requirement for formal written consent in line with the Health Research Authority decision tool[20].

Results

Across all three patient cohorts a full report including an output for all anatomical levels was produced in 387/436 (89%). There were 14 scans (3%) with no analysis possible, 1 scan ($<1\%$) failing manual quality assurance (QA) check for spurious anatomical locations selected, and 34 (8%) with a partial result only (i.e. ≥ 1 anatomical level not measured). Of scans with partial AI reports produced, 30/34 (88%) had only the hiatus measurement unavailable. The AI tool produced a measurement at ≥ 1 anatomical level that passed manual QA check in 422/436 (96.8%).

Demographics across the three study cohorts included are presented in Table 1.

Diagnostic accuracy

Of the 195 included scans, 1317/1365 (96%) individual anatomical measurements were analysed after manual QA check with 48 exclusions at focal levels for dissection, incomplete analysis or inaccurate segmentation (Figure 1).

The absolute difference, proportionate difference and ICC between the expert reader and AI assessment of the aorta dimensions at each anatomical level on ECG-gated contrast imaging was at least 'good' throughout (Table 2). AI analysis was most accurate relative to the expert reader in the mid ascending aorta. Table 2 demonstrates the mean absolute and median proportionate differences in aortic measurements at each anatomical level, which was maximally 2.6mm (± 2.1 mm) at the level of the SoV. On Bland-Altman analysis, measurements crossed the threshold for maximally accepted limits of agreement (>5 mm) at the level of the SoV and STJ (Figure 2). Further detailed comparisons of expert reader and AI measurements are available in supplementary table S1.

Across the 195 studies analysed, relative to the expert reader reference standard the AI analysis would have missed 14 patients (7%) with a diameter ≥ 40 mm in the ascending aorta (10 at the level of the SoV, 4 at the mid ascending aorta and 2 at both), and 10 cases with a diameter ≥ 30 mm in the descending aorta (5 at the level of the isthmus, 4 at the level of the mid descending aorta and 1 at both). Of those with a dilated ascending thoracic aorta missed by the AI tool, 6/14 (43%) had a maximum diameter of 40mm reported by the expert reader. The maximum diameter reported by the expert reader that was missed by the AI tool was 43mm in the ascending aorta and 36mm in the descending (at the level of the isthmus). The mean difference in these scans between AI and expert reader was 0.9mm. Additionally, 5/14 of those with a missed dilated ascending aorta (including both measuring 43mm) would have been flagged as abnormal based on descending aorta measurements of ≥ 30 mm. The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the AI tool at detecting aortic dilatation are presented in Table 3.

Figure 1. Study flowchart for (A) Diagnostic accuracy, (B) repeated measures analysis and (C) clinical impact.

