

Article

Not peer-reviewed version

Enhancing Radiologist Efficiency with AI: A Multi-Reader Multi-Case Study on Aortic Dissection Detection and Prioritisation

[Martina Cotena PhD](#)*, [Angela Ayobi](#), Colin Zuchowski, Jacqueline C. Junn, [Brent D. Weinberg](#), [Peter D. Chang](#), [Daniel S. Chow](#), Jennifer E. Soun, [Mar Roca-Sogorb](#), [Yasmina Chaibi](#), [Sarah Quenet](#)

Posted Date: 16 October 2024

doi: 10.20944/preprints202410.1263.v1

Keywords: aortic dissection; automated detection; deep learning; prioritised worklist; emergency radiology; multi-reader multi-case study



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Enhancing Radiologist Efficiency with AI: A Multi-Reader Multi-Case Study on Aortic Dissection Detection and Prioritisation

Martina Cotena PhD ^{1,*}, Angela Ayobi MEng ¹, Colin Zuchowski MD ², Jacqueline C. Junn MD ², Brent D. Weinberg MD ², Peter D. Chang MD ^{3,4}, Daniel S. Chow MD ^{3,4}, Jennifer E. Soun MD ^{3,4}, Mar Roca-Sogorb PhD ¹, Yasmina Chaibi PhD ¹ and Sarah Quenet MEng ¹

¹ Avicenna.AI, 375 Avenue du Mistral, 13600 La Ciotat, France.

² Department of Radiology and Imaging Sciences, Emory University School of Medicine, 1364 Clifton Road Northeast, Suite BG20, Atlanta, GA 30322, USA.

³ Department of Radiological Sciences, University of California Irvine, Irvine, CA 92697, USA.

⁴ Center for Artificial Intelligence in Diagnostic Medicine, University of California Irvine, Irvine, CA 92697, USA.

* Correspondence: martina.cotena@avicenna.ai

Abstract: *Background and Objectives:* Acute aortic dissection (AD) is a life-threatening condition and early detection can significantly improve patient outcomes and survival. This study evaluates the clinical benefits of integrating a deep-learning (DL)-based application for automated detection and prioritisation of AD on chest CT angiographies (CTAs), focusing on the reduction of scan-to-assessment (STAT) and interpretation times (IT). *Materials and Methods:* This retrospective Multi-Reader, Multi-Case (MRMC) study compared AD detection with and without artificial intelligence-(AI) assistance. Ground truth was established by two U.S. board-certified radiologists, while three additional expert radiologists participated as readers. All participants assessed the same CTAs without AI-assistance (pre-AI arm) and, after a 1-month washout period, with the help of the device outputs (post-AI arm). STAT and IT were compared between the two phases. *Results:* The study included 285 CTAs (95 per reader, per arm), with a mean patient age of 58.5 years ± 14.7 (SD), 52% men, 37% prevalence. AI assistance significantly reduced STAT for detecting 33 true positive AD cases, from 15.84 minutes (95% CI: 13.37–18.31min) without AI to 5.07 minutes (95% CI: 4.23–5.91min) with AI, a 68% reduction ($p < 0.01$). IT also decreased significantly, from 21.22 seconds (95% CI: 19.87–22.58s) without AI to 14.17 seconds (95% CI: 13.39–14.95s) with AI ($p < 0.05$). *Conclusions:* Integrating a DL-based algorithm for AD detection on chest CTAs significantly reduces both STAT and IT. By prioritising the most urgent cases, AI ensures faster diagnosis and improves the workflow efficiency in clinical radiology practice, compared to standard First-In First-Out (FIFO) workflow.

Keywords: aortic dissection; automated detection; deep learning; prioritised worklist; emergency radiology; multi-reader multi-case study

1. Introduction

Aortic dissection (AD) is a severe thoracic aortic disorder and a cardiovascular emergency with a high mortality risk [1]. Its incidence is increasing, rising each year 15 cases per 100,000 patients [2–4]. If left untreated, acute AD has a 33% mortality rate within the first 24 hours, rising to 50% by 48 hours and up to 75% in undiagnosed cases of ascending AD [2,5–7]. Nearly 22% of patients die before reaching medical care [3,4]. Timely diagnosis through imaging and prompt patient management are crucial, as mortality in AD patients increases by 1–2% per hour in the first 24 hours [8].

Computed tomography angiography (CTA) is the gold standard for diagnosing suspected AD due to its non-invasiveness and ability to produce rapid, high-quality images [4]. However, the Royal College of Radiologists reports that radiologists face significant pressure from increasing workloads

and demands for efficiency, leading to fatigue, errors, and delays in diagnosis [9,10]. Some studies reported dire consequences of potential delays in treatment, often resulting from misdiagnosis, late diagnosis, or a low clinical index of suspicion in the Emergency Department [11,12].

Deep learning-based artificial intelligence (AI) systems are emerging in radiology, showing promise in neurology, cardiology, thoracic imaging, and cancer screening and detection [13–17]. AI tools can help identify AD features on CTA, reducing the risk of missed lesions [18]. New AI tools have been developed and validated to assist radiologists by prioritising AD positive cases, ensuring they receive urgent attention [19,20]. These AI systems effectively identify most dissections and all available aortic ruptures, placing critical cases at the top of radiologists' worklists. Prioritisation helps streamline the workflow, enabling timely and accurate diagnosis and treatment for patients needing immediate care [21]. This approach is a more effective alternative to the First-In, First-Out (FIFO) methodology, which has significant drawbacks, including its failure to consider case urgency or severity and potentially leading to increased risks in patient care [22].

There is limited literature on the advantages of computerised tools for AD detection on radiologist workflow and patient outcomes. A recent study highlighted that an automated tool not only exhibited good technical performance but also significantly decreased the time between study intake to radiologist read [21]. To elucidate the clinical benefits of integrating an automated tool for AD detection and prioritisation, we conducted a retrospective Multi-Reader Multi-Case (MRMC) study. This study aimed to assess the impact of the validated AI algorithm, CINA-CHEST for AD (Avicenna.AI, La Ciotat, France), on radiologists' efficiency and the time required to identify AD-positive cases. We simulated two clinical workflows for AD detection: a conventional FIFO approach without AI assistance and an AI-enhanced approach based on prioritisation. We hypothesised that the use of AI plays a crucial role in prioritising critical AD cases, improving the timeliness of prompt diagnosis.

