tomography

MDPI

Systematic Review

AI-Driven CT Analysis for Aortic Disease: A Task-Based Systematic Review of Segmentation, Measurement, and Detection Applications

Francesco Paparo [1,*], Nicola Panvini [1], Enrico Francesco Melani [1], Marco Magliocco [2], Giuseppe Cittadini [3], Lucio Molini [1], Giovanni Pratesi [4,5] and Giovanni Spinella [4,5]

[1] Radiology Unit, Department of Diagnostic Imaging, E.O. Ospedali Galliera, Mura delle Cappuccine 14, 16128 Genoa, Italy; nicola.panvini@galliera.it (N.P.); enrico.melani@galliera.it (E.F.M.); lucio.molini@galliera.it (L.M.)
[2] Department of Experimental Medicine (DIMES), University of Genoa, 16132 Genoa, Italy; marco.magliocco@edu.unige.it
[3] General and Interventional Radiology Unit, AOM-IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy; giuseppe.cittadini@hsanmartino.it
[4] Department of Surgical and Integrated Diagnostic Sciences (DISC), University of Genoa, 16132 Genoa, Italy; giovanni.pratesi@aomliguria.it (G.P.); giovanni.spinella@unige.it (G.S.)
[5] Vascular and Endovascular Surgery Unit, AOM-IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy
* Correspondence: francesco.paparo@galliera.it

Simple Summary

Aortic aneurysms and dissections are life-threatening conditions that physicians assess with computed tomography (CT) scans. Outlining the aorta and measuring its size on these scans by hand is slow and varies between readers. We reviewed 118 studies in which artificial intelligence (AI) was used to outline the aorta, measure it, or detect disease on CT. Reported accuracy was high for most tasks, and in the few studies that compared AI directly with radiologists, for measuring the aortic diameter and for detecting dissection, the two performed similarly. However, most tools were tested on selected data from single hospitals, few studies measured whether AI changes diagnoses or patient care, and only two were tested in everyday clinical use. Larger independent studies in routine practice are needed before these tools can be trusted to guide patient care.

Abstract

Background/Objectives: Computed tomography (CT) and CT angiography (CTA) are central to assessing aortic aneurysm and dissection, but manual segmentation and geometric analysis remain time-consuming and operator-dependent, while evidence on artificial intelligence (AI) is fragmented. This task-based systematic review synthesised AI applications to CT/CTA of the thoracic and abdominal aorta by primary task: segmentation, measurement, or detection and classification. **Methods:** PubMed/MEDLINE and Europe PMC were searched from 2015 to September 2026 following PRISMA 2020; the protocol was registered retrospectively (INPLASY2026110004). Segmentation and measurement studies were appraised with CLAIM and detection studies with QUADAS-2/QUADAS-AI, and results were synthesised without meta-analysis. **Results:** Of 118 included studies (42 segmentation, 38 measurement, 38 detection), half were published since 2024. Median Dice similarity coefficient was 0.90 or higher for the outer wall, lumen and true and false lumina, but 0.86 for intraluminal thrombus and 0.63 for false lumen thrombus. Automated maximum diameter showed a median absolute bias of 0.26 mm (95% limits of agreement about ±5 mm). Detection of aortic dissection reached a median AUC of 0.96–0.97, mostly in enriched samples. External validation was reported in 14%, 34% and 55% of segmentation, measurement and detection studies, 33 of 38 detection studies were at high risk

Check for updates

Academic Editor: Eugenio Vocaturo

Received: 7 October 2026
Revised: 3 November 2026
Accepted: 12 November 2026
Published: 18 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

of bias, and only two systems were evaluated prospectively. **Conclusions:** AI supports technical automation of aortic CT/CTA segmentation and measurement and is progressing toward multistructure geometric phenotyping and clinical decision support, but independent multicentre validation, consecutive cohorts and prospective reader-in-the-loop studies are required before clinical integration. Detection/classification tools require independent diagnostic and workflow validation.

Keywords: artificial intelligence; deep learning; computed tomography angiography; aorta; abdominal aortic aneurysm; aortic dissection; image segmentation; systematic review

1. Introduction

Aortic diseases—including abdominal aortic aneurysm (AAA), thoracic aortic aneurysm (TAA), aortic dissection, and the broader spectrum of acute aortic syndromes—require imaging that defines vascular anatomy rapidly and reproducibly. Computed tomography (CT) and, particularly, CT angiography (CTA) are central to emergency diagnosis, operative planning, and postoperative surveillance, because they depict the aorta and its branches with high spatial and temporal resolution and define disease extent, luminal and wall morphology, and the anatomical information required for open and endovascular intervention [1,2]. The clinical value of CT/CTA, however, depends on the extraction of accurate and reliable geometric information. Although maximum diameter remains the primary metric for diagnosis, procedural planning, and longitudinal surveillance, additional descriptors—volume, vessel length, tortuosity, thrombus burden, calcification, and branch vessel anatomy—provide a more comprehensive depiction of aortic morphology. Deriving these measurements typically requires accurate delineation of complex three-dimensional structures, including the outer aortic wall, the true and false lumina, and intraluminal thrombus (ILT). Segmentation is therefore not an isolated technical task but the foundation on which quantitative morphological assessment, treatment planning, and the monitoring of anatomical change are built [2–4].

This requirement exposes a persistent workflow problem. Manual and semi-automated contouring and measurement demand specialist expertise, are labour-intensive, and are susceptible to intra- and inter-observer variation. The challenge is particularly acute in time-sensitive acute aortic syndromes, where delayed recognition worsens outcomes, and in longitudinal surveillance after endovascular repair, where consistent comparison across serial examinations is essential [5]. Anatomical and technical complexity compounds the problem: tortuosity, calcification, ILT, stent graft artefacts, altered postoperative anatomy, non-ECG-gated acquisition, and variable contrast opacification can all affect the reproducibility of segmentation and measurement. Acquisition parameters matter as well, since slice thickness and spacing measurably influence both segmentation accuracy and derived diameter error [6]. Automated analysis has consequently been identified as a clinically relevant route toward more standardised CT/CTA assessment, provided that agreement with expert reference standards and robustness across acquisition conditions are demonstrated [3,4].

Artificial intelligence (AI), predominantly deep learning (DL), has rapidly expanded the technical repertoire available for aortic CT/CTA analysis. Convolutional neural networks and U-Net-derived architectures dominate segmentation, while other approaches support landmark detection, centreline extraction, direct measurement, screening, detection, and classification. More recent work extends beyond conventional contrast-enhanced CTA to non-contrast CT, image synthesis, multi-task and weakly supervised learning, transformer-based models, and foundation model paradigms [2]. The evidence base is promising but

heterogeneous. Meta-analyses in aortic dissection report high pooled detection accuracy and high segmentation overlap, with lower performance for true and false lumen than for whole-aorta delineation [7,8]. The same syntheses, however, rate the certainty of evidence as low: most datasets are retrospective, single-centre, or enriched case–control samples; external validation is uncommon, and differences in architectures, reference standards, and outcome reporting limit direct comparison and leave generalisability unresolved [7–9].

Existing reviews have made important, albeit deliberately focused, contributions. Guo et al. reviewed DL-based ILT segmentation in AAA, comparing 2D and 3D architectures across preoperative and postoperative applications [3]. Lo et al. focused on AI-assisted automated diameter measurement in thoracic aneurysm and dissection and its integration into clinical workflow [4], and Spinella et al. examined how CT acquisition parameters affect DL aortic segmentation [6]. In aortic dissection, successive meta-analyses have pooled CT-based detection and classification accuracy [9], segmentation and diagnostic performance [7], and head-to-head comparisons between DL and radiologists on non-contrast CT [8]. Broader reviews have addressed AI within multimodal aortic imaging and CTA workflows [1,2] and its potential to shorten diagnostic timelines in acute type A dissection [5].

Taken together, these reviews are organised by pathology, anatomical target, or technical frontier rather than by the function the AI system performs. A complementary synthesis can connect them through the primary task performed by the AI system and the geometric output that task enables, which is how such tools would ultimately be positioned within the imaging workflow. This task-based systematic review therefore examines AI-driven CT/CTA analysis from the ascending to the abdominal aorta across aneurysmal disease and aortic dissection. Studies are organised according to their primary task: (i) segmentation, (ii) measurement and geometric analysis, or (iii) detection and classification. Within each task, we describe the AI architectures used, the aortic segment and pathology studied, the geometric measures produced or used, the reference standard adopted for validation, the intended clinical use, and the reported technical performance. By integrating these dimensions, the review aims to clarify where current AI approaches may improve the precision, reproducibility, and efficiency of aortic CT/CTA assessment, and which methodological gaps still separate them from clinical implementation.

2. Materials and Methods

2.1. Protocol and Reporting

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [10]. The review protocol was not registered prospectively; a protocol describing the methods as conducted was registered retrospectively on INPLASY (INPLASY2026110004; doi:10.37766/inplasy2026.11.0004).

2.2. Eligibility Criteria

Studies were eligible if they met all three core criteria: an AI model was trained or applied to CT or CTA images; the model produced an anatomical output, defined as segmentation, geometric measurement, or spatially localised detection or classification; and model performance was quantitatively evaluated against a reference standard, defined as a manual or expert-adjudicated segmentation, measurement, or diagnosis, or an expert-reader assessment of the model output with an ordinal or metric scale.

Eligible studies addressed the thoracic and/or abdominal aorta in any clinical condition and in normal anatomy, in preoperative and postoperative settings. Both clinical cohorts and methodological studies based on public datasets were eligible. No restriction was

applied to study design or sample size; the date and language limits applied at the search stage are described in Section 2.3.

Records were excluded if they met any of the following criteria: models whose endpoint was the prediction of a future event—growth, rupture, or postoperative outcome—rather than the characterisation of the current examination, which falls outside the scope of this review; studies confined to the aortic valve, root, annulus, or coarctation; studies of non-aortic vascular districts; imaging modalities other than CT, or hybrid modalities in which CT was not the substrate of the analysis; radiomic or feature-based analyses without an anatomical output; studies in which the AI component did not operate on the image itself, including models based on clinical variables alone and natural language processing of reports; studies addressing image reconstruction, image quality, radiation dose, or contrast-medium optimisation; studies confined to the quantification of vascular calcification; surrogate models of computational simulation; studies without any model performance endpoint; and reviews, editorials, letters, corrections, conference abstracts, preprints of published articles, and redundant publications.