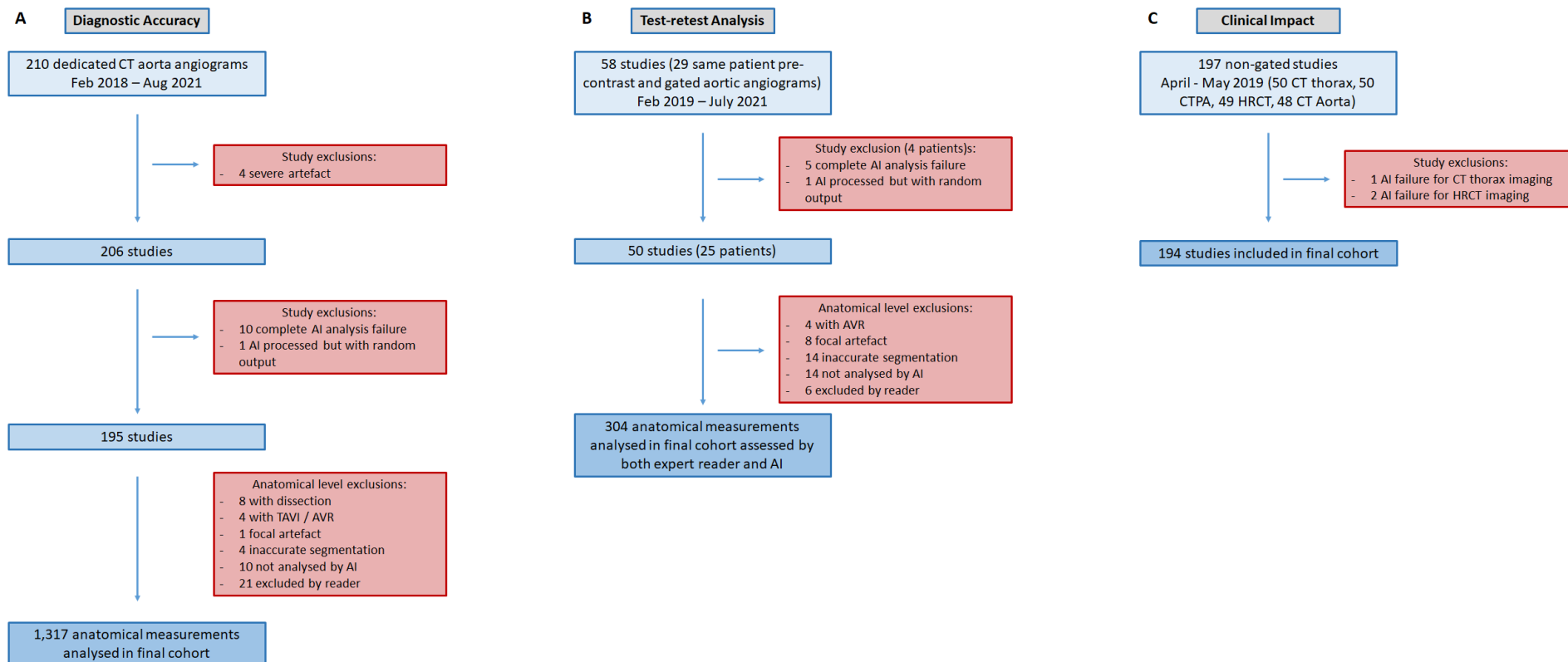


Table 1. Demographics of the three study cohorts.

Patient cohort	Age (mean [s.d.])	Sex (n [%] male)
Diagnostic accuracy (n=210)	75 [± 13]	127 [60]
Repeat measures (n=30)	63 [± 17]	20 [67]
Clinical impact		
HRCT (n=49)	66 [± 13]	25 [51]
CT chest (n=50)	67 [± 13]	31 [62]
CTPA (n=50)	63 [± 17]	27 [54]
CT aorta (n=48)	67 [± 15]	25 [52]

Table 2. Comparison of expert reader and AI measurements at each thoracic anatomical aortic level analysed on ECG-gated contrast imaging, assessed against absolute difference (mm), proportionate difference (%) and the average of each anatomical levels single measurer intra-class correlation coefficient (ICC) with its associated rating.

Anatomical location	Absolute difference (mm; mean [\pm s.d])	Proportionate difference (%; median [IQR])	ICC (95% CI, p)	ICC Rating
SoV	2.6 [2.1]	6 [3 – 11]	0.78 (0.67-0.85, p<0.001)	Good
STJ	1.9 [1.9]	4 [3 – 8]	0.83 (0.74-0.88, p<0.001)	Good
Mid Asc	1.3 [1.3]	3 [0 – 6]	0.92 (0.89-0.94, p<0.001)	Excellent
Mid Arch	1.7 [1.7]	4 [3 – 8]	0.76 (0.70-0.82, p<0.001)	Good
Isthmus	1.7 [1.7]	4 [3 – 8]	0.81 (0.76-0.86, p<0.001)	Good
Mid Desc	1.3 [1.4]	4 [0 – 8]	0.88 (0.78-0.92, p<0.001)	Good
Hiatus	1.4 [1.6]	4 [0 – 8]	0.81 (0.76-0.86, p<0.001)	Good

Figure 2. Bland-Altman analysis comparing expert reader with AI analysis at each anatomical level for ECG-gated contrast imaging (SoV = sinus of Valsalva; STJ = sinotubular junction; MidAsc = mid ascending aorta; MidArch = mid aortic arch; MidDesc = mid descending aorta).