2. Materials and Methods

2.1. Data Collection

The study was conducted in accordance with the 1975 Helsinki Declaration (as revised in 2013). Prior to investigator assessments, all data were anonymised in compliance with HIPAA and General Data Protection Regulation (GDPR) (EU) 2016/679. Informed consent was waived when it was deemed necessary, following national legislation and institutional protocols. Anonymised CTA cases were retrospectively provided by a large U.S. teleradiology network and acquired between July 2017 and December 2018 from multiple clinical sites across 90 cities in the U.S. The dataset included CTA images from 4 scanner vendors including 16 different scanner models: 5 GE Medical Systems, 2 Philips, 6 Siemens and 3 Canon/Toshiba. Among these patients, we identified the cohort based on the following criteria: 1) age ≥ 18 years old, 2) chest or thoraco-abdominal CTA scans only, 3) slice thickness ≤ 3 mm or less with no gap between successive slices, and 4) soft tissue reconstruction kernel. Exclusion criteria included: 1) not compatible with the recommended acquisition protocol, 2) thoracic aorta out of the field of view, and 3) significant acquisition artefacts impeding CTA interpretation.

2.2. The Ground Truth

Two U.S. board-certified expert radiologists (P.D.C. and D.S.C.), with 7- and 6-years of experience in clinical practice respectively, independently analysed CTAs to determine the presence of aortic dissections (ADs). In case of disagreement, a third U.S. board-certified expert radiologist (J.E.S.), with 8 years of experience in clinical practice, analysed the cases and the ground truth (presence or absence of AD, along with the classification of AD type) was determined by majority agreement. All classifications of AD—hyperacute, acute, subacute, or chronic—were considered positive. Additionally, the radiologists documented observed confounding factors, such as thoracic or abdominal aneurysms, intramural hematoma, calcifications, and post-surgery instances (e.g., presence of stents).

2.2.1. AI Algorithm for AD

An FDA-approved and CE-marked commercially available DL-powered application for AD, CINA-CHEST v1.0.3 (Avicenna.AI, La Ciotat, France), was utilised in this study. The application automatically processes CTA and displays notifications of suspected findings (if any) alongside image series information. For cases flagged as positive by the application, the type of AD (Type A or B) is displayed and localised with a red bounding box. These cases are subsequently prioritised in the worklist according to their classification as either positive or negative. A more detailed explanation of the deep learning algorithms and the integration process of CINA-CHEST for AD was provided in a recent study [19].

2.3. Multi-Reader Multi-Case (MRMC) Study

A retrospective, multi-center, fully-crossed MRMC study was conducted to evaluate the clinical efficacy of CINA-CHEST for AD on clinical workflow. This study included two phases, the pre-AI phase (Unaided Arm) in which radiologists detected AD without the aid of the software’s outputs and the post-AI phase (Aided Arm) in which radiologists detected AD assisted by the application outputs.

Three radiologists (J.C.J, B.W. and C.Z.; two U.S. board-certified radiologists and one fellow in general radiology) participated in the study and were different from the ones who established the ground truth, they had respectively 9, 5 and 2 years of experience in radiology clinical practice. The readers analysed every CTA twice, once unaided and once aided by the software’s outputs. A 1-month washout period between the two sessions mitigated recall bias. The study design and the radiologist reading workflow are summarised in Figure 1.

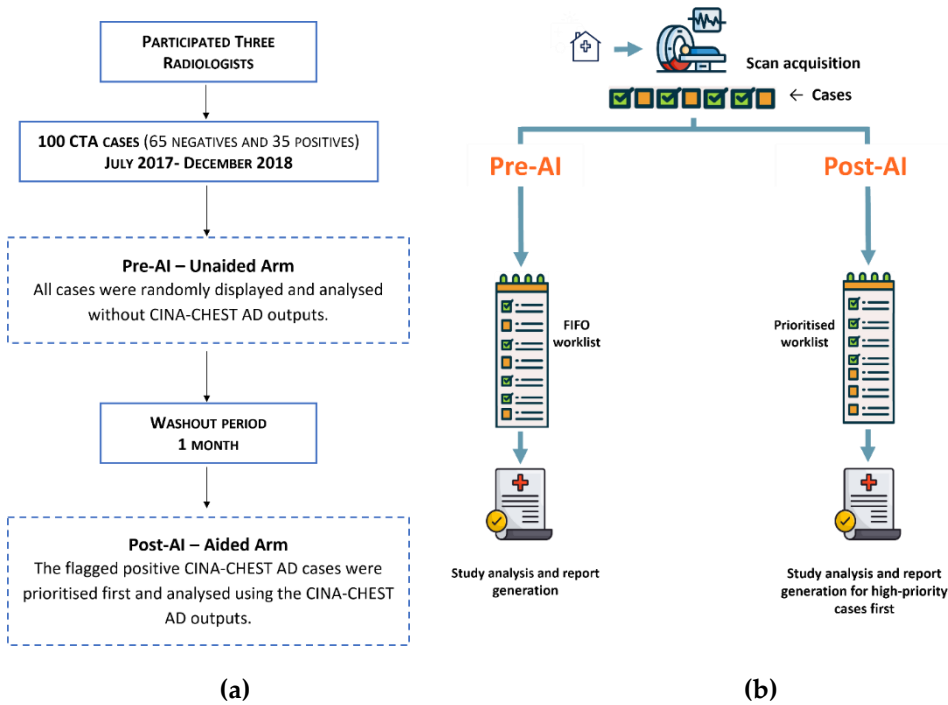


Figure 1. a: The current study design overview, b: Workflow diagram illustrating the traditional radiologist reading process alongside the AI-assisted approach.

In the pre-AI phase, cases were presented in random order without alerts, simulating a conventional first-in, first-out (FIFO) worklist [22]. Radiologists analysed the cases as they would do in their routine daily practice, with each CTA appearing in the radiologists’ worklist sequentially and the most recently completed examination positioned at the bottom of the list. In the post-AI phase, the AI application flagged potential positive AD cases, prioritising them at the top of the worklist for evaluation (Figure 1b, Figure 2).