Eligible studies were assigned to one of three task categories—segmentation, measurement, and detection or classification—according to the primary output evaluated. Studies reporting more than one task were assigned to the task on which the stated contribution and the conclusions rested.

2.3. Information Sources and Search Strategy

PubMed/MEDLINE and Europe PMC were searched on 5 September 2026 for records published from 1 January 2015 onward. This lower limit was chosen to coincide with the adoption of deep learning in medical image analysis [11]. Europe PMC was searched to capture records not indexed in MEDLINE. No language filter was applied in PubMed/MEDLINE; the Europe PMC search was restricted to records in English.

Both search strings combined three concept blocks with the Boolean operator AND: artificial intelligence, aorta, and computed tomography. No block restricting the type of task was included, so that the search was not limited by application; predictive models were therefore retrieved and excluded at screening. The full search strings are reported in Supplementary Table S1.

IEEE Xplore was evaluated as an additional source: exploratory searching indicated that the yield of eligible clinical studies would not justify the screening burden of its predominantly conference proceedings content, and it was not searched. Database searching was supplemented by backward and forward citation searching of the included studies and of the previous reviews on AI applications to aortic CT identified during screening [2–9]; seven of these reviews had searched Embase, Scopus or Web of Science [3–9].

2.4. Selection Process

Records were deduplicated across sources before screening, with cross-database duplicates and preprint versions of published articles removed. Titles and abstracts of all records, from both databases, were screened by consensus by two reviewers (FP and NP), who examined each record together and recorded a single joint decision in the record register; the same procedure was applied to the re-examination of excluded records. Joint screening was chosen so that the eligibility criteria would be applied uniformly and every borderline record discussed at the time of the decision; because this procedure does not generate two independent judgements, the risk of omitting eligible records on a shared misjudgement was addressed by the structured re-examination described below and by citation searching.

To limit misclassification at the title and abstract stage, screened records were graded into five bands according to the estimated risk of an incorrect decision. All records in the

two highest-risk bands were re-examined in full, intermediate bands were re-scanned, and a sample of the lowest-risk band was verified. In addition, all records excluded on the grounds of a predictive endpoint were re-assessed and the exclusion confirmed.

Full texts were retrieved for all records passing title and abstract screening and for records identified by citation searching and were assessed against the eligibility criteria by the same two reviewers in consensus. Reasons for exclusion at the full-text stage were recorded for every report and are listed in Supplementary Table S2. The study selection process is summarised in the PRISMA flow diagram..

2.5. Data Collection

Data were extracted from the full text into three task-specific forms sharing a common backbone. Extraction conventions and closed term lists were defined a priori and documented within the extraction file. Data were extracted by one reviewer (FP) with the support of Elicit (Elicit Research, Oakland, CA, USA), a web-based AI-assisted research platform, and every extracted value was verified against the full text of the source article by a second reviewer (NP); discrepancies were resolved by discussion against the full text. The complete extraction tables are provided as Supplementary Tables S3–S5 (spreadsheet).

The common backbone captured the country and number of contributing centres; study design; population, including sample size, the training, validation and test split or cross-validation scheme, and whether the split was made at patient or image level; CT technique, including contrast administration; aortic segment, pathology and operative status; AI method; model provenance, coded as developed, adapted, commercial or mixed; external validation; and workflow time.

Task-specific blocks captured, for segmentation, the structures segmented, the primary structure, and the overlap and distance metrics; for measurement, the measurand, the agreement statistic and the reference method; and for detection, the target condition, the unit of analysis, sensitivity, specificity, area under the receiver operating characteristic curve (AUC), and comparison with human readers.

Conventions were adopted to prevent the aggregation of non-comparable quantities. Performance metrics were recorded together with the name of the metric, since not all studies report the Dice similarity coefficient and since the Hausdorff distance and its 95th percentile are not interchangeable. The primary structure of each segmentation study was coded to a closed list of anatomical compartments. Studies reporting only a mean metric across many classes were coded as such. The unit of analysis was recorded for detection studies. Workflow time was recorded only where a human or semi-automatic comparator was measured within the study; inference time reported in isolation was annotated as such.

2.6. Quality Appraisal

Quality appraisal was matched to the primary task. Detection and classification studies were assessed for risk of bias and applicability concerns with the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool [12], supplemented by the QUADAS-AI items on data partitioning, leakage and external testing [13,14]. The overall risk of bias was judged low when all domains were at low risk, high when at least one domain was at high risk, and unclear otherwise; it was based on the risk-of-bias domains only, and applicability concerns were recorded and reported separately without contributing to the overall judgement. Studies whose endpoint was localisation only, or that reported no diagnostic accuracy, were not formally appraised.

Segmentation and measurement studies, for which no validated risk-of-bias tool exists, were appraised descriptively with the Checklist for Artificial Intelligence in Medical Imaging (CLAIM), 2024 update [15], used as a structured source of items and not as a risk-of-bias

instrument. Three areas were rated on a three-level scale (no concern, concern, critical concern): reference standard and data; validation and generalisability; and reporting and reproducibility. A critical concern was defined a priori as a limitation that compromises the validity of the reported performance: a reference standard generated or corrected by the model under evaluation or by the authors' own software, or no reference segmentation or measurement, with agreement assessed by reader scoring or visual review only; test data not independent of training, including image- or scan-level partitioning and the selection of thresholds or parameters on the test data, or evaluation restricted to selected slices or cases; or performance reported only graphically or as group means. Each study was then assigned a limitation tier by a fixed rule: higher if any area carried a critical concern; lower if no area carried a critical concern, an external or independent test was reported, and the human reference was established by at least two readers, including sequential verification by a second reader, or inter-reader variability was documented; moderate in all other cases. Restrictions of the study population were recorded as applicability notes and did not contribute to the tier. Tiers describe methodological and reporting limitations relative to the included body of literature and are not risk-of-bias judgements; the area ratings and the rule are given in Supplementary Table S10 and the resulting tiers in Supplementary Tables S6 and S7.

Appraisal was performed by two reviewers (FP and EFM) in consensus. Condensed appraisal tables are provided in Supplementary Tables S6–S8; the main text presents the pattern of appraisal findings within each framework.

2.7. Synthesis Methods

A meta-analysis was not undertaken because of heterogeneity in targets, metrics, and reference standards; the synthesis was conducted without meta-analysis and reported following the Synthesis Without Meta-analysis (SWiM) guideline [16].

Results were synthesised narratively and in structured tables, separately for each task: Table 1 for segmentation, Table 2 for measurement, and Table 3 for detection, with rows grouped by pathology and operative status. For each task, the primary performance metric was summarised as the median and range across studies reporting that metric, with the number of contributing studies stated. Summaries were computed within groups of comparable targets rather than across all studies, and studies reporting only a mean metric across multiple anatomical classes were described separately and excluded from the aggregate distributions. Studies were additionally grouped by aortic segment and by the presence of external validation, to describe where the evidence is concentrated and where it is absent. These medians are unweighted and descriptive: they are computed across studies that differ in sample size, population, acquisition protocol, reference standard and validation design, and they serve to locate where the evidence is concentrated rather than to compare or rank models. As an additional descriptive analysis, the headline summary statistics of each task were also computed within subgroups defined by three study characteristics. Sample size was dichotomised at 200 patients or scans for segmentation and measurement and at 500 for detection. Validation design was classified according to whether an external test was reported; for measurement, models developed in the study were further distinguished from commercial or open-source software applied without training. Detection studies were additionally divided by CT technique, non-contrast versus contrast-enhanced. Results are reported in Supplementary Table S9.

Table 1. Studies on AI-based aortic segmentation ($n = 42$).