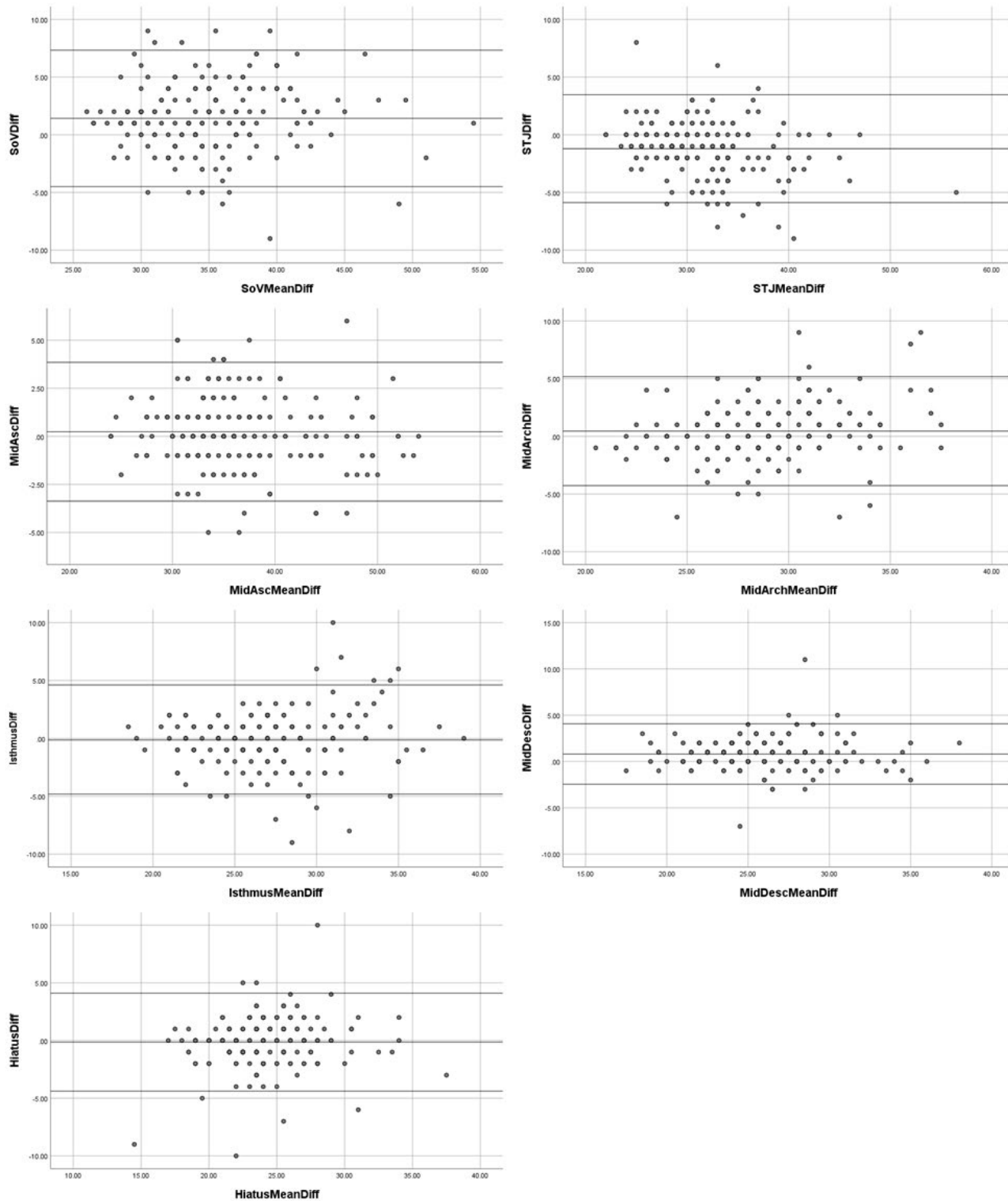


Table 3. The AI-Rad Companion sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for detecting aortic dilatation versus expert cardiothoracic reader on ECG-gated contrast and non-ECG-gated non-contrast imaging.

		Sensitivity	Specificity	PPV	NPV
ECG-gated contrast imaging	Thoracic	77	99	98	86
	Aorta	(66 – 85%)	(95 – 100)	(90 – 100)	(80 – 90)
	Ascending	74	98	96	87
	Aorta	(62 – 84)	(94 – 100)	(87 – 99)	(82 – 91)
	Descending	68	98	85	94
	Aorta	(66 – 83)	(94 – 99)	(68 – 94)	(90 – 96)
Non-ECG-gated non-contrast imaging	Thoracic	100	77	67	100
	Aorta	(63 – 100)	(50 – 93)	(46 – 83)	-
	Ascending	90	92.3	90	92
	Aorta	(56 – 100)	(64.0 – 99.8)	(58 – 98)	(65 – 99)
	Descending	100	59	50	100
	Aorta	(59 – 100)	(33 – 82)	(36 – 64)	-

Repeated measures

After exclusions (Figure 1), 25 non-ECG-gated pre-contrast thorax and ECG-gated CT aortogram acquisitions from the same patient scan had both AI and expert radiologist assessment. 16 (64%) were male, mean age of 61 ± 17 years. Repeated measures assessment by both expert reader and AI analysis is presented in Table 4.

Diagnostic accuracy was measured on the non-ECG-gated pre-contrast acquisitions, with the mean absolute difference varying from 1.2 to 4.3mm across all anatomical levels (Table S2). The ICC between the expert reader and AI assessment were graded as moderate at all locations except the mid ascending and isthmus where it was graded as good. The Bland-Altman limits of agreement crossed the pre-set threshold of 5mm at all anatomical levels (Table S2 and Figure S1). When testing accuracy for the detection of aortic dilatation vs expert reader across the full thoracic aorta, sensitivity was 100% and specificity 76.5% (Table 3).

Table 4. Comparison of expert reader versus AI reliability as assessed with the intra-class correlation coefficient (ICC) for repeated measures performance using the single measures rating, absolute agreement and two-way mixed effects model.