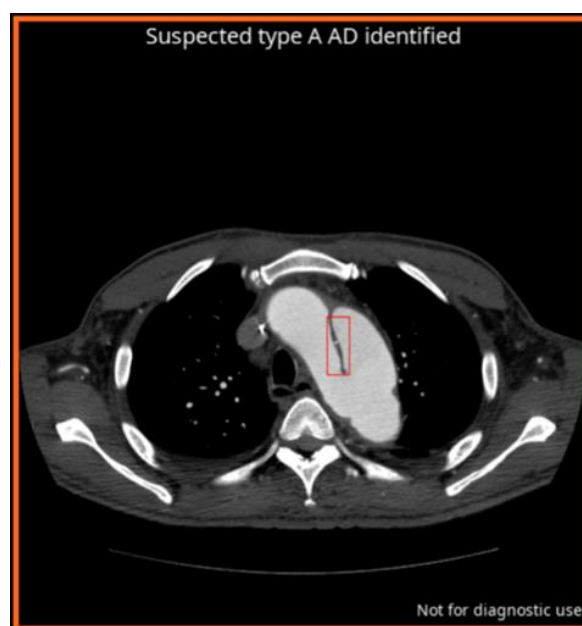


Figure 2. CINA-CHEST for Aortic Dissection (AD) outputs. The Red bounding box shows the localisation of the automatically detected AD. The type of AD detected is provided with mention "Suspected Type A AD identified".

All readers assessed the CTAs for AD. Uncertain cases were marked as indeterminate, while cases without dissection were labelled "No dissection." Indeterminate cases were excluded from the analysis. Readers were blinded to each other, the ground truth, and patient clinical data.

The time taken by each reader to analyse each case was automatically measured, starting from the moment the reader began the analysis and ending when they validated the result and proceeded to the next case. Based on this data, we evaluated two key metrics: the primary endpoint, equivalent to Scan-to-Assessment Time (STAT) for true positive cases, defined as the cumulative duration (in minutes) from when a study becomes available for interpretation on the clinical workstation to the moment the final diagnosis is made; and the secondary endpoint, per-case Interpretation Time (IT), which is the time interval (in minutes) beginning when a radiologist opens the corresponding CTA and ending when they submit their final response. This latter represents the duration required for the radiologist to analyse one case, interpret it, and confirm their findings before moving on to the next case.

The effectiveness of CINA-CHEST for AD in reducing the time required to identify and assess AD on CTA was evaluated under the assumption that all the scans were previously acquired and became available for interpretation on the workstation simultaneously, simulating the most critical situation in an emergency hospital department with high workload.

2.4. Statistical Analysis

Initially, the results computed by CINA-CHEST for AD were compared to the ground truth, with the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, and accuracy calculated for the entire dataset. The 95% confidence intervals (CI) for sensitivity, specificity, and accuracy were determined using the Clopper–Pearson method, based on the exact binomial distribution. The performance of each of the three U.S. board-certified radiologists was first individually evaluated against the ground truth and subsequently assessed across the two phases of the study, before and after the implementation of AI, by measuring the AUROC, sensitivity, specificity, and accuracy in each phase.

Furthermore, to compare workflow metrics pre- and post-AI, the STAT for each true positive case was calculated for each arm as follows (Equation 1): $STAT(n) \text{ in minutes} = \sum_{i=1}^n IT(i)$ where $STAT(n)$ represents the scan-to-assessment time for the n -th case and $IT(i)$ denotes the interpretation

time for the i -th case. Stratified analyses were conducted to explore potential variations across various subgroups based on readers' expertise, with a per-group comparison of junior (< 5 years of clinical experience) versus senior readers (≥ 5 years of experience). Additionally, a comparison of STAT between pre- and post-AI phases including all positive and negative cases was performed. In addition, the per-case IT was calculated for each arm, as follows (Equation 2): $IT\ (seconds) = T_{Diagnosis} - T_{study\ open}$. The differences between the aided and unaided arms for mean STAT and mean IT were assessed.

To evaluate the statistically significant difference (significant reduction, $\alpha \leq 0.05$, two-sided) between the aided arm and the unaided arm, a mixed effects repeated measures model was implemented with reader, case and AI usage (Aided versus Unaided) included as fixed effect terms in the model and a paired sample t-test was conducted [23]. A p-value < 0.05 was considered to represent statistical significance. All statistical analyses were conducted using MedCalc Statistical Software (v22.023, MedCalc Software Ltd).

3. Results

3.1. AI and Readers Performance

Radiologists reviewed a total of 100 CTAs (65 negatives and 35 positives according to the ground truth) in the pre-AI phase and post-AI phase. Five negative cases were excluded from the cohort due to readers' indeterminate responses, resulting in a final analysis of 285 CTA cases (95 cases per reader, including 35 positive cases), for each arm (Table 1). The mean age of the patients was 58 years \pm SD = 14.7, and 52.63% were male. CINA-CHEST for AD performance against the ground truth resulted in a AUROC of 0.971 (95% CI: 0.915 - 0.995), accuracy of 97.89% (95% CI: 92.6 - 99.74%), sensitivity of 94.29% (95% CI: 80.84 - 99.3%), and specificity of 100% (95% CI: 94.04 - 100.00%). The software misclassified CTAs in 2 out of 95 patients (2.1%), resulting in 2 false negatives and 33 true positives. The first false negative result was attributed to the presence of an intramural hematoma (IMH) along with a graft or stent, while the second was due to the presence of IMH alone.

Table 1. Table 1. Data characteristics. Scanner makers and Slice thickness distributions.

Scanner makers	Occurrence (%)	Slice Thickness	Occurrence (%)
GE MEDICAL SYSTEMS	59 (62.11%)	ST < 1 mm	4 (4%)
SIEMENS	21 (22.1%)	1 \leq ST \leq 2.5mm	83 (87%)
CANON (Formerly TOSHIBA)	10 (10.53%)	ST \leq 3mm	8 (9%)
PHILIPS	5 (5.26%)		
TOTAL 95			

The accuracy of three readers when using CINA-CHEST for AD (Aided phase) was 97.89% (95% CI: 92.6 - 99.74%), 97.89% (95% CI: 92.6 - 99.74%), and 98.94% (95% CI: 94.27 - 99.97%) respectively, compared to 98.95% (95% CI: 94.33 to 99.97%), 97.89% (95% CI: 94.33 to 99.74%), and 97.89% (95% CI: 92.60 - 99.74%) without the use of the software (Unaided phase). The sensitivity, specificity, and AUROC of the three readers across the two phases of the study, are shown in Table 2. Statistical analysis found no significant difference ($p > 0.05$) between the two phases. All readers misclassified one case as positive in the pre-AI phase, which was confirmed as true negative when using the device (post-AI phase). Additionally, two of the three readers classified two cases as indeterminate in the pre-AI phase, later identified as true negatives in the post-AI phase, however these two cases were not included in the final analysis.

Table 2. Readers' performances without and with CINA-CHEST for AD identification. The results include 95 CTA readings per reader and per arm. AUROC: area under the receiver operating characteristics curve.