Study	Data	Condition	Structures	Model	Ext. val.	Overlap (Primary)
Aortic dissection ($n = 14$)						
Abaid 2024 [17]	100 scans [a]	TBAD	TL, FL, FLT	2D U-Net (VGG19)	No	DSC 0.90 (FL)
Cao 2019 [18]	276 pts	TBAD	A, TL, FL	3D U-Net, multi-task	No	DSC 0.93 (TL)
Chen 2021 [19]	120 pts	TBAD	TL, FL, Br	Cascaded CNN	No	DSC 0.96 TL; 0.95 FL
Feng 2023 [20]	167 pts	TAAD, post-repair	A, TL, FL, FLT	Cascaded CNN, attention	No	DSC 0.90 (FLT)
Guo 2024 [21]	24 pts	TAAD	TL, FL	nnU-Net (2D/3D)	No	DSC 0.88 (TL)
Hahn (LD) 2020 [22]	45 pts	uTBAD	A, TL, FL	2D U-Net (VGG11)	No	DSC 0.87 TL; 0.89 FL
Imran 2024 [23]	59 pts	uTBAD	A, 13 Br	CIS-UNet (CNN-Swin)	No	DSC 0.73 (mean, 14 classes)
Imran 2026 [24]	100 scans	uTBAD	10 zones, 13 Br	nnU-Net ensemble	No	DSC 0.78 (mean, 23 classes)
Jung 2024 [25]	253 pts + 98 scans [a]	TAAD and TBAD	TL, FL, FLT	3D transformer + U-Net	Yes	DSC 0.92 TL; 0.88 FL; 0.63 FLT (ext 0.94; 0.92; 0.59)
Kesavuori 2023 [26]	206 scans	TBAD (incl. IMH)	A	3D and 2D V-Net	No	DSC 0.96 (W)
Lyu 2021 [27]	42 pts	AD, type not stated	A	PSPNet + HED	No	DSC 0.92 (L)
Wobben 2021 [28]	40 pts	uTBAD	TL, FL, FLT	Residual 3D U-Net	No	DSC 0.63 (FLT)
Yao 2021 [29]	100 scans	TBAD	TL, FL, FLT	3D U-Net, two-stage	No	DSC 0.52 (FLT) [c]
Yu 2021 [30]	139 pts	TBAD without FLT	A, TL, FL	3D U-Net + DenseNet	No	DSC 0.96 TL; 0.93 FL
Abdominal aortic aneurysm ($n = 21$)						
Abdolmanafi 2022 [31]	81 pts	AAA	W, L, ILT, Ca	ResNet-18 FCN	Yes	IoU 0.98 (W) (ext 0.88)
Brutti 2022 [32]	85 scans	AAA	ILT	Multi-view U-Net	No	DSC 0.89 (ILT)
Caradu 2021 [33]	100 scans	AAA	L, ILT	PRAEVAorta (Nurea, Bordeaux, France)	No	DSC 0.95 (W)
Chandrashekar 2023 [34]	75 + 165 pts	AAA [b]	L, ILT	GAN	Yes	DSC 0.86 (L) (ext 0.84)
Derycke 2026 [35]	328 scans	AAA	L, ILT, Ca, Br	nnU-Net	No	DSC 0.97 L; 0.94 ILT; 0.87 Ca
Fantazzini 2020 [36]	80 scans	AAA	L	Multi-view 2D U-Net	No	DSC 0.93 (L)
Hendricks 2026 [37]	75 pts	AAA, pre-EVAR	W	nnU-Net	Yes	DSC 0.97 (W) [g]
Hwang 2022 [38]	60 pts	AAA, post-EVAR	ILT	Mask R-CNN	No	DSC 0.83 (ILT)
Jung 2022 [39]	60 pts	AAA, post-EVAR	ILT	Mask R-CNN + Bi-ConvLSTM	No	DSC 0.87 (ILT)
Lopez-Linares 2018 [40]	13 pts	AAA, post-EVAR	ILT	DetectNet + HED	No	DSC 0.82 (ILT)
Lopez-Linares 2019 [41]	28 scans	AAA, post-EVAR	ILT	3D CNN, synthetic training	No	DSC 0.84 (ILT)
Lyu 2023 [42]	30 scans	AAA	L	3D U-Net, attention/cascade	No	DSC 0.92 (L)
Lyu 2024 [43]	80 scans	AAA	ILT	2D/3D fusion CNN	No	DSC 0.88 (ILT)
Ma 2023 [44]	30 pts	AAA [b]	A	2D/3D fusion CNN	No	DSC 0.89 (W)
Ma 2025 [45]	30 pts + 90 scans [a]	AAA, pre-EVAR [b]	A	CNNs and transformers, weak labels	No	DSC 0.90 (W)
Mu 2023 [46]	70 pts	AAA	L, ILT	Cascaded 3D U-Net	No	DSC 0.80 (ILT)
Roby 2025a [47]	90 pts	AAA	L, W	Patch-based 2D U-Net	Yes	DSC 0.96 L; 0.97 W [g]
Roby 2025b [48]	22 scans	AAA	L, W	2D U-Net, mixture of experts	No	DSC 0.96 L; 0.96 W
Salvi 2021 [49]	30 pts	AAA	A	3D U-Net	No	DSC ≈ 0.26 (W) [d]
Siriapisith 2022 [50]	220 + 40 pts	AAA, pre- and post-EVAR [b]	W	3D CNNs (5)	No	DSC 0.97 (W)
Wang 2022 [51]	340 pts	AAA	L, ILT	DeepLabv3+	No	IoU 0.87 (ILT)
Thoracic aortic aneurysm or dilatation ($n = 2$)						
Aghayeva 2025 [52]	500 scans	Ascending dilatation	A (ascending)	CNN + 2D U-Net	No	DSC 0.99 (W) [e]
Comelli 2021 [53]	72 pts	ATAA	L (ascending)	ENet	No	DSC 0.91 (L)
Mixed or unselected cohorts ($n = 5$)						
Chen 2022 [54]	300 pts	Normal, AD [b]	A	nnU-Net + GAN	No	DSC 0.94–0.95 (W)
El Rahal 2026 [55]	60 scans	Normal, aneurysm, TAAD [b]	A	TotalSegmentator (open source software)	Yes	NR [f]
Liu 2026 [56]	56 scans [a] + 4 cases	Normal, aneurysm, AD	A	MAFR-UNet	No	DSC 0.89 (L)
Sieren 2022 [57]	191 pts	Normal, aneurysm, AD	A	3D U-Net	No	DSC 0.95 (W)
Vagenas 2025 [58]	56 scans [a] + 20 pts	Normal, AAA	A	SwinUNETR	No	DSC 0.91 (L)

Summary: unweighted median DSC [range], number of studies, by primary compartment: outer wall 0.96 [0.89–0.99], 11; lumen 0.92 [0.86–0.97], 8; true lumen 0.92 [0.87–0.96], 6; false lumen 0.90 [0.88–0.95], 5; ILT 0.86 [0.80–0.94], 8; false lumen thrombus 0.63 [0.52–0.90], 4. Co-primary classes

contribute their per-class values. Excluded: multi-class mean only (Imran 2024 [23]; Imran 2026 [24]), IoU only (Abdolmanafi 2022 [31]; Wang 2022 [51]), Dice loss only (Salvi 2021 [49]), reader-based score only (El Rahal 2026 [55]). Rows are grouped by condition and sorted alphabetically. Data: study sample as patients (pts) or scans; "+" adds an independent or public dataset. Overlap: value for the primary structure on the internal test set or cross-validation; external test value in brackets (ext) when available, two decimals; co-primary classes are listed separately; metric as reported by the authors (DSC or IoU), not converted. Full extraction in Table S3. [a] Public dataset. [b] Non-contrast CT included. [c] Averaged including 32 scans without false lumen thrombus. [d] Only the Dice loss is reported; value derived as 1—loss on the out-of-sample test. [e] Slice-level, on classifier-selected slices only. [f] No reference segmentation; reader-based ordinal score anchored to the maximum contour error (mm); excluded from the summary distributions. [g] External test only; internal results not reported. Abbreviations: A: aorta as a single class; AAA: abdominal aortic aneurysm; AD: aortic dissection; ATAA: ascending thoracic aortic aneurysm; Bi-ConvLSTM: bidirectional convolutional long short-term memory; Br: branches; Ca: calcification; CIS-UNet: context-infused Swin U-Net; CNN: convolutional neural network; DSC: Dice similarity coefficient; EVAR: endovascular aneurysm repair; Ext. val.: external validation; FL: false lumen; FLT: false lumen thrombus; GAN: generative adversarial network; HED: holistically nested edge detection; ILT: intraluminal thrombus; IMH: intramural haematoma; IoU: intersection over union; L: lumen; MAFR-UNet: multi-scale adaptive feature reassembly U-Net; NR: not reported; PRAEVAorta: commercial software (Nurea, Bordeaux, France); PSPNet: pyramid scene parsing network; SwinUNETR: Swin UNEt TRansformers; TAAD/TBAD: type A/type B aortic dissection; uTBAD: uncomplicated TBAD; TL: true lumen; TotalSegmentator: open-source software (University Hospital Basel, Basel, Switzerland); VGG: Visual Geometry Group network; W: outer wall.

Table 2. Studies on AI-based aortic measurement ($n = 38$).

Study	Data	Condition	Measurand	Model	Reference	Agreement
Aortic dissection ($n = 4$)						
Bäumler 2026 [59]	396 scans	uTBAD	Landmarks (9)	3D DQN (DRL)	Manual, 5	Mean error 3.4 mm (ext 3.0–3.9)
Krebs 2024 [60]	59 pts	uTBAD	Zonal volume	CIS-UNet	Manual, 3	NR
Nakano 2026 [61]	112 pts	TAAD and TBAD	TL/FL volume	nnU-Net	Manual, 1	Rel. error 3–136%
Pepe 2023 [62]	162 scans	Chronic AD	Measurement plane	3D regression CNN	Manual, 3	MAE 5.8°/16.6° (θ/φ)
Abdominal aortic aneurysm ($n = 12$)						
Broda 2025 [63]	89 pts	AAA, post-EVAR/FEVAR	Dmax, volume	ARVA [c]	Manual, 4	NR
Caradu 2022 [64]	48 pts	AAA, post-EVAR	Dmax, volume	PRAEVAorta [c]	Semi-auto, 1	MAE 0.63 mm
Coatsaliou 2024b [65]	82 pts	AAA, post-EVAR [b]	Dmax, volume	PRAEVAorta 2 [c]	AI, arterial phase	r 0.72–0.78 [e]
Hatzl 2024 [66]	347 pts	AAA, pre-/post-EVAR	Dmax	PRAEVAorta 2 [c]	Core lab	Bias 0.28 (−4.36 to 4.93) mm
Kim (T) 2024 [67]	300 pts	AAA, pre-EVAR	EVAR planning	nnU-Net	Manual, NR	Bias 0.23 (−7.61 to 8.06) mm
Lareyre 2022 [68]	34 pts	AAA, pre-EVAR	Dmax	Hybrid, U-Net	Manual, 2	MAE 0.8 mm
Pouncey 2026 [69]	50 pts	Complex AAA, pre-op	Visceral branches	PRAEVAorta [c]	Manual, 3	Bias −0.11 (−5.34 to 5.13) mm
Robbi 2025 [70]	20 scans [a]	AAA, pre-EVAR	EVAR planning	nnU-Net + SegResNet	Manual, 2	Bias −0.37 (−2.78 to 2.04) mm
Sharma 2025 [71]	142 pts	AAA, untreated	Dmax, volume	PRAEVAorta [c]	Semi-auto, 3	RC 0 mm [d]
van Tongeren 2025 [72]	49 pts	AAA, post-EVAR	Sac volume	PRAEVAorta 2 [c]	Semi-auto, 3	ICC 0.94
Wegner 2025 [73]	50 pts	Complex AAA, pre-/post-FEVAR	Dmax	ARVA [c]	Manual, 2	MAE 2.4/1.6 mm (pre/post)
Weiss 2026 [74]	110 pts + 60 scans	AAA, pre-/post-EVAR	Volume	nnU-Net	Manual, 1	Bias −0.32 mL (ext r 0.96)
Thoracic aortic aneurysm or dilatation ($n = 8$)						
Fink 2024 [75]	100 pts	TAA/dilatation, incl. repair, AD [b]	Dmax	3D U-Net + DRL	Manual, 2	Bias −1.3 (−9.8 to 7.2) mm
Graby 2023 [76]	436 scans	Dilatation [b]	Dmax, 7 sites	AI-Rad Companion [c]	Manual, NR	MAE 1.3–2.6 mm
Katakol 2023 [77]	67 pts	TAA	Dmax, 9 sites	Multitask 3D U-Net	Manual, 3	Bias −1.38 to 0.40 mm
Livolsi 2026 [78]	335 pts	ATAA	Dmax	Attention U-Net	Manual, NR	Bias 0.92 (−5.30 to 7.13) mm
Lo Piccolo 2023 [79]	995 pts	Suspected dilatation [b]	Dmax, 9 sites	AI-Rad Companion [c]	Visual review, 2	NR [f]
Marway 2025 [80]	177 pts	ATAA	Dmax, 6 sites	Multitask 3D U-Net	Manual, NR	Bias 0.04 (−4.2 to 4.3) mm