Anatomical Location	Expert Reader		AI Analysis	
	ICC (95% CI, p)	Rating	ICC (95% CI, p)	Rating
SoV	0.89 (0.71-0.96, $p < 0.001$)	Good	0.78 (0.53-0.91, $p < 0.001$)	Good
STJ	0.78 (0.57-0.90, $p < 0.001$)	Good	0.69 (0.40-0.85, $p < 0.001$)	Moderate
Mid Asc	0.96 (0.92-0.98, $p < 0.001$)	Excellent	0.88 (0.74-0.95, $p < 0.001$)	Good
Mid Arch	0.79 (0.57-0.91, $p < 0.001$)	Good	0.74 (0.49-0.88, $p < 0.001$)	Moderate
Isthmus	0.87 (0.69-0.94, $p < 0.001$)	Good	0.79 (0.58-0.91, $p < 0.001$)	Good
Mid Desc	0.84 (0.68-0.93, $p < 0.001$)	Good	0.66 (0.35-0.84, $p < 0.001$)	Moderate
Hiatus	0.84 (0.66-0.93, $p < 0.001$)	Good	0.57 (0.22-0.79, $p = 0.002$)	Moderate

Clinical impact

Figure 3 presents the potential impact of AI aortic measurement on detecting new thoracic aorta dilatation on scans reported by non-expert cardiothoracic readers, sub-divided by acquisition type, with AI success rate. Across all scans successfully analysed, the AI tool newly identified aortic dilatation in 53 (27%) patients where it was not previously documented in the baseline clinical report or electronic patient records (X^2 51, $p < 0.001$; Figure 4). This includes dilatation of up to 47mm in the ascending aorta and 37mm in the descending aorta in patients without a known aneurysm.

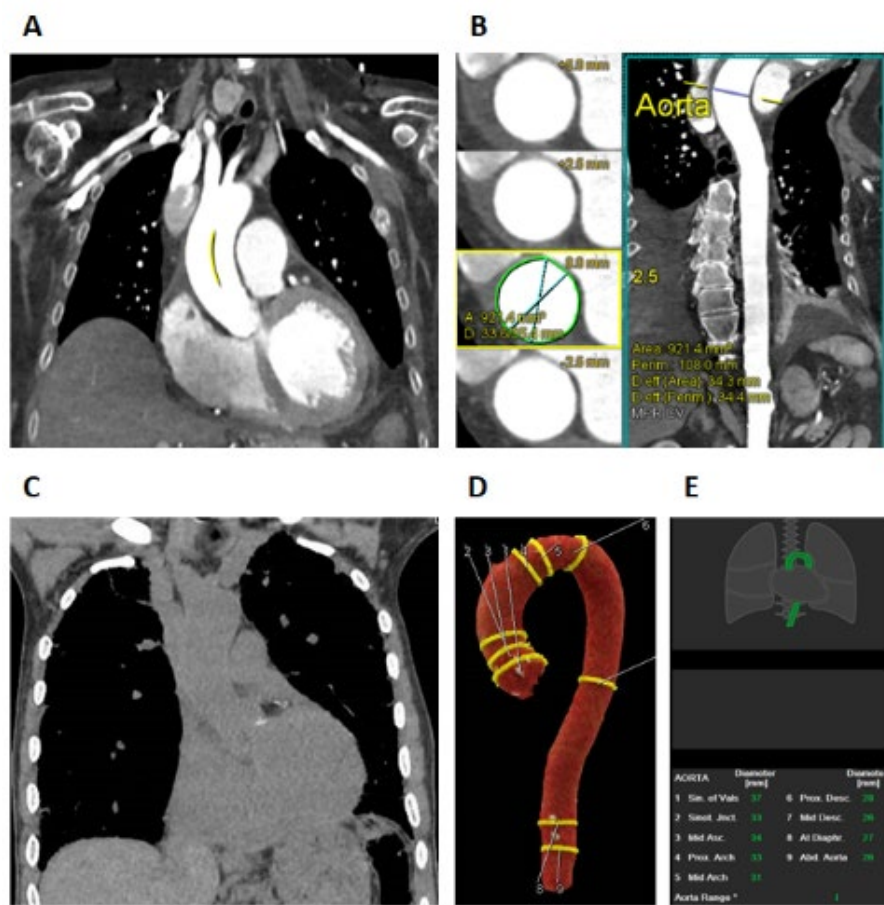
Figure 3. Detection of new aortic dilatation via AI analysis of non-ECG gated chest CT imaging sent for AI analysis sub-divided by acquisition type.



Figure 4. Detection of new aortic dilatation with AI analysis vs known aortic aneurysm or routine clinical reports by non-expert readers of non-dedicated aorta thoracic imaging.

	AI Analysis		
Known or Clinically Reported Aortic Dilatation	n = 194	Aortic dilatation	Normal
	Aortic dilatation	12	0
	Normal	53	129

Figure 5: Example analyses of expert reader and AI aortic measurements, including (A) ECG-gated CT aorta (coronal plane), (B) measured manually in double oblique view with *Syngo.Via* (Siemens Healthineers), (C) mediastinal window of a high-resolution CT chest (HRCT), with (D) and (E) AI analysis and report of aorta assessment of HRCT in (C).