Parameter % [95% CI]	Reader 1		Reader 2		Reader 3	
	Pre-AI	Post-AI	Pre-AI	Post-AI	Pre-AI	Post-AI
Accuracy	98.95%	97.89%	97.895%	97.895%	97.895%	98.94%
	[94.33-	[92.60-	[92.60-	[92.60-	[92.60-	[94.27-
	99.97%]	99.74%]	99.74%]	99.74%]	99.74%]	99.97%]
Sensitivity	100%	100%	100%	94.27%	97.143%	97.143%
	[89.99-	[89.99-	[89.99-	[80.84-	[85.08-	[85.08-
	100.0%]	100.0%]	100.0%]	99.3%]	99.92%]	99.92%]
Specificity	98.33%	96.66%	96.66%	100%	98.33%	100%
	[91.20-	[88.47-	[88.47-	[94.03-	[91.20-	[94.03-
	99.96%]	99.59%]	99.59%]	100.0%]	99.96%]	100.0%]
AUROC	0.992	0.983	0.983	0.971	0.977	0.986
	[0.947-1.0]	[0.933-	[0.933-	[0.915-	[0.924-0.997]	[0.937-
		0.999]	0.999]	0.995]		0.999]

3.2. Comparison of Scan-to-Assessment Time

3.2.1. STAT for AD True Positives Cases

A comprehensive analysis was conducted to compare the STAT only for AD true positive cases before and after the implementation of the software. The study focused on the efficiency of time to assessment across 33 true positive AD cases per arm and per reader, resulting in a total of 99 cases analysed in pre- and post-AI phases.

The analysis across the three readers demonstrated a significant reduction in STAT when using the software compared to the unaided arm ($p < 0.01$). In the pre-AI phase, the mean STAT for true positive cases was 15.84 minutes (SD: 12.22 min; 95% CI: 13.37 - 18.31 min). In contrast, the post-AI phase showed a notable reduction in mean STAT to 5.07 minutes (SD: 4.24 min; 95% CI: 4.23 - 5.91 min). This represents a decrease of 10.77 minutes (SD: 12.96 min; 95% CI: -13.36 to -8.18 min) (Figure 3).

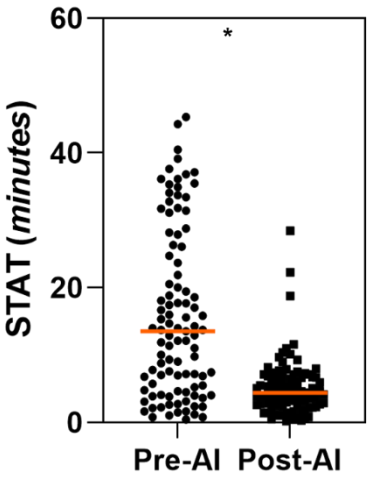


Figure 3. Comparison of Scan-to-Assessment Time (STAT) only for AD True Positive Cases Before and After AI Implementation. STAT (in minutes) was measured for true positive AD cases in both pre- and post-AI phases, based on assessments by three independent readers evaluating 33 CTA scans per condition. Each data point represents the STAT for an individual case, with the central line indicating the median. $*p < 0.05$ for statistically significant difference between the two conditions according to paired t-test.

Furthermore, a per-reader analysis of AD true positive cases indicated that all three readers experienced a significant reduction in STAT between the pre- and post-AI phases ($p < 0.05$). Specifically, the reduction ranged from 5.08 (SD: 7.05 min; 95% CI: -7.64 to -2.53 min) to 15.19 minutes (SD: 14.69 min; 95% CI: -20.39 to -9.98 min), (Figure 4 and Table 3).

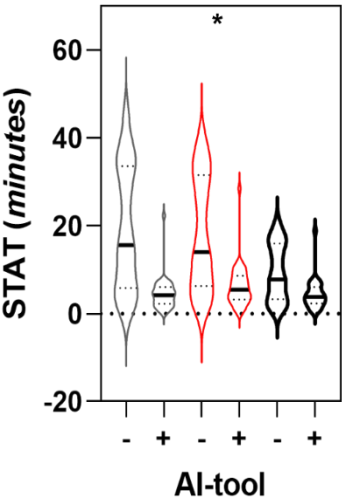


Figure 4. Per-reader comparison of Scan-to-Assessment Time (STAT) for True Positive AD Cases Before and After AI Implementation. STAT (in minutes) was measured for True positive AD cases in pre-AI (-) and post-AI (+) phases. Three independent readers evaluated 33 CTA scans per condition. The main central line corresponds to the median value. * $p < 0.05$ for statistically significant difference between the two conditions according to paired t-test.

Table 3. Comparison of Scan-to-assessment Time (STAT) for True positive AD cases between the Unaided and Aided Arms. The overall analysis across all readers (N = 3 radiologists) includes a total of 99 cases for each arm and a per-reader comparison with 33 cases per arm. Results are expressed in minutes. *The difference is statistically significant ($p < 0.05$) according to the paired t-test.

STAT for True Positive AD cases	Unaided Arm Time (min) Mean \pm SD [95% CI]	Aided Arm Time (min) Mean \pm SD [95% CI]	Aided - Unaided Difference (min) Mean \pm SD [95% CI]
All readers (N = 99)	15.84 \pm 12.22 [13.37, 18.31]	5.07 \pm 4.24 [4.23, 5.91]	-10.77* \pm 12.96 [-13.36, -8.18]
Reader 1 (N = 33)	9.45 \pm 6.41 [7.17, 11.72]	4.36 \pm 3.36 [3.17, 5.56]	-5.08* \pm 7.05 [-7.64, -2.53]
Reader 2 (N = 33)	19.72 \pm 14.09 [14.06, 25.38]	4.53 \pm 3.83 [3.17, 5.88]	-15.19* \pm 14.69 [-20.39, -9.98]
Reader 3 (N = 33)	18.36 \pm 12.29 [13.79, 22.94]	6.32 \pm 5.06 [4.53, 8.10]	-12.04* \pm 13.85 [-18.62, -5.46]

Finally, a per-group analysis on AD true positive cases comparing junior versus senior readers revealed a significant reduction in STAT for both groups ($p < 0.05$). For the junior group, the difference between pre- and post-AI conditions was -5.08 minutes (SD: 7.05 min; 95% CI: -7.64 to -2.53 min). For the senior group, the difference was -13.63 minutes (SD: 14.25 min; 95% CI: -17.12 to -10.13 min) (Table 4).

Table 4. Per-group comparison of Scan-to-assessment time (STAT) Junior versus Senior group for the Aided and Unaided Arm. All readers were taken into account (N = 3 readers) with a total of 33 True positive cases per arm. The results are expressed in minutes. *The difference is statistically significant ($p < 0.05$) according to the paired t-test for mean difference.