Table 2. *Cont.*

Study	Data	Condition	Measurand	Model	Reference	Agreement
Pradella 2021 [81]	371 pts	Dilatation	Dmax, 8 sites	AI-Rad Companion [c]	Manual, 2	Bias −0.45 to −0.02 mm
Rueckel 2021 [82]	18 pts	Ascending ectasia/ATAA	Dmax, 9 sites	AI-Rad Companion [c]	Manual, 3	MAE ≈ 1–4.7 mm
Mixed or unselected cohorts (n = 14)						
Adam 2021 [83]	345 pts	Healthy, aneurysm/AD, post-repair	Dmax	3D V-Net (ARVA)	Manual, 7	MAE 1.2 mm
Artzner 2022 [84]	122 pts	Consecutive, AD, IMH, post-TEVAR [b]	Dmax, 9 sites	AI-Rad Companion [c]	Manual, 2	ICC 0.96–0.98
Bratt 2021 [85]	2835 + 57 pts	Stable aortas	Dmax, volume	2D U-Net	Semi-auto, NR	LoA ± 3.1 mm [d]
Gatinot 2026 [86]	32 pts	Post-EVAR/FEVAR/TEVAR, controls [b]	Dmax, EVAR planning	PRAEVAorta 2 [c]	Semi-auto, 2	MAE 2.5/1.4 mm (CTA/NCCT)
Hamelink 2023 [87]	240 pts	Age-stratified, dilatation [b]	Dmax, 8 sites	AI-Rad Companion [c]	Manual, 1	Bias 0 mm
Heo 2025 [88]	300 pts	Trauma, no aortic disease	Diameter	SANet	Manual, NR	MAE 2.11 mm
Hong 2026 [89]	190 pts	Clinical, ±dilatation [b]	Dmax, 3 segments	TotalSegmentator	Semi-auto, NR	ICC 0.95–0.99
Koo 2025 [90]	704 pts	Healthy	Zonal D, volume	nnU-Net + atlas	Corrected AI, 2	Bias 0.08 mm (zone 0)
Park 2026 [91]	1204 pts	Lung screening [b]	Dmax, 4 segments	AVIEW Aorta [c]	Manual, 3	Bias −5.23 to 1.30 mm
Paulsen 2026 [92]	11,418 + 7627 scans	General population	Dmax, 6 sites	TotalSegmentator	Manual, 1	Bias −2.10 to −0.33 mm
Postiglione 2024 [93]	216 pts	Pre-/post-repair, incl. AD	Dmax, volume	3D U-Net (ARVA)	Manual, 1 per segment	Bias −0.1 (−5.5 to 4.9) mm
Saitta 2022 [94]	465 scans	Healthy, TAA	Landing zones	3D U-Net + geometry	Semi-auto, 1	Bias 0.19–1.39 mm
Tahoces 2021 [95]	140 scans	Mixed, incl. post-repair	Root centre and plane	VGG16 + Faster R-CNN	Manual, 2	Error 6.6 mm
Yang 2023 [96]	162 pts	No severe aortic disease [b]	D, length, volume	3D VB-Net	Manual, 2	NR

Summary: maximum diameter, unweighted median [range], number of studies: absolute bias 0.26 [0–1.3] mm, 8; half-width of the 95% limits of agreement 5.2 [2.4–8.5] mm, 7; mean absolute error 1.5 [0.6–2.5] mm, 8. Studies reporting values only by site or segment, other measurands, or no agreement statistic are excluded. Rows are grouped by condition and sorted alphabetically. Data: study sample as patients (pts) or scans; "+" adds an independent or training dataset. Reference: method and number of readers. Agreement: one statistic per study for the main measurand (maximum diameter or total volume), chosen in this order: bias with 95% limits of agreement, mean absolute error, ICC, correlation coefficient; ranges span sites or segments; external test value in brackets (ext) when available. Full extraction in Table S4. [a] Public dataset. [b] Non-contrast CT included. [c] Commercial software: PRAEVAorta and PRAEVAorta 2 (Nurea, Bordeaux, France); ARVA (Incepto Medical, Paris, France); AI-Rad Companion (Siemens Healthineers, Erlangen, Germany); AVIEW Aorta (Coreline Soft, Seoul, Republic of Korea). [d] Reproducibility only (repeat readings or serial scans); agreement with the reference not reported. [e] Agreement between acquisition phases, arterial-phase AI as reference. [f] No agreement statistic; 95.5% of measurements judged correct on visual review. Abbreviations: AAA: abdominal aortic aneurysm; AD: aortic dissection; ARVA: Augmented Radiology for Vascular Aneurysm (Incepto Medical, Paris, France); ATAA: ascending thoracic aortic aneurysm; D: diameter; Dmax: maximum diameter; DQN: deep Q-network; DRL: deep reinforcement learning; EVAR/FEVAR/TEVAR: endovascular/fenestrated endovascular/thoracic endovascular aortic repair; ICC: intraclass correlation coefficient; IMH: intramural haematoma; LoA: limits of agreement; MAE: mean absolute error; NCCT: non-contrast CT; NR: not reported; RC: repeatability coefficient; r: correlation coefficient; SANet: shallow attention network; TAA: thoracic aortic aneurysm; TAAD/TBAD: type A/type B aortic dissection; TL/FL: true/false lumen; TotalSegmentator: open-source software (University Hospital Basel, Basel, Switzerland); uTBAD: uncomplicated TBAD; θ/φ: polar and azimuthal angles of the measurement plane normal (degrees); Rel. error: relative volume error (%); Error (Tahoces 2021 [95]): mean Euclidean distance between automated and expert aortic root landmark. (Tahoces 2021 [95]).

Table 3. Studies on AI-based aortic detection ($n = 38$).

Study	Data	Target	CT	Unit	Se/Sp	External Se/Sp	AUC (ext)
Aortic dissection or acute aortic syndrome ($n = 22$)							
Chang 2025 [97]	1275 scans	AAS	CTA	Exam	NR [g]	0.94/0.93	NR (0.96)
Cheng 2024 [98] [r]	320 pts	AD	NCCT	Patient	0.92/0.96	0.90/0.94	0.90 [b]
Dong 2024 [99] [r]	480 pts	AD	NCCT	Image	0.98/0.98	0.90/0.93	0.99 (0.94)
Gao 2026 [100]	300 pts	AD	NCCT	Box	NR [c]	–	NR [c]
Harris 2019 [101]	778 + 118 + 34,689 scans	AD, rupture	CE-CT	Exam	0.90/0.94	0.88/0.96	0.98 (NR)
Hata 2021 [102] [r]	170 pts	AD	NCCT	Exam	0.92/0.88	–	0.94
Hu 2025 [103] [r]	24,100 pts [d]	AAS	NCCT	Patient	0.98/0.95	0.95–0.98/0.93–0.95	0.98 (0.94–0.97)
Huang 2022 [104]	428 pts	AD, type A/B	CTA	Patient	0.96/1.00	A 0.95/0.99; B 0.79/0.94	0.98 (NR)
Kim (TB) 2024 [105]	121 pts	AD, IMH	NCCT	Image	NR [c]	–	NR [c]
Laletin 2024 [106]	1303 scans	AD	CTA	Exam	0.94/0.97	NS	0.96
Liu 2024 [107]	814 pts	AD, type A/B	CTA	Patient	A 0.97/0.98; B 0.95/1.00	–	NR
Norajitra 2025 [108] [r]	263 pts + 138 scans [a]	AD	CE-CT	Exam	0.97/1.00	0.92/1.00	0.99 (0.97)
Raj 2024 [109]	195 pts + 1189 scans [a]	AD	CE-CT	Exam	0.88/0.69	1.00/0.87	0.89 (0.99)
Tang 2026 [110] [r]	1138 pts	AD	NCCT	Patient	0.94/0.91	0.98/0.94	0.96 (0.97)
Wada 2023 [111]	436 pts	AD, IMH, rupture	CTA	Patient	NR [g]	0.99/0.88	NR (0.995)
Wang 2025 [112]	701 pts	AD	NCCT	Patient	0.76/0.96	0.88/0.91	0.94 (0.94)
Wu 2026 [113]	197 pts	TBAD	CTA	Patient	NR [g]	1.00/0.95	NR [c]
Xiong 2022 [114] [r]	237 pts	AD	NCCT	Patient	0.94/0.73	0.89/0.73	0.80 (0.86)
Yang 2026 [115]	493 pts	AAS, 4-class	CTA	Patient	NR [c]	–	0.9
Yi 2022 [116] [r]	452 pts	AD	NCCT	Patient	0.86/0.92	0.98/0.55 [e]	0.95 (0.97)
Zhang 2026 [117]	464 pts	AD	NCCT	Patient	0.93/0.90	0.89/0.80	0.94 (0.85)
Zhuang 2025 [118]	386 pts	AD, type A/B	CTA	Patient	0.98/0.97	–	0.995
Abdominal aortic aneurysm ($n = 8$)							
Camara 2022 [119]	400 pts	AAA	CTA	Image	0.99/0.99	–	0.99
Chatterjee 2024 [120]	66 scans [a] + 9172 pts	AAA, screening	NCCT	Exam	NR [g]	0.96/0.96	NR (0.99)
Chen 2025 [121] [r]	4023 scans	Incidental AAA	Mixed	Exam	NR [c]	NS	NR
Golla 2021 [122]	293 scans	AAA	CE-CT	Exam	0.90/NR	0.97/NR	0.93 (0.97)
Mohammadi 2019 [123]	10 pts	AAA, size class	Mixed	Patch	0.99/NR	–	NR
Rengarajan 2020 [124]	150 pts	Symptomatic or ruptured AAA	CTA	Patient	0.78/0.92	–	0.89
Spinella 2023 [125]	73 pts	AAA	CTA	Patient	NR [g]	0.98/0.96	NR
Tomihama 2023 [126]	400 pts	AAA	CTA	Image	0.99/0.99	–	0.999
Endoleak after EVAR ($n = 4$)							
Coatsaliou 2024a [127] [r]	56 pts	Endoleak	CTA	NS	0.90/0.91	NS	NR
Hahn (S) 2020 [128]	191 pts	Endoleak	CTA	Exam	0.90/0.88	–	0.94
Talebi 2020 [129] [r]	70 pts	Endoleak	CTA	Patient	1.00/NR	–	0.99
Yang 2024 [130]	167 pts	Endoleak	NCCT	Patient	0.88/0.88	–	0.95
Thoracic aortic aneurysm or dilatation ($n = 2$)							
Macruz 2022 [131] [r]	1715 scans	TAA, ascending/ descending	Mixed	Exam	Asc. 0.92/0.67; desc. 0.88/0.57 [f]	Asc. 0.93/0.81; desc. 0.91/0.65 [f]	NR
Pradella 2022 [132] [r]	18,243 scans	Unreported dilatation	Mixed	Exam	–	NR	NR
Mixed conditions ($n = 2$)							
Baltaci 2025 [133]	126 pts [a]	AAA, AD, 3-class	CT, NS	Image	0.87–0.88/NR	–	NR
Fujiwara 2026 [134] [r]	350 scans	Thoracic, abdominal, iliac aneurysm	NCCT	Lesion	NR [g]	0.81/NR	NR