Discussion

This is the first study to compare expert cardiothoracic readers with an AI software segmentation of the aorta for accuracy and potential clinical impact in a UK population and across all chest CT imaging. Its accuracy varied across anatomical levels, however it demonstrates that the Siemens AI-Rad Companion Chest CT segmentation software has the potential to newly identify a dilated aorta on routine thoracic imaging with a high degree of specificity (99.1% on a per-patient basis) but with sub-optimal sensitivity (76.5%).

The measurements from the AI performed modestly against expert manual segmentation (imaging example of output provided in Figure 5). Comparative measurements between AI and expert reader for the thoracic aorta demonstrated good to excellent interclass correlation (Table 2). On ECG-gated contrast imaging mean absolute differences in AI measurements vs expert reader varied from 1.3mm at the mid ascending and descending aorta to 2.6mm at the SoV. The SoV also had the largest mean measurement variation from the use of AI-Rad Companion Chest CT in other studies. An example includes mean deviations at the SoV of up to 4.7mm when compared against manual radiologist measurements reported by Rueckel *et al.* [22]. Further, when radiology reporters were given the option to correct the AI anatomical level by Rueckel *et al.*, 50% of the AI measurements at the SoV were manually adjusted. This compares with 34% of all measurements manually adjusted across other anatomical levels, suggesting a systematic AI error in the assessment of the SoV. Rueckel *et al.* propose that the poorer analysis at the SoV is due to the geometrically complex shape of the proximal aorta, which can change with respiration, blood pressure and the bio-elasticity of the aorta[22]. This has been considered in other research outlining the difficulty in measuring the thoracic aorta, which considered the SoV a difficult measurement level due to its asymmetric structure[23].

Monti *et al.* demonstrated no significant difference in manual and automatic (Siemens, AI-Rad Companion Chest CT) maximum aortic diameter measurements[14]. However, they found a negative bias in automatic measurements of –1.5mm with a coefficient of reproducibility of 8mm concordance between AI and manual measurements[14]. In addition, Monti *et al.* and Pradella report lower accuracy at the SoV attributed to the aortic centreline not being

perpendicular to the aortic origin with a subsequently tilted measurement plane[14, 15]. This is consistent with the geometrical complexities considered by Quaint and Rueckel[22, 23]. This may explain the disproportionate number of cases measured at $\geq 40\text{mm}$ missed by the AI at the SoV when compared to manual reporting (10/14 [71%] of all dilated ascending aortas missed). These 14 patients represent the 7% of scans where the AI tool would have missed a dilated ascending aorta defined by a $\geq 40\text{mm}$ threshold. Pradella *et al.* found similar results for AI misclassified aneurysms at 8.5%, though this is unclear if aneurysms were over- or under-reported[15]. In our study the mean difference in AI report vs expert reader was low (0.9mm), whilst in 6/14 (43%) cases missed by the AI tool the maximum diameter reported by the expert reader was at the pre-defined level for dilated (40mm). Additionally, 5/14 (36%) would have been highlighted as abnormal by the AI tool based on descending aorta measurements of $\geq 30\text{mm}$. The maximum diameter missed in the ascending aorta for the diagnostic accuracy cohort was 43mm and in both of these cases would again have been flagged to the reporter based on a dilated descending aorta.

We demonstrate that the AI was most accurate in ECG-gated contrast imaging at the mid-ascending level. We found that the AI performed better in the ascending aorta over the descending. This is reflected in a study by Macruz *et al.* which reports an ICC greater than 0.8 for the ascending aorta and 0.7 for the descending[24].

Absolute differences of up to 2.6mm were observed in our study across all anatomical levels on ECG-gated contrast CTs. It has been argued that CT measured increases of up to 4mm may not always represent pathological dilation of the aorta. Discrepancies may result from multiple factors affecting measurement such as image resolution, type of cursor, inclusion of the aortic wall on non-contrast studies and motion artefact[25, 26]. These will all lead to inter-scan, inter-observer and intra-observer variability. Quint *et al.* reviewed variability between manual segmentation using double oblique views and though there were mean differences of no more than 1.3mm at each segmentation level, the 95% confidence interval was up to 5.1mm [23]. Frazao *et al.* found CT and MRI provide larger estimates of aortic root measurements when compared to transthoracic echocardiography assessment (mean difference $4.9 \pm 2.7\text{mm}$)[27].

However, for ECG-gated contrast studies the AI tool crossed the pre-defined maximum acceptable limits of agreement ($>5\text{mm}$) at the SoV and STJ. In non-ECG-gated non-contrast studies it crossed this threshold at all levels, though interpretation of this finding is limited by sample size. On this basis, this tool would be considered insufficiently accurate for use in routine surveillance imaging to track progressive change in known aortic dilatation. In such cases subtle increases in size may impact major clinical management decisions. There is of course also a sliding scale in severity of dilatation, as evidenced by the mortality rate that rises with increasing dilatation. This is represented by the diagnostic overlap of diameter thresholds, e.g. 40mm for a 'dilated' mid ascending aorta vs 50mm for it to be considered 'aneurysmal' at the same anatomical location.