Readers' experience	Unaided Arm Time (min) Mean \pm SD [95% CI]	Aided Arm Time (min) Mean \pm SD [95% CI]	Aided - Unaided Difference (min) Mean \pm SD [95% CI]
Junior (N = 33)	9.45 \pm 6.41 [7.17, 11.72]	4.36 \pm 3.36 [3.17, 5.56]	-5.08* \pm 7.05 [-7.64, -2.53]
Senior (N = 66)	19.04 \pm 13.41 [15.74, 22.67]	5.43 \pm 4.54 [4.31, 6.54]	-13.63* \pm 14.25 [-17.12, -10.13]

3.2.2. STAT for All Cases

The STAT was evaluated across all 95 cases, including both positives and negatives, to assess the overall impact of AI implementation. These results represent the combined performance of all three readers. In the pre-AI phase, the mean STAT was 17.17 minutes (SD: 12.16 min; 95% CI: 15.75 - 18.60 min). With the assistance of the AI software, the mean STAT was reduced to 12.54 minutes (SD: 7.15 min; 95% CI: 11.71 - 13.86 min). This reduction of around 4.62 minutes (SD: 13.06 min; 95% CI: -6.14 to -3.10 min) was statistically significant ($p < 0.01$) (Table 5).

The STAT was also analysed on a per-reader basis for all cases. For the first reader, the time difference between the pre-AI and post-AI conditions was 0.45 minutes or 27.26 seconds (SD: 7.78 min; 95% CI: 1.13 - 2.04 min), which was not statistically significant ($p > 0.05$). In contrast, the second and third readers showed significant differences between the two study phases: -9.84 minutes (SD: 14.33 min; 95% CI: -12.76 to -6.92 min) ($p < 0.01$) and -4.48 minutes (SD: 13.99 min; 95% CI: -7.33 to -1.63 min) ($p < 0.01$), respectively (Table 5).

Table 5. Comparison of Scan-to-assessment Time (STAT) for all cases between the Aided and Unaided Arms. The overall analysis across all readers (N = 3 radiologists) includes a total of 285 cases for each arm. A per-reader comparison was conducted with 95 cases (positives and negatives) per arm. The results are expressed in minutes. *The difference is statistically significant ($p < 0.05$) according to the paired t-test.

STAT for all cases	Unaided Arm Time (min) Mean \pm SD [95% CI]	Aided Arm Time (min) Mean \pm SD [95% CI]	Aided - Unaided Difference (min) Mean \pm SD [95% CI]
All readers (N = 285)	17.17 \pm 12.16 [15.75, 18.60]	12.54 \pm 7.15 [11.71, 13.86]	-4.62* \pm 13.06 [-6.14, -3.10]
Reader 1 (N = 95)	10.18 \pm 6.10 [8.94, 11.43]	10.64 \pm 5.55 [9.51, 11.76]	0.45 \pm 7.78 [1.13, 2.04]
Reader 2 (N = 95)	21.46 \pm 13.50 [18.71, 24.21]	11.62 \pm 6.61 [10.28, 13.04]	-9.84* \pm 14.33 [-12.76, -6.92]
Reader 3 (N = 95)	19.85 \pm 12.32 [17.34, 22.36]	15.37 \pm 8.22 [13.70, 17.94]	-4.48* \pm 13.33 [-7.33, -1.63]

3.3. Comparison of Per-Case Interpretation Time (IT)

The IT was evaluated for a total of 570 cases (95 cases per arm, per reader). The mean per-case IT was 21.22 seconds or 0.35 minutes (SD: 11.62s; 95% CI: 19.87 – 22.58s) in the pre-AI phase, whereas in the post-AI (Aided) phase, the value decreased to 14.17 seconds or 0.24 minutes (SD: 6.7s; 95% CI: 13.39 – 14.95s). This reduction in IT represents a statistically significant difference of 7.04 seconds or 0.11 minutes ($p < 0.01$) (Figure 5).

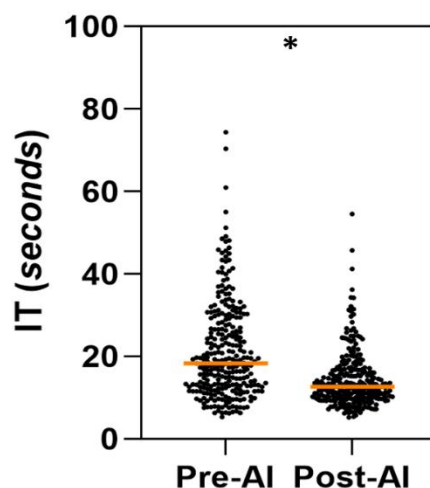


Figure 5. Per-case Interpretation Time (IT) in pre-AI and post-AI phases. IT times (in seconds) measured for both phases, based on assessments from three independent readers across 95 CTA scans per condition. The central line corresponds to the median value. * $p < 0.05$ for statistically significant difference between the two conditions according to paired t-test.

4. Discussion

In this retrospective, multi-center, fully-crossed MRM study, three radiologists interpreted 95 cases twice to evaluate the clinical efficacy of CINA-CHEST for AD. The primary goal was to determine if the AI tool could reduce the STAT for AD-positive cases, compared to the STAT of a traditional FIFO workflow. This study underscored the challenges radiologists face in promptly identifying urgent cases in critical situations with very busy worklists. To the best of our knowledge, this study is the first to demonstrate improved STAT and IT in identification and prioritisation of AD on CTA scans using AI.

The CINA-CHEST for AD software showed robust standalone performance, achieving an AUROC of 0.971, 97.89% accuracy, 94.29% sensitivity, and 100% specificity. In the post-AI phase, readers' performance remained consistently high. The junior reader maintained 100% sensitivity across both phases, with only a minor non-significant drop in specificity (98.33% to 96.66%). The two senior readers experienced a small improvement in accuracy post-AI, with one of them reaching 98.94% accuracy compared to 97.89% pre-AI. The overall AUROC remained high for all readers in both phases, indicating that the software effectively supported accurate diagnoses without significant changes in their performance. Notably, in the pre-AI phase, all readers misclassified one case as positive, which was later confirmed as a true negative using the help of the device in the post-AI phase. Additionally, two of the three readers classified two cases as indeterminate in the pre-AI phase, later identified as true negatives post-AI, though these two cases were excluded from the final analysis. These findings underscore the AI tool's ability to maintain diagnostic accuracy and reliability, even for a pathology generally considered easy to identify, without introducing variability in clinical judgment.