Summary: aortic dissection or acute aortic syndrome, patient or examination level, unweighted median [range], number of studies: internal AUC 0.96 [0.80–0.995], 14; external AUC 0.97 [0.85–0.995], 9; external sensitivity 0.93 [0.88–1.00] and specificity 0.92 [0.55–1.00], 12. Endoleak: AUC 0.95 [0.94–0.99], 3. Image-, patch-, box- or lesion-level results, multi-class AUC, ranges across cohorts or classes, and studies without the metric are excluded. Rows are grouped by condition and sorted alphabetically. Data: study sample as patients (pts) or scans; "+" adds an external or public dataset. Unit: level at which sensitivity and specificity were computed. Se/Sp and AUC: internal test set or cross-validation; external test (independent centre, public dataset, or temporal/independent same-centre cohort) in the External Se/Sp column and in brackets for AUC; "–", no external test; two decimals, three where rounding would give 1.00; ranges span external cohorts or classes. Full extraction in Table S5. [r] Comparison with radiologists or with the original radiology report. [a] Public dataset. [b] Evaluation set not stated. [c] Metric not reported; alternative metrics: Gao 2026 [100], mAP 0.34; Kim (TB) 2024 [105],

mAP 0.95; Wu 2026 [113], external FROC-AUC 0.996; Yang 2026 [115], accuracy 0.75; Chen 2025 [121], PPV 0.31 for discordant cases. [d] Development and external cohorts; real-world and prospective cohorts (175,691 scans) not included. [e] External test with thick slices (mean 7.25 mm); readers' performance also decreased. [f] Reference: original radiology report. [g] External test only; internal results not reported. Abbreviations: AAA: abdominal aortic aneurysm; AAS: acute aortic syndrome; AD: aortic dissection; AUC: area under the ROC curve; CE-CT: contrast-enhanced CT (non-angiographic or any phase); CTA: CT angiography; EVAR: endovascular aneurysm repair; IMH: intramural haematoma; Mixed: contrast-enhanced and non-contrast CT; NCCT: non-contrast CT; NR: not reported; NS: not stated; Se/Sp: sensitivity/specificity; TAA: thoracic aortic aneurysm; TBAD: type B aortic dissection.Use of generative artificial intelligence. During the preparation of this manuscript, the authors used Elicit (accessed September 2026) to support structured extraction of study-level data. Elicit was not used for study selection, task classification, quality appraisal, or interpretation. All extracted data were verified against the primary publications by the authors, who take full responsibility for the content of the publication.

3. Results

3.1. Study Selection, Characteristics of Included Studies, Pathology and CT Technique

Database searching retrieved 944 records: 477 from PubMed/MEDLINE and 467 from Europe PMC. After removal of 341 duplicates, 603 records were screened on title and abstract, and 418 were excluded. Of the 185 reports sought for retrieval, one could not be obtained. The remaining 184 were assessed in full text, and 67 were excluded. The most frequent reason for exclusion was a prognostic or predictive endpoint ($n = 30$). This was followed by the absence of an AI model operating on the image ($n = 15$) and a non-CT or hybrid modality ($n = 5$). All reasons are reported in Figure 1 and Supplementary Table S2. Citation searching identified two further reports: one was included, and one was excluded as a surrogate simulation model. Overall, 118 studies were included (Tables 1–3). Of these, 42 were assigned to segmentation, 38 to measurement, and 38 to detection or classification (Figure 1).

The included studies were published between 2018 and 2026, and 60 (51%) appeared from 2024 onward. Most were retrospective clinical studies ($n = 101$); 15 were methodological studies based on public datasets, used alone or together with local data, and two did not state their design. Prospective evaluation was limited to a four-week clinical deployment of one detection system [101] and a prospective comparative and pilot implementation phase of another [103]; three further studies analysed data from a prospective registry, cohort, or trial retrospectively [66,71,78]. Research activity was concentrated in Europe ($n = 42$), the United States ($n = 27$), and China ($n = 25$). Thirty-one studies drew on data from two or more institutions; the remaining 87 were single-centre, although several supplemented local data with public datasets. Sample sizes ranged from 10 [123] to more than 24,000 patients [103], with the unit of enrolment varying between patients, scans, and images; 33 of the 42 segmentation studies analysed 200 or fewer patients or scans, whereas the largest samples came from detection and opportunistic measurement studies.

Aortic dissection or acute aortic syndrome was the target condition in 40 studies (14 segmentation, four measurement, 22 detection), AAA in 45 (21, 12, and 12, including four on endoleak after EVAR), thoracic aneurysm or dilatation in 12 (two, eight, and two), and mixed or unselected cohorts, including healthy subjects, in 21 (five, 14, and two). Dissection therefore dominated detection, AAA dominated segmentation, and mixed or population cohorts were concentrated in measurement. Twenty-three studies included postoperative examinations after EVAR, FEVAR, TEVAR, or surgical repair, 12 of them exclusively, and all but one of these addressed AAA after EVAR. Eighty-one studies analysed contrast-enhanced CT only, predominantly CTA; 19 analysed non-contrast CT only, 15 analysed both, and three did not state the use of contrast medium. Non-contrast CT was concentrated in detection: 14 of the 19 studies based exclusively on non-contrast CT were detection studies, 11 of them targeting aortic dissection with same-session contrast-enhanced CT as the reference. ECG-

gated or ECG-triggered acquisition, in all or part of the sample, was reported in 16 studies, mostly thoracic measurement studies.

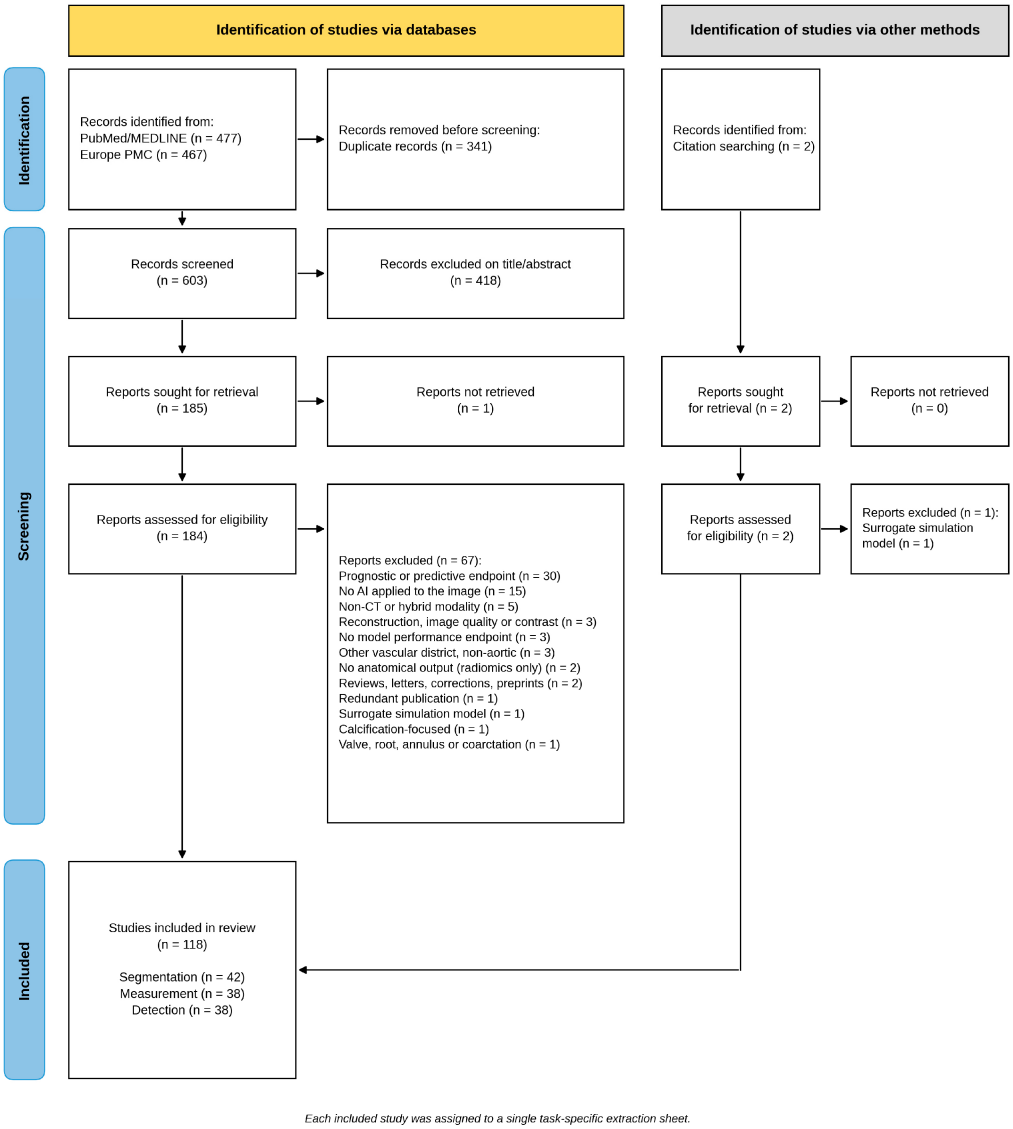


Figure 1. PRISMA 2020 flow diagram of study identification and selection.