It could be argued though that failure to meet the limits of agreement for accuracy at all anatomical levels may be interpreted in a wider context. For example, the tool may still address the proposed goal of identifying unknown dilated aortas on a per-patient basis on routine non-dedicated aorta chest CTs reported by non-expert readers. To this end, the AI tool's specificity and sensitivity would suggest it may provide a simple and pragmatic way to deliver this.

Automated assessment of the thoracic aorta across all chest imaging may identify a significant proportion of patients (27% in our population) with previously undiagnosed aortic dilatation. This is a patient group where the aorta is typically not routinely assessed in detail or reported by non-cardiothoracic radiologists. Measuring the aorta accurately takes a considerable amount of time using specific software to achieve double oblique views. As such, it demonstrates the role software may have in screening for thoracic aortic aneurysms in routine CT chest imaging, whilst not necessarily being sufficiently accurate for subsequent surveillance imaging.

AI not only highlights the shortcomings in manual reporting of routine chest CT studies, but offers a solution to the outlay in reporting time it would require. The incidence of aortic aneurysm is considered to be underreported due to the indolent nature of the disease. Our findings may be relevant in both demonstrating the opportunity to newly diagnose a proportion of patients with a dilated aorta, treat patients before potentially fatal complications, and to better assess its prevalence within the population. This includes the identification of a population at increased risk of other future MACE[2], likely due to shared cardiovascular risk

factors and may enable more aggressive, targeted primary prevention. This surreptitious screening from data already acquired for other purposes may have different cost implications versus a dedicated screening programme. Whether its use in this fashion translates into improved clinical outcomes or cost-effectiveness requires further investigation. This would need to ensure it factored in both the AI failure rate and requirement for a semi-automated QA assessment of the of AI output by the reporting radiologist. If this iteration were introduced into clinical practice, it may be appropriate to manually review scans with an ascending aorta measurement of 35-40mm to ensure normality.

Our identification of 27% of patients with new aortic dilatation is comparable with others. Monti *et al.* found 31% of their scans showing aneurysms, though the exact mix of heterogeneous scans (ECG-gated aorta vs non-ECG gated) is not outlined. Given the low number of their scans (29%) that demonstrated no pathology, their findings do not reflect the expected diagnosis rate of new disease in screening chest CT and raises the question of a higher pre-test probability in their cohort[14]. Similarly, Artzener *et al.* demonstrated 20.5% of patients with a dilated aorta, though their cohort contained only 50% random CT chest acquisitions with the other 50 selected for aortopathy and mediastinal pathology[21].

It is important to note that the moderate sensitivity on contrast imaging does highlight the potential to miss patients with a dilated aorta if relying solely on AI performance. In our cohort this would miss a thoracic aorta of no larger than 43mm, which occurred in two patients, both of whom would have been flagged by the AI report for a descending aorta of >30mm. Importantly, the detection rate for new diagnosis of a dilated aorta was significantly higher than that of the manual report by non-expert readers. The high specificity of the tool also highlights appropriate resource utilisation as there is a low likelihood of subsequent surveillance imaging being unnecessary based on an error in the baseline AI report.

The AI software successfully returned a reading at ≥ 1 anatomical level in 96.8% of our scans, though full reports were present in 89%. The success rate of the AI's segmentation in our study is representative of its performance in the wider literature. Pradella *et al.* found that automatic software (Chest AI, version 0.2.9.2, Siemens Healthineers, Germany) returned a result in 94% of

all measurements in a study with 405 ECG-gated CT aorta studies[15], and Artzener *et al.* returned results in 99.2% of scan parameters [21]. 250 CT examinations underwent Siemens AI-Rad Companion Chest CT software analysis by Monti *et al.* with 93% of scans adequate for analysis and successful transfer. Monti *et al.* found that the software performed poorly with pathological aortas. In particular this included dissection[14], which were excluded in our analysis. Artzener *et al.* included aortic dissections within their cohort of patients and found the true lumen was only measured if the false lumen was thrombosed, otherwise the true and false lumen would be measured as a single diameter[21]

The exclusion of imaging with significant artefact in our study may provide an over-estimation of diagnostic accuracy performance. However, artefact also limits manual reporting and as such may be considered a non-diagnostic scan. Additionally, many AI tools currently embedded into clinical practice require a degree of manual QA in combination with AI outputs to reduce the potential for inaccurate reporting. How the relationship between AI and manual reporting differs over severity of artefact would be grounds for further research.

An interpretation of AI tool reliability was assessed across a pre- and post-contrast acquisition taken during the same patient study. The AI tool was not as reliable as the expert reader, with lower ICC ratings. The expert reader was more reliable at all anatomical landmarks apart from matching the AI at the SoV and isthmus. As discussed above, the comparison of images with and without contrast may lead to the inclusion of the aorta wall on non-contrast scans but not in contrast scans[26]. Expert readers may provide a better judgement over the AI of measuring to the aorta wall.