By simulating real-world clinical workflow, we found that AI expedited identification, and prioritisation of the most critical AD cases. With AI-assisted detection, radiologists identified all positive cases in an average of 5.07 minutes, approximately 11 minutes faster than with traditional FIFO methods, resulting in a 68% improvement in efficiency. Similarly, a global comparison of STAT,

including all positive and negative cases, resulted in a significant reduction of 26.8% (4.6 minutes) in the aided arm, demonstrating that AI streamline the detection of critical cases but also enhances overall workflow efficiency for both urgent and routine cases.

Prioritising radiology worklists may enhance patient care and reduce radiologist workload, in contrast to the traditional FIFO workflow, which is driven by often incomplete and ambiguous priority categories (e.g., stat, ASAP, now, critical) determined by the ordering physician's urgency information [24]. By actively prioritising positive cases, AI may enable radiologists to spend more time on critical instances which are often misdiagnosed due to its nonspecific symptoms [25,26]. According to the International Registry of Acute Aortic Dissection, the median time from presentation in the emergency department to a definitive diagnosis of acute aortic dissection is 4.3 hours, due to the high patient load [27,28]. Hence, the implementation of an AI algorithm capable of detecting AD features on CT images could greatly reduce delays in treating potentially serious aortic lesions and shorten the patient's hospital length of stay [17,29].

To date, no other studies have focused specifically on the prioritisation of AD detection in relation to STAT. Previous studies have assessed the performances of AI triage solutions for AD detection, mostly evaluating the diagnostic ability and accuracy of the algorithm compared to those of the radiologists. However, these findings might not determine the true clinical benefit on radiology workflow and patient management [19,30–32]. Harris et al. developed a convolutional neural network model trained to detect AD and rupture, resulting in a median reduction of 395 seconds in the delay time, the interval between when a study is received by the system and when it is opened by a radiologist. This is the only study found in the literature that begins to address study prioritisation for AD, however, its primary focus is on evaluating the technical processing performances of the device [21].

Conversely, there are several studies evaluating AI-based prioritisation effectiveness for different pathologies, such as pulmonary embolism (PE), intracranial haemorrhage (ICH) or cancer. For instance, AI prioritised worklists significantly reduced the time to diagnosis of incidental PE on CT scans in cancer patients. The median turnaround time (TAT) for true positive examinations flagged by the AI software, was 91 minutes, resulting in a significant reduction of TAT from several days to 1.5 hours, compared to the traditional FIFO workflow [33]. Moreover, the implementation of AI in radiology significantly reduced scan-to-alert time (from scan initiation to AI alert) for PE patients, with an average AI alert time of under 6 minutes. These findings highlight the critical importance of a prioritisation model, which may improve patients' chance of survival [34]. Similarly, regarding ICH, the incorporation of a machine learning algorithm into the clinical radiology workflow, decreased significantly the mean report TAT from 75 to 69 minutes in emergency settings, leading to faster critical case identification and improved patient outcomes in urgent care scenarios [35]. Even though these studies were conducted for other pathologies different to AD, they demonstrate that AI-based prioritisation can significantly reduce TATs and improve patient outcomes, highlighting the potential for similar benefits in other diagnostic areas.

On the other hand, regarding individual reader performance, our stratified analyses revealed that each of the three junior and senior readers experienced a significant reduction in STAT for true positive cases during the post-AI phase, with reductions ranging from 5 to 15 minutes compared to the pre-AI phase. When including positive and negative cases in the analysis, one of the three readers showed no significant difference between the aided and unaided study arms, and even a negligible, non-significant increase in STAT. This variability reflects differences in how AI integration could affect individual readers' workflows, however, despite individual discrepancies on average the overall STAT was statistically reduced. In addition, for true positive cases, the senior group experienced a greater gain in STAT (71.5%) from the AI tool than the junior group (53.8%), suggesting that more experienced radiologists gained a greater efficiency boost from the AI implementation compared to their less experienced counterparts. Other studies evaluating IT rather than TAT have shown similar trends, Bennani et al., reported greater IT improvements for general radiologists (34%) compared to residents (30%) in an MRMC study on thoracic abnormalities, and Muller et al., noted a small increase in IT for one resident with AI aid, though residents perceived a better overview of

cases [36]. The greater efficiency improvement observed in senior radiologists suggests that their advanced skills and familiarity with complex cases enable them to leverage AI tools more effectively than junior radiologists. However, the impact of AI can depend on user experience and clinical context.

Finally, an analysis was performed specifically on IT, the average time a radiologists spend interpreting a case regardless of its position in the working list. We found that the aid of the AI software improves radiologists' IT by 33%, which highlights the helpfulness of AI in AD detection and consequently in identifying the most complex cases, enabling radiologists to spend more time on these critical instances. Previous studies have shown that AI tools can reduce IT for chest CT scans by 7% to 44% compared to traditional methods in detecting and measuring lung nodules. These results align with our findings, as we used a similar paired study design where the same reader analysed each case twice [37]. In a prospective study of 390 patients, cardiothoracic radiologists interpreting chest CT exams with and without AI assistance observed a 22.1% reduction in IT. The AI system automatically labeled, segmented, and measured both normal structures and abnormalities across the cardiac, pulmonary, and musculoskeletal systems [38]. These findings suggest that AI automation enhances efficiency by streamlining diagnostics and optimising radiology workflows.

In summary, the significant reduction in IT and STAT underscores the transformative impact of AI on radiology. By streamlining the diagnostic process and improving STAT, AI not only enhances efficiency but also shifts the paradigm from traditional FIFO methods to more effective prioritisation. This advancement highlights the remarkable potential of AI to revolutionize radiology workflows, making it an invaluable tool in modern medical practice. This study was conducted across multiple clinical sites, scanner makers, and countries, this reader study benefits from a diverse dataset encompassing various imaging parameters and patient profiles. Moreover, it engaged a panel of readers representing different expertise levels encountered in real world clinical practice.

This study has a few limitations. First, the workflow impact of the AI tool was only assessed among three radiologists of varying experience and so the effect may differ among other radiologists. Future research should evaluate AI's impact across a broader group of radiologists. Additionally, the database analysed only consisted of 100 cases, with 5% excluded due to 'Indeterminate' responses, and future studies should include a larger sample size. Conducting a prospective study in future would provide a more robust validation of these findings in a real-world setting.