3.2. AI Tasks, Architectures and Validation

Each study was assigned to a single primary task, but the tasks were closely interdependent. Segmentation was the usual intermediate step in measurement (35 of 38 studies), the remaining three localising anatomical landmarks or measurement planes directly [59,62,95], and about half of the detection studies (21/38) included a segmentation stage; in nine of these, the detection decision was a rule applied to a segmentation-derived quantity, a maximum diameter threshold for aneurysm or a false lumen or regional volume for dissection. Most models were developed within the study ($n = 82$); 21 studies evaluated commercial software, two applied the open-source pre-trained TotalSegmentator, and 13 used mixed pipelines in which a non-learned component, such as centreline extraction, ellipse fitting, a Hough transform, or multi-atlas zoning, produced part of the output. Commercial software was concentrated in measurement (16/38 studies, against one in segmentation and four in detection), and three platforms, PRAEVAorta ($n = 9$), AI-Rad Companion ($n = 7$), and ARVA ($n = 2$), accounted for 18 of the 21 commercial evaluations. Among developed models, U-Net-derived encoder–decoder networks predominated, and the self-configuring

nnU-Net framework was used in 20 studies across all three tasks; transformer components appeared in seven studies, mostly as hybrid CNN–transformer segmentation networks. Detection relied on image classification CNNs and on object detection networks of the YOLO, R-CNN, and RetinaNet families. Deep reinforcement learning for landmark localisation was used in nine studies, seven of them through AI-Rad Companion, and generative adversarial networks in three studies to synthesise contrast-enhanced-like images from non-contrast CT.

The validation designs of the included studies are described in Section 3.6 and Table 4.

Table 4. Validation design of the included studies by primary task. Values are numbers of studies.

Validation Design	Segmentation (n = 42)	Measurement (n = 38)	Detection (n = 38)	All (n = 118)
Geographically external test of a model developed in the study (other institutions or public datasets)	3	3	14	20
Temporally separated or otherwise independent cohort from the same centre	2	0	6	8
Pre-trained commercial or open-source tool applied to data declared independent of its training	1	10	1	12
Any external validation	6 (14%)	13 (34%)	21 (55%)	40 (34%)
Partly external test (part of the test set from a centre absent from training)	0	1	0	1
Independence of test data from training not established [a]	0	10	3	13
Internal hold-out or cross-validation only	36	14	14	64

[a] Measurement: [63,65,68,72,73,75,82,84,86,91]; detection: [106,121,127]; 11 of the 13 evaluated commercial or manufacturer-linked software whose training data were not reported.

3.3. Results by Primary Task: Segmentation-Focused Applications

The 42 segmentation studies are summarised in Table 1. Aortic dissection was the target condition in 14, AAA in 21, thoracic aneurysm or dilatation in two, and mixed or unselected cohorts in five. The primary structure was the outer aortic wall or whole aorta in 12 studies, ILT in eight, the non-dissected lumen in five, the true lumen in two, the whole lumen of a dissected aorta in two, false lumen thrombus in three, and the false lumen in one; in seven further studies two or more classes had equal rank, and two reported only a mean across many classes. The Dice similarity coefficient (DSC) was reported in 38 studies; two reported only intersection over union (IoU), one only the Dice loss, and one had no reference segmentation.

Summarised within comparable compartments, median DSC was highest for the outer wall (0.96, range 0.89–0.99; 11 studies) and the non-dissected lumen (0.92, 0.86–0.97; 8). In dissection, median DSC was 0.92 for the true lumen (0.87–0.96; 6) and 0.90 for the false lumen (0.88–0.95; 5). It was lower for ILT (0.86, 0.80–0.94; 8) and markedly lower for false lumen thrombus (0.63, 0.52–0.90; 4). Distance metrics were not summarised: they were absent from the main text of 15 studies, the 95th-percentile Hausdorff distance (HD95) was reported in only nine, and in six studies the units were not stated.

In dissection, the luminal compartments were segmented consistently across 2D and 3D architectures, with DSC of 0.87–0.96 for the true lumen and 0.89–0.95 for the false lumen [17–19,21,22,27,30]. For the outer surface, 2D and 3D V-Nets both achieved DSC 0.96, yet the authors judged this accuracy insufficient for unsupervised use [26]. False lumen thrombus was the least tractable target, with DSC 0.52–0.63 in three studies [25,28,29]; a transformer–U-Net cascade that maintained luminal performance on external ImageT-

BAD data (DSC 0.94 and 0.92) reached only 0.59 for thrombus [25], the exception being postoperative type-A dissection (DSC 0.90) [20]. Multi-class segmentation of branches and zones reached a mean DSC of 0.73 across 14 classes [23] and 0.78 across 23 classes in the AortaSeg24 challenge, with DSC decreasing with structure size [24].

In AAA, overlap for the lumen and outer wall was consistently high (DSC 0.92–0.97) [33,36,37,42,47,48], and nnU-Net tested only on three centres not used in training reached DSC 0.97 with HD95 1.4 mm [37]. Two studies showed how performance can shift beyond the development data: IoU fell from 0.98 internally to 0.88 at an external centre [31], and a 3D U-Net trained on 10–20 masks reached an out-of-sample DSC of only about 0.26 [49]. ILT was consistently harder to segment than the lumen, with DSC 0.80–0.94 against 0.95–0.97 for the lumen in the same models [35,46] and 0.87–0.89 in ILT-focused models [32,43,51]; after EVAR, the stent graft and altered sac anatomy lowered ILT overlap further (DSC 0.82–0.87) [38–40], although coordinate-channel 3D CNNs reached an outer-wall DSC of 0.97 on both CTA and non-contrast CT, including the post-EVAR subset [50]. Annotation burden and non-contrast CT were addressed by training on synthetic images only (DSC 0.84 against 0.88 with real scans) [41], by weak supervision from pseudo-labels, which matched fully supervised training while reducing annotation time from 84.6 to 15.2 h [45], by a cycle-consistent GAN that synthesised contrast-enhanced-like images from non-contrast CT (lumen DSC 0.86 internally and 0.84 externally) [34], and by a feature-fusion network applied directly to non-contrast CT (DSC 0.89) [44].

In the ascending aorta, a slice selection CNN with a 2D U-Net reached DSC 0.99 on screening chest CT, computed only on classifier-selected slices [52], and ENet reached DSC 0.91 on ECG-gated CTA [53]. Whole-aorta segmentation across normal and diseased anatomy reached DSC 0.89–0.95 [54,56–58], with a consistent drop in pathological subsets, for example from 0.91 on healthy public data to 0.85 on a private AAA cohort [58]. TotalSegmentator, evaluated without a reference segmentation by reader-based scoring anchored to the contour error in millimetres, scored lower in dissection and aneurysm than in young patients without aortic disease, with systematic failure modes: in aneurysms, it followed the contrast-filled lumen and omitted thrombus and outer wall, and in dissection it did not recognise the intimal flap [55].

3.4. Results by Primary Task: Measurement and Geometric Analysis Applications

The 38 measurement studies are summarised in Table 2. Aortic dissection was the target condition in four, AAA in 12, thoracic aneurysm or dilatation in eight, and mixed or unselected cohorts in 14. The geometric outputs were dominated by aortic diameter (31/38), usually the maximum diameter perpendicular to the centreline and, in 11 studies, reported at standardised guideline positions. Volume was reported in 15 studies, most often total aneurysm sac, lumen, and thrombus volume below the lowest renal artery, or zonal volume in dissection; other outputs included landmark positions, measurement plane orientation, and EVAR and TEVAR planning parameters such as neck and landing zone diameters, lengths, angulation, and tortuosity. Longitudinal change between serial examinations was assessed in eight studies. The reference was manual measurement in 25 studies and semi-automatic measurement in nine; in the remaining four it was a core laboratory, visual review, a manually corrected AI output, or the AI output on another acquisition phase. Agreement was expressed as bias with 95% limits of agreement (LoA), mean absolute error (MAE), intraclass correlation (ICC), or correlation coefficients; no agreement statistic against the reference was reported in four studies.

For maximum diameter, bias was small, with a median absolute bias of 0.26 mm (range 0–1.3 mm; 8 studies), but the LoA were wide, with a median half-width of 5.2 mm (2.4–8.5 mm; 7 studies); the median MAE was 1.5 mm (0.6–2.5 mm; eight studies). Where

inter-reader variability was measured in the same study, the AI error was generally of the same order as the disagreement between human readers [68,69,87,93].

In AAA, a hybrid expert system and U-Net pipeline reached a median MAE of 0.8 mm, against an inter-operator MAE of 1.2 mm [68], and in an external multicentre evaluation of PRAEVAorta 2 against a core laboratory the bias was 0.28 mm (LoA −4.36 to 4.93 mm), with 95.4% of measurements within 5 mm but 10% of scans excluded after visual review for evident segmentation errors [66]. In complex aneurysms before and after fenestrated repair, ARVA reached a median MAE of 2.4 and 1.6 mm, against an inter-reader error of 0.8–1.1 mm, and consistently underestimated the diameter in the 11% of outliers [73]. EVAR planning pipelines reported a diameter bias below 0.4 mm but LoA of up to ±7.8 mm and weaker agreement for lengths [67,70], and PRAEVAorta identified 99.5% of visceral vessels with a diameter bias matching inter-observer agreement [69]. Volumetric surveillance after EVAR was the most distinctive application: PRAEVAorta reached $r > 0.99$ for sac volume with an analysis time of 2.5 versus 22 min per patient [64], and an nnU-Net model maintained r 0.96 on external CTA [74]. PRAEVAorta 2, however, reached an ICC of only 0.85 for volume change at one year, and its classification into regression, stability, or growth agreed with the semi-automatic reference in only 37 of 49 patients (weighted kappa 0.33) [72]. Automated volume classified more sacs as growing than diameter did, and most discordant cases had an endoleak [63]; fully automated measurement yielded complete agreement on the indication for repair, against kappa 0.53–0.70 for semi-automatic measurement [71]. Measurements from venous and non-contrast phases differed from the arterial-phase output of the same software by more than 5 mm in 16–17% of diameters and by more than 5% in 26–32% of volumes [65].