Other studies have assessed the reproducibility of automated thoracic segmentation software. Though Pradella *et al.* achieved near perfect reproducibility, their method for assessment was to send a single phase through the AI software twice[15]. This may not truly represent the reproducibility of an AI tool. We approached this evaluation in an alternative fashion in our study with separate contrast and non-contrast sequences of the same aorta study sent for AI segmentation. Whilst the ICC in our study is lower and analysed from a smaller patient cohort of 25 patients (50 studies), it likely reflects variation in the approach taken to testing

reproducibility of the software. However, it is important to recognise the limitations of our approach to reproducibility analysis. Our approach compared a non-ECG-gated with an ECG-gated acquisition. As a result it would be reasonable to expect some variation in measurements based on changes in aortic dimension throughout the cardiac cycle. Ideally patients would have undergone sequential ECG-gated acquisitions getting on and off the scanner between each scan to enable true test-retest analysis. However, in the context of our retrospective validation study that attempted to use routinely acquired clinical imaging this provided as close as we could achieve to a meaningful assessment of repeated measures testing.

This study was limited by its retrospective and single-centre design, though this was off-set to some degree by the use of consecutive imaging acquired. Further limitations include the statistical analysis being undertaken unblinded to imaging and the sample size of the repeated measures and diagnostic accuracy in non-ECG-gated non-contrast imaging cohort was low. Further, in our repeated measures analysis the manual reporting of the sequential ECG-gated non-contrast and contrast CT Aorta acquisitions was completed by a single reader. This may lead to observer bias, though a period of >4 weeks elapsed between reporting the non-contrast and contrast scans and was performed blinded to the baseline measurements to minimise this effect.

The exclusion of aortas with dissection or prior valve surgery may partially limit the hypothetical benefit of our findings to the opportunistic screening of all patients. Further research evaluating the AI tool in these patients would be required to evaluate the AI for its role in monitoring patients post intervention, having been beyond the scope of this study. Additionally, the AI tool was analysed in CTs acquired using a single vendor (Siemens) and may require validation using other CT systems to ensure the generalisability of our findings.

Conclusions

In conclusion, AI was feasible in the majority of scans and had good agreement with exception of the root. The high specificity for detecting aortic dilatation suggests the tool may have a role in opportunistic screening for aortic dilatation in CT chests performed for other clinical

indications. Further work is require to determine if the tool's sensitivity is acceptable in a cost-effective analysis.

References

1. Mathur A, Mohan V, Ameta D, *et al.* (2016) Aortic aneurysm. *J Transl Int Med* 4:35–41. <https://doi.org/10.1515/jtim-2016-0008>
2. Erbel R, Aboyans V, Boileau C, *et al.* (2014) 2014 ESC guidelines on the diagnosis and treatment of aortic diseases. *Eur Heart J* 35:2873–2926. <https://doi.org/10.1093/eurheartj/ehu281>
3. Hiratzka LF, Bakris GL, Beckman JA, *et al.* (2010) 2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM guidelines for the diagnosis and management of patients with thoracic aortic disease: Executive summary: A report of the american college of cardiology foundation/american heart association task force on pra. *Circulation* 121:266–369. <https://doi.org/10.1161/CIR.0b013e3181d4739e>
4. Hager A, Kaemmerer H, Rapp-Bernhardt U, *et al.* (2002) Diameters of the thoracic aorta throughout life as measured with helical computed tomography. *Journal of Thoracic and Cardiovascular Surgery* 123:1060–1066. <https://doi.org/10.1067/mtc.2002.122310>
5. Sharples L, Sastry P, Freeman C, *et al.* (2022) Aneurysm growth, survival, and quality of life in untreated thoracic aortic aneurysms: the effective treatments for thoracic aortic aneurysms study. *Eur Heart J* 43:2356–2369
6. Erbel R, Aboyans V, Boileau C, *et al.* (2014) 2014 ESC guidelines on the diagnosis and treatment of aortic diseases. *Eur Heart J* 35:2873–2926
7. Gouveia e Melo R, Silva Duarte G, Lopes A, *et al.* (2022) Incidence and Prevalence of Thoracic Aortic Aneurysms: A Systematic Review and Meta-analysis of Population-Based Studies. *Semin Thorac Cardiovasc Surg* 34:1–16. <https://doi.org/10.1053/J.SEMTCVS.2021.02.029>
8. Erbel R, Eggebrecht H (2006) Aortic dimensions and the risk of dissection. *Heart* 92:137–142
9. Wang TKM, Desai MY (2020) Thoracic aortic aneurysm: Optimal surveillance and treatment. *Cleve Clin J Med* 87:557–568. <https://doi.org/10.3949/CCJM.87A.19140-1>
10. Melvinsdottir IH, Lund SH, Agnarsson BA, *et al.* (2016) The incidence and mortality of acute thoracic aortic dissection: Results from a whole nation study. *European Journal of Cardio-thoracic Surgery* 50:1111–1117. <https://doi.org/10.1093/ejcts/ezw235>
11. Sastry P, Hughes V, Hayes P, *et al.* (2015) The ETAA study protocol: a UK-wide observational study of “Effective Treatments for Thoracic Aortic Aneurysm.” *BMJ Open* 5:8147. <https://doi.org/10.1136/bmjopen-2015>
12. Sedghi Gamechi Z, Bons LR, Giordano M, *et al.* (2019) Automated 3D segmentation and diameter measurement of the thoracic aorta on non-contrast enhanced CT. *Eur Radiol* 29:4613–4623. <https://doi.org/10.1007/s00330-018-5931-z>
13. Diagnostic imaging Dataset Annual Statistical Release 2020/21, NHS England and NHS Improvement, Published 18th November 2021. Page 20. Version 1.0. Accessed 13 Apr 2022 <https://www.england.nhs.uk/statistics/statistical-work-areas/diagnostic-imaging-dataset/diagnostic-imaging-dataset-2020-21-data/>