5. Conclusions

In conclusion, our MRMC study is the first to highlight the positive impact of AI on clinical workflow for detecting AD. The AI tool significantly reduces the time radiologists need to identify positive cases in the emergency department, enabling prioritisation of these critical cases. Timely diagnosis and intervention are crucial for improving outcomes in AD. By automatically flagging and prioritising suspected cases, AI enhances workflow efficiency, allowing radiologists to focus on the most urgent cases first and improve emergency medical care.

Author Contributions: Conceptualization, Y.C., P.D.C., A.A. and S.Q; Methodology, A.A. and S.Q.; Software, M.R.-S.; Validation, Y.C., A.A., and V.L.; Formal analysis, A.A. and M.C.; Investigation, Y.C., J.C.J. B.W. and C.Z.; Resources, P.D.C., D.S.C., J.E.S., and J.C.J., B.W. and C.Z.; Data curation, M.R.-S. and A.A. Writing—original draft preparation, M.C.; Writing—review and editing, M.C., A.A., Y.C., M.R.-S., J.E.S. and S.Q.; Visualization, M.C.; Supervision, Y.C. and A.A.; Project administration, S.Q. All authors have read and agreed to the published version of the manuscript.

Funding: The authors state that this work has not received any funding.

Institutional Review Board Statement: Not applicable. The study was conducted in accordance with the Declaration of Helsinki. Prior to investigator assessments, all data were anonymized in compliance with HIPAA and GDPR regulations. Informed consent was waived when it was deemed necessary, following national legislation and institutional protocols, before the data were transferred to Avicenna.AI. The investigators had no possibility to ascertain the identity or patient's private data.

The investigators had no access to the keys of coded data. Consequently, the data were deemed exempt from IRB approval.

Informed Consent Statement: Informed consent was waived when mandatory, aligning with both national legislation and institutional protocols.

Data Availability Statement: The study data is the property of Avicenna.AI and is not publicly accessible. It can be obtained from the corresponding author upon reasonable request and with the approval of the Regulatory Affairs Department of Avicenna.AI.

Acknowledgments: The authors are grateful to all participants who made this work possible, in particular to the clinicians involved in this study and those who contributed to establishing the ground truth. They also appreciate the patients who provided the CTA scan data anonymously.

Conflicts of Interest: The authors of this manuscript declare relationships with the following companies: Avicenna.AI.

References

1. Konstantinides SV, Meyer G, Becattini C, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). *European Heart Journal* 2020;41(4):543–603. <https://doi.org/10.1093/eurheartj/ehz405>.
2. Gawinecka J, Schönrath F, Eckardstein A von. Acute aortic dissection: pathogenesis, risk factors and diagnosis. *Swiss Medical Weekly* 2017;147(3334):w14489–w14489. <https://doi.org/10.4414/smww.2017.14489>.
3. Acharya M. Diagnosis and acute management of type A aortic dissection. *Br J Cardiol* 2023. <https://doi.org/10.5837/bjc.2023.012>.
4. Kesävuori R, Kaseva T, Salli E, Raivio P, Savolainen S, Kangasniemi M. Deep learning-aided extraction of outer aortic surface from CT angiography scans of patients with Stanford type B aortic dissection. *European Radiology Experimental* 2023;7(1):35. <https://doi.org/10.1186/s41747-023-00342-z>.
5. Isselbacher EM, Preventza O, Hamilton Black J, et al. 2022 ACC/AHA Guideline for the Diagnosis and Management of Aortic Disease: A Report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. *Circulation* 2022;146(24):e334–482. <https://doi.org/10.1161/CIR.0000000000001106>.
6. Sebastià C, Pallisa E, Quiroga S, Alvarez-Castells A, Dominguez R, Evangelista A. Aortic dissection: diagnosis and follow-up with helical CT. *Radiographics* 1999;19(1):45–60; quiz 149–50. <https://doi.org/10.1148/radiographics.19.1.g99ja0945>.
7. De León Ayala IA, Chen Y-F. Acute aortic dissection: an update. *Kaohsiung J Med Sci* 2012;28(6):299–305. <https://doi.org/10.1016/j.kjms.2011.11.010>.
8. Coady MA, Rizzo JA, Goldstein LJ, Elefteriades JA. NATURAL HISTORY, PATHOGENESIS, AND ETIOLOGY OF THORACIC AORTIC ANEURYSMS AND DISSECTIONS. *Cardiology Clinics* 1999;17(4):615–35. [https://doi.org/10.1016/S0733-8651\(05\)70105-3](https://doi.org/10.1016/S0733-8651(05)70105-3).
9. Taylor-Phillips S, Stinton C. Fatigue in radiology: a fertile area for future research. *Br J Radiol* 2019;92(1099):20190043. <https://doi.org/10.1259/bjr.20190043>.
10. The Royal College of Radiologists. Turnaround times – what are we seeing? | The Royal College of Radiologists n.d. <https://www.rcr.ac.uk/news-policy/policy-reports-initiatives/turnaround-times-what-are-we-seeing/> (accessed July 24, 2024).
11. Zschke L, Habazettl H, Thürau J, et al. Acute type A aortic dissection: Aortic Dissection Detection Risk Score in emergency care – surgical delay because of initial misdiagnosis. *European Heart Journal Acute Cardiovascular Care* 2020;9(3_suppl):S40–7. <https://doi.org/10.1177/2048872620914931>.
12. Froehlich W, Tolenaar JL, Harris KM, et al. Delay from Diagnosis to Surgery in Transferred Type A Aortic Dissection. *Am J Med* 2018;131(3):300–6. <https://doi.org/10.1016/j.amjmed.2017.11.009>.
13. Dey D, Slomka PJ, Leeson P, et al. Artificial Intelligence in Cardiovascular Imaging. *J Am Coll Cardiol* 2019;73(11):1317–35. <https://doi.org/10.1016/j.jacc.2018.12.054>.
14. Ojeda P, Zawaideh M, Mossa-Basha M, Haynor D. The utility of deep learning: evaluation of a convolutional neural network for detection of intracranial bleeds on non-contrast head computed tomography studies. In: Angelini ED, Landman BA, eds. *Medical Imaging 2019: Image Processing*. San Diego, United States: SPIE; 2019;128. <https://doi.org/10.1117/12.2513167>.
15. Hunter B, Chen M, Ratnakumar P, et al. A radiomics-based decision support tool improves lung cancer diagnosis in combination with the Herder score in large lung nodules. *eBioMedicine* 2022;86. <https://doi.org/10.1016/j.ebiom.2022.104344>.
16. Schmuelling L, Franzeck FC, Nickel CH, et al. Deep learning-based automated detection of pulmonary embolism on CT pulmonary angiograms: No significant effects on report communication times and patient