In the thoracic aorta, four studies evaluated AI-Rad Companion at guideline measurement positions. Bias was close to zero, but the LoA reached ±11 mm at the sinus of Valsalva [81] and exceeded ±5 mm at the root and at every level on non-contrast CT [76]; a third of measurements required correction, mostly at the root, although AI assistance reduced reading time from 13:01 to 4:46 min and inter-reader variability from 1.16 to 0.42 mm [82], and after retraining 95.5% of 8948 measurements were judged correct on visual review [79]. Developed multitask 3D U-Nets combining segmentation with landmark detection reached an MAE of 1 mm or less at six of nine standard positions [77] and a bias of 0.04 mm, with 96% of measurements accepted on human-in-the-loop validation [80]; error was consistently largest at the sinus of Valsalva, the sinotubular junction, and the arch [77,80]. Wider LoA were reported for the ascending aorta (−5.30 to 7.13 mm) [78] and on paired contrast-enhanced and non-contrast CT (−9.8 to 7.2 mm), with lower agreement in the descending aorta and after repair or dissection [75].

In dissection, measurement moved from diameter to compartmental volume and landmarks. Zonal volumetry reached DSC of 0.51–0.91 across 14 SVS/STS zones without comparison against a reference [60], and relative volume error reached 136% in the false lumen, concentrated in distal segments with small or thrombosed false lumina [61]; a regression CNN predicted the orientation of the double-oblique measurement plane with MAE 5.8° and 16.6° [62], and deep reinforcement learning localised nine landmarks with a mean error of 3.4 mm, against 3.9 mm between observers [59]. In mixed cohorts, ARVA reached a median MAE of 1.2 mm, rising to 1.9 mm in diseased and 2.2 mm in post-endovascular aortas [83], and 1.6 mm in a larger two-centre test set, identical to that of clinicians [93]; AI-Rad Companion reached ICC 0.96–0.98 on consecutive and pathological thoracic CT [84]. On low-dose non-contrast CT from a population cohort, the bias was 0 mm and the MAE 1.3 mm, against an inter-reader MAE of 1.4 mm [87], whereas on lung cancer screening CT agreement was poor at the arch, where AI overestimated by 5.2 mm [91]. TotalSegmentator with lumen refinement measured more than 19,000 CTA with an MAE

of 0.97–2.32 mm by site [92], but without adaptation it overestimated diameters, more on non-contrast CT, which the authors attributed to wall inclusion [89]; after endovascular repair, PRAEVAorta 2 reached a median MAE of 2.5 mm on CTA and 1.4 mm on non-contrast CT, but at least one measurement was missing in two thirds of the non-contrast cases [86]. The remaining studies generated zone-specific reference values from 704 healthy aortas [90], measured the descending aorta in trauma patients (MAE 2.11 mm) [88], halved measurement variability on serial CTA of stable aortas [85], measured TEVAR landing zones with small diameter bias but larger discrepancies in lengths and angles [94], localised the aortic root centre with an error of 6.6 mm against 4.6 mm between experts [95], or reported agreement only graphically [96].

3.5. Results by Primary Task: Detection and Classification Applications

The 38 detection and classification studies are summarised in Table 3. They addressed aortic dissection or acute aortic syndrome (22 studies), AAA (8), endoleak after EVAR (4), thoracic aneurysm or dilatation (2), and mixed conditions (2). Sensitivity and specificity were computed at patient ($n = 17$), examination ($n = 12$), or image level ($n = 5$), at the level of bounding box, patch, or lesion in one study each, and at an unstated level in one; image-level results were therefore not pooled with patient-level results. Fourteen studies compared the model with radiologists, nine on the same images and five against the original radiology report.

For aortic dissection or acute aortic syndrome at patient or examination level, the median internal AUC was 0.96 (range 0.80–0.995; 14 studies). On external testing, the median AUC was 0.97 (0.85–0.995; 9), with median sensitivity 0.93 (0.88–1.00) and specificity 0.92 (0.55–1.00; 12 studies).

On contrast-enhanced CT, models ranged from binary identification to pipelines integrating segmentation and Stanford subtyping. Two such pipelines reached sensitivity 0.95–0.97 and specificity 0.98–0.99 for type A dissection, with sensitivity 0.79 for type B in one of them [104,107], a third reached AUC 0.995 in cross-validation only [118], and two models applying a volume rule to segmentations reached an external AUC of 0.97–0.99, although sensitivity fell to 0.81 in atypical presentations [108,109]. Triage-oriented evaluations were the closest to clinical use. In a four-week prospective deployment in a teleradiology network, sensitivity was 0.88 and specificity 0.96, and model-positive studies were opened sooner (median 265 versus 660 s) [101]. A commercial device evaluated as a standalone tool on 1303 consecutive CTA reached sensitivity 0.94, specificity 0.97, and AUC 0.96, with all eight false negatives being type-B dissections [106], and a cascaded model for acute aortic syndrome tested at 14 US sites reached sensitivity 0.94, specificity 0.93, and AUC 0.96, with lower sensitivity for penetrating ulcer (0.83) [97]. Two studies departed from binary detection, localising dissection components in known type-B dissection [113] or classifying four acute aortic syndrome categories from imaging and blood tests with accuracy 0.75 [115].

Eleven studies detected dissection on non-contrast CT, mostly using same-session contrast-enhanced CT as the reference, with performance comparable to or better than that of radiologists reading the same non-contrast images: AUC 0.94 for classification and false lumen volume approaches [102,112], external sensitivity 0.90 and specificity 0.94 against three radiologists [98], and, for YOLO-based detectors, sensitivity 0.90 against 0.70–0.72 [99] and AUC 0.97 against 0.77–0.82 for radiologists [110]. External performance was not uniformly maintained: on an external set with thick slices, specificity fell to 0.55, as did the sensitivity of radiologists [116]; a multitask GAN that synthesised contrast-enhanced images reached external AUC 0.86 with specificity 0.73 [114]; and a model distilled from a contrast-enhanced teacher network fell from AUC 0.94 in cross-validation to 0.85 externally [117]. The largest programme developed an nnU-Net-based multitask

model on data from nine centres and more than 24,000 patients, reaching internal AUC 0.98 and external AUC 0.94–0.97; in a prospective comparative study it shortened the time to the correct diagnostic pathway in patients with an initially incorrect suspicion, from 219.7 to 61.6 min [103].

For AAA, image-level classifiers reported the highest accuracy, up to sensitivity and specificity 0.99 [119], but from image-level splits [119,123,133]; reanalysed on a selected set, the sensitivity of the same model fell to 0.88, with errors concentrated in small aneurysms with mural thrombus [126]. Screening pipelines applying a diameter threshold to automated segmentation reached sensitivity 0.98 and specificity 0.96 across six centres [125] and 0.96 and 0.96 on 9172 CT colonography examinations, but with a positive predictive value of 13% [120]. A commercial tool applied to emergency abdominal CT identified five unreported AAA, although only discordant cases were verified [121], and a model on geometric and biomechanical descriptors distinguished symptomatic or ruptured from asymptomatic AAA with AUC 0.89, outperforming maximum diameter alone [124]. For endoleak after EVAR, AUC was 0.94–0.99 on CTA [128,129] and 0.95 on non-contrast CT with a manually segmented sac [130]; a rule applied to commercial segmentation reached sensitivity 0.89 against 0.68 for the original reports, which missed mainly low-flow type-II endoleaks [127]. For thoracic aneurysm, a 3D U-Net on 1400 consecutive routine chest CT reached sensitivity 0.93 and specificity 0.81 for ascending and 0.91 and 0.65 for descending aneurysm against the radiology report [131], and AI-Rad Companion applied to 18,243 CT reported as normal flagged 5.5% as dilated, with a positive predictive value of 45% corresponding to 452 unreported dilatations, while negative results were not verified [132]. A whole-aorta pipeline on non-contrast CT detected thoracic, abdominal, and iliac aneurysms with lesion-level sensitivity 0.81 against the report, only 0.56 for iliac aneurysms [134].

3.6. Validation and Reporting Patterns Across Tasks

External validation, defined as testing on data that had not contributed to training, was reported in 40 studies, and in part of the test set of one further study [93]; the remaining 64 studies relied on internal hold-out testing or cross-validation, and in 13 studies, 11 of them evaluating commercial or manufacturer-linked software, independence of the test data from training could not be established. The proportion of studies with any external validation differed markedly by task, from 6/42 (14%) in segmentation to 13/38 (34%) in measurement and 21/38 (55%) in detection, but so did its form (Table 4). Of the 40 studies, 20 tested a model developed in the study on data from other institutions or on public datasets, eight used a temporally separated or otherwise independent cohort from the same centre, and 12 applied a pre-trained commercial or open-source tool to data declared independent of its training. Geographically external testing of a developed model accounted for most external validation in segmentation (three of six) and detection (14 of 21), whereas in measurement 10 of the 13 externally validated studies were evaluations of commercial software.

Performance generally held when a model was tested outside its development data, but not uniformly: overlap was maintained in some external segmentation tests [25] and fell substantially in others [31], and some detection models lost AUC or specificity with different slice thickness or case mix [116,117]. The reference standard depended on the task: manual or semi-automatic annotation, often by one or two readers and in some cases a manually corrected AI output, for segmentation and measurement; same-session contrast-enhanced CT for non-contrast detection; and the original radiology report, sometimes with verification limited to model-positive or discordant cases, for opportunistic and population studies [121,132]. Processing failures and exclusions were reported inconsistently but were not negligible where stated, ranging from 7% of examinations not processed [132] and

10–23% of scans excluded or flagged as outliers after visual review [66,72,73] to 15% of examinations with incomplete measurements [81] and two thirds of non-contrast cases with at least one missing measurement [86]. Workflow time against a human or semi-automatic comparator was reported in 27 studies, but the comparator was timed within the study in only 13; where measured, automated analysis reduced processing time several-fold, for example, from 35 min to 90 s for AAA segmentation [33], from 82.5 to 21.7 min for dissection segmentation and measurement [30], and from 22 to 2.5 min for post-EVAR volumetry [64]. Studies of AI assistance to readers, rather than standalone performance, were few: they showed reduced reading time and inter-reader variability [82], human-in-the-loop acceptance of 96% of measurements [80], a shorter time to the correct diagnostic pathway in acute aortic syndrome [103], and additional aneurysms identified on AI-assisted re-reading [134].

Subgroup summaries are reported in Supplementary Table S9. In segmentation, subdividing the studies beyond the anatomical compartment left most groups with one to four studies, and only four of the studies contributing to the compartment medians reported an external test; the compartment gradient was preserved within every group. In measurement of maximum diameter, the median absolute bias was below 0.6 mm throughout, whereas the median half-width of the limits of agreement was wider in studies without a stated external test (7.0 versus 4.7 mm) and in samples above 200 patients (5.7 versus 4.3 mm). In detection of aortic dissection or acute aortic syndrome, external AUC, sensitivity and specificity were lower on non-contrast than on contrast-enhanced CT (median external specificity 0.85 versus 0.94), whereas the internal AUC differed little between samples of ≤500 and >500 patients (0.94 versus 0.96). Given the small number of studies in each group, these are descriptive observations.

3.7. Quality Appraisal

Quality appraisal was matched to the primary task (Figure 2) and is detailed in Supplementary Tables S6–S8. Among the 80 segmentation and measurement studies, CLAIM-informed limitations were lower in 10 (13%), moderate in 50 (63%), and higher in 20 (25%), with a similar distribution across the two tasks. The higher tier was driven by the reference standard in 13 studies (reference generated or corrected by the authors' own software or by the model, or no quantitative agreement statistic) and by data independence or selection in seven (image- or scan-level partitioning, parameters set on the test data, or evaluation on selected slices or cases). The recurrent limitations were a reference standard produced by a single reader or operator, with reader variability rarely documented and, in some cases, generated or corrected with the authors' own or proprietary software [34,41,47,48,65,90], and validation confined to cross-validation or small internal test sets, occasionally partitioned at slice or scan level [54,56,77] or restricted to selected slices or regions [44,52]. Detection studies were appraised with a different instrument and construct, and their categories are not comparable with the CLAIM-informed tiers. Among the 38 detection studies, the overall QUADAS-2/QUADAS-AI risk of bias was high in 33 (87%), low in three [103,104,112], and unclear in one, because independence of the test data from the vendor's training data was not stated [106]; one protocol paper reported no diagnostic accuracy and was not formally appraised [100]. The dominant concern was patient selection: 22 studies used case–control, matched, or enriched sampling, or drew positive and negative cases from different sources, a design that inflates prevalence and spectrum relative to the intended clinical use. Secondary concerns were a reference standard based on the original radiology report, with partial verification or incorporation of the report into the reference [101,120,121,128,132,134], data partitioning at image or patch level with possible leakage between training and test sets [105,119,123,126,133], and selection of operating

thresholds or models on data later used for testing [95,101,102,110]. The three low-risk studies still raised applicability concerns: a single-centre temporal test [104], a CTA-referred population [112], and few events in the prospective pilot implementation [103].

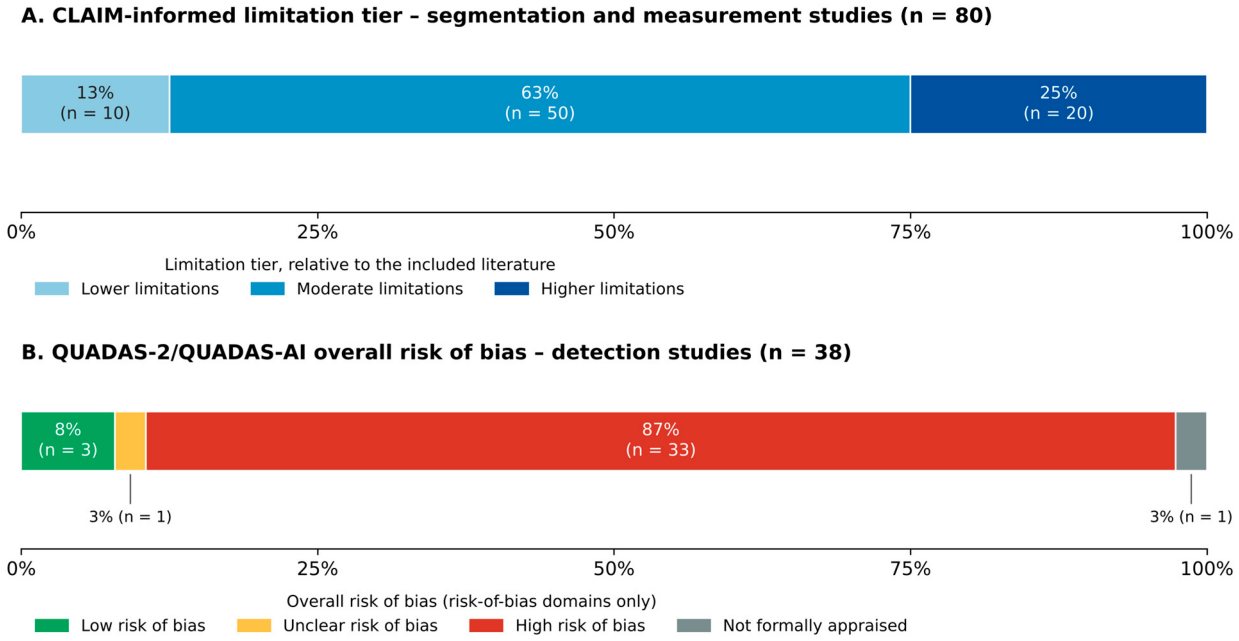


Figure 2. Quality appraisal by task. (**A**) CLAIM-informed limitation tiers for segmentation and measurement studies ($n = 80$), a descriptive classification of methodological and reporting limitations relative to the included literature (decision rule in Section 2.6 and Supplementary Table S10). (**B**) QUADAS-2/QUADAS-AI overall risk of bias for detection studies ($n = 38$), based on the risk-of-bias domains only. The two panels use different instruments and constructs and different colour scales; categories are not comparable between panels. Percentages are calculated within each panel.

4. Discussion

4.1. Principal Findings

This task-based systematic review identified 118 studies applying AI to CT or CTA of the thoracic and abdominal aorta: 42 on segmentation, 38 on measurement, and 38 on detection or classification. The field is recent and expanding, with half of the studies published from 2024 onward, but the evidence is unevenly distributed. Aortic dissection dominated detection (22 of 38 studies) and abdominal aortic aneurysm dominated segmentation (21 of 42). The three tasks were also closely interdependent. Of the 38 measurement studies, 35 derived their measures from an automated segmentation. About half of the detection studies included a segmentation stage, and in nine the detection rule was applied directly to a segmentation-derived quantity. Segmentation is therefore less an end point than the common substrate on which measurement and much of detection are built, and its errors may propagate downstream.

Across tasks, reported technical performance was high. Median DSC was 0.90 or higher for the outer wall, the non-dissected lumen, and the true and false lumina. The median absolute bias for maximum diameter was below 0.3 mm. For detection of aortic dissection or acute aortic syndrome, the median AUC was 0.96 internally and 0.97 on external testing. The maturity of this evidence, however, is limited: external validation of any kind was reported in 14% of segmentation, 34% of measurement, and 55% of detection studies, and geographically external testing of a model developed in the study in only three, three and 14 studies, respectively (Table 4); only two systems were evaluated prospectively in clinical use [101,103]; 33 of the 38 detection studies were at high risk of bias on QUADAS-

2/QUADAS-AI, and, on a different and descriptive scale, 70 of the 80 segmentation and measurement studies carried moderate or higher CLAIM-informed limitations. The gap between high reported performance and limited evidence of transportability is a relevant finding of this review (Figure 3).

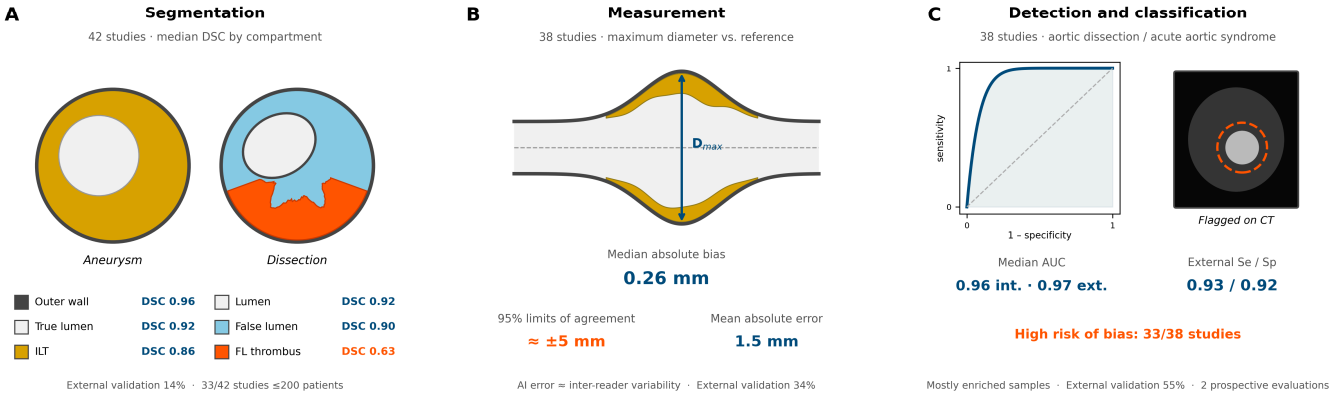


Figure 3. Summary of reported AI performance by primary task. (**A**) Segmentation: schematic cross-sections of an abdominal aortic aneurysm and of an aortic dissection, with the median Dice similarity coefficient (DSC) across studies for each anatomical compartment. (**B**) Measurement: agreement of automated maximum diameter (Dmax) with the reference standard, expressed as median absolute bias, median half-width of the 95% limits of agreement (LoA) and median mean absolute