14. Monti CB, van Assen M, Stillman AE, *et al.* (2022) Evaluating the Performance of a Convolutional Neural Network Algorithm for Measuring Thoracic Aortic Diameters in a Heterogeneous Population. *Radiol Artif Intell* 4:. <https://doi.org/10.1148/ryai.210196>
15. Pradella M, Weikert T, Sperl JJ, *et al.* (2021) Fully automated guideline-compliant diameter measurements of the thoracic aorta on ECG-gated CT angiography using deep learning. *Quant Imaging Med Surg* 11:4245–4257. <https://doi.org/10.21037/qims-21-142>
16. Siemens Healthineers (2021) Features, Data, and Algorithms AI-Rad Companion Chest CT VA13 Whitepaper, Accessed 25 May 2022. <https://cdn0.scrvt.com/39b415fb07de4d9656c7b516d8e2d907/3c06d6256fff0b5d/c01241ad9c70/DH-AI-Rad-Companion-Chest-CT-Whitepaper--2-.PDF>
17. Ghesu FC, Georgescu B, Zheng Y, *et al.* IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 1 Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans
18. Yang D, Xu D, Zhou SK, *et al.* (2017) Automatic Liver Segmentation Using an Adversarial Image-to-Image Network
19. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 15:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
20. NHS Health Research Authority (2020) Decision tools - is my study research? <http://www.hra-decisiontools.org.uk/research/>. Accessed 20 Sep 2021
21. Artzner C, Bongers MN, Kärgel R, *et al.* (2022) Assessing the Accuracy of an Artificial Intelligence-Based Segmentation Algorithm for the Thoracic Aorta in Computed Tomography Applications. *Diagnostics* 12:1790. <https://doi.org/10.3390/diagnostics12081790>
22. Rueckel J, Reidler P, Fink N, *et al.* (2021) Artificial intelligence assistance improves reporting efficiency of thoracic aortic aneurysm CT follow-up. *Eur J Radiol* 134:. <https://doi.org/10.1016/j.ejrad.2020.109424>
23. Quint LE, Liu PS, Booher AM, *et al.* (2013) Proximal thoracic aortic diameter measurements at CT: Repeatability and reproducibility according to measurement method. *International Journal of Cardiovascular Imaging* 29:479–488. <https://doi.org/10.1007/s10554-012-0102-9>
24. Macruz FB de C, Lu C, Strout J, *et al.* (2022) Quantification of the Thoracic Aorta and Detection of Aneurysm at CT: Development and Validation of a Fully Automatic Methodology. *Radiol Artif Intell* 4:. <https://doi.org/10.1148/ryai.210076>
25. Elefteriades JA, Farkas EA (2010) Thoracic Aortic Aneurysm. Clinically Pertinent Controversies and Uncertainties. *J Am Coll Cardiol* 55:841–857
26. Elefteriades JA, Mukherjee SK, Mojibian H (2020) Discrepancies in Measurement of the Thoracic Aorta: JACC Review Topic of the Week. *J Am Coll Cardiol* 76:201–217
27. Frazao C, Tavoosi A, Wintersperger B, *et al.* (2020) Multimodality Assessment of Thoracic Aortic Dimensions. *J Thorac Imaging*

Figure legends

Table 1. Demographics of the three study cohorts.

Table 2. Comparison of expert reader and AI measurements at each thoracic anatomical aortic level analysed on ECG-gated contrast imaging, assessed against absolute difference (mm), proportionate difference (%) and the average of each anatomical levels single measurer intra-class correlation coefficient (ICC) with its associated rating.

Table 3. The AI-Rad Companion sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for detecting aortic dilatation versus expert cardiothoracic reader on ECG-gated contrast and non-ECG-gated non-contrast imaging.

Table 4. Comparison of expert reader versus AI reliability as assessed with the intra-class correlation coefficient (ICC) for repeated measures performance using the single measures rating, absolute agreement and two-way mixed effects model.