- turnaround in the emergency department nine months after technical implementation. *European Journal of Radiology* 2021;141:109816. <https://doi.org/10.1016/j.ejrad.2021.109816>.
17. Petry M, Lansky C, Chodakiewitz Y, Maya M, Pressman B. Decreased Hospital Length of Stay for ICH and PE after Adoption of an Artificial Intelligence-Augmented Radiological Worklist Triage System. *Radiology Research and Practice* 2022;2022:1–7. <https://doi.org/10.1155/2022/2141839>.
 18. Xu X, He Z, Niu K, Zhang Y, Tang H, Tan L. An Automatic Detection Scheme of Acute Stanford Type A Aortic Dissection Based on DCNNs in CTA Images. *Proceedings of the 2019 4th International Conference on Multimedia Systems and Signal Processing*. Guangzhou China: ACM; 2019;16–20. <https://doi.org/10.1145/3330393.3330405>.
 19. Laletin V, Ayobi A, Chang PD, et al. Diagnostic Performance of a Deep Learning-Powered Application for Aortic Dissection Triage Prioritization and Classification. *Diagnostics* 2024;14(17):1877. <https://doi.org/10.3390/diagnostics14171877>.
 20. U.S FOOD & DRUGS ADMINISTRATION. BriefCase for AD - K222329. 510(k) Premarket Notification. 2022.
 21. Harris RJ, Kim S, Lohr J, et al. Classification of Aortic Dissection and Rupture on Post-contrast CT Images Using a Convolutional Neural Network. *J Digit Imaging* 2019;32(6):939–46. <https://doi.org/10.1007/s10278-019-00281-5>.
 22. Baltruschat I, Steinmeister L, Nickisch H, et al. Smart chest X-ray worklist prioritization using artificial intelligence: a clinical workflow simulation. *Eur Radiol* 2021;31(6):3837–45. <https://doi.org/10.1007/s00330-020-07480-7>.
 23. FRALICK D, ZHENG JZ, Wang B, TU XM, FENG C. The Differences and Similarities Between Two-Sample T-Test and Paired T-Test. *Shanghai Arch Psychiatry* n.d.;29(3):184–8. <https://doi.org/10.11919/j.issn.1002-0829.217070>.
 24. Gaskin CM, Patrie JT, Hanshew MD, Boatman DM, McWey RP. Impact of a Reading Priority Scoring System on the Prioritization of Examination Interpretations. *AJR Am J Roentgenol* 2016;206(5):1031–9. <https://doi.org/10.2214/AJR.15.14837>.
 25. Fleischmann D, Afifi RO, Casanegra AI, et al. Imaging and Surveillance of Chronic Aortic Dissection: A Scientific Statement From the American Heart Association. *Circ Cardiovasc Imaging* 2022;15(3):e000075. <https://doi.org/10.1161/HCI.0000000000000075>.
 26. Lovatt S, Wong CW, Schwarz K, et al. Misdiagnosis of aortic dissection: A systematic review of the literature. *Am J Emerg Med* 2022;53:16–22. <https://doi.org/10.1016/j.ajem.2021.11.047>.
 27. Lloyd-Jones DM. Cardiovascular Health and Protection Against CVD. *Circulation* 2014;130(19):1671–3. <https://doi.org/10.1161/CIRCULATIONAHA.114.012869>.
 28. Freundt M, Kolat P, Friedrich C, et al. Preoperative Predictors of Adverse Clinical Outcome in Emergent Repair of Acute Type A Aortic Dissection in 15 Year Follow Up. *J Clin Med* 2021;10(22):5370. <https://doi.org/10.3390/jcm10225370>.
 29. Mastrodicasa D, Codari M, Bäumlér K, et al. Artificial Intelligence Applications in Aortic Dissection Imaging. *Seminars in Roentgenology* 2022;57(4):357–63. <https://doi.org/10.1053/j.ro.2022.07.001>.
 30. Huang L-T, Tsai Y-S, Liou C-F, et al. Automated Stanford classification of aortic dissection using a 2-step hierarchical neural network at computed tomography angiography. *Eur Radiol* 2022;32(4):2277–85. <https://doi.org/10.1007/s00330-021-08370-2>.
 31. Hata A, Yanagawa M, Yamagata K, et al. Deep learning algorithm for detection of aortic dissection on non-contrast-enhanced CT. *Eur Radiol* 2021;31(2):1151–9. <https://doi.org/10.1007/s00330-020-07213-w>.
 32. Soffer S, Klang E, Shimon O, et al. Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis. *Sci Rep* 2021;11(1):15814. <https://doi.org/10.1038/s41598-021-95249-3>.
 33. Topff L, Ranschaert ER, Bartels-Rutten A, et al. Artificial Intelligence Tool for Detection and Worklist Prioritization Reduces Time to Diagnosis of Incidental Pulmonary Embolism at CT. *Radiology: Cardiothoracic Imaging* 2023;5(2):e220163. <https://doi.org/10.1148/ryct.220163>.
 34. Shapiro J. Shorter Time to Assessment and Anticoagulation with Decreased Mortality in Patients with Pulmonary Embolism Following Implementation of Artificial Intelligence Software 2024.
 35. Davis MA, Rao B, Cedeno PA, Saha A, Zohrabian VM. Machine Learning and Improved Quality Metrics in Acute Intracranial Hemorrhage by Noncontrast Computed Tomography. *Current Problems in Diagnostic Radiology* 2022;51(4):556–61. <https://doi.org/10.1067/j.cpradiol.2020.10.007>.
 36. Müller FC, Raaschou H, Akhtar N, Brejnebo M, Collatz L, Andersen MB. Impact of Concurrent Use of Artificial Intelligence Tools on Radiologists Reading Time: A Prospective Feasibility Study. *Academic Radiology* 2022;29(7):1085–90. <https://doi.org/10.1016/j.acra.2021.10.008>.
 37. Brown M, Browning P, Wahi-Anwar MW, et al. Integration of Chest CT CAD into the Clinical Workflow and Impact on Radiologist Efficiency. *Academic Radiology* 2019;26(5):626–31. <https://doi.org/10.1016/j.acra.2018.07.006>.

38. Yacoub B, Varga-Szemes A, Schoepf UJ, et al. Impact of Artificial Intelligence Assistance on Chest CT Interpretation Times: A Prospective Randomized Study. *AJR Am J Roentgenol* 2022;219(5):743–51. <https://doi.org/10.2214/AJR.22.27598>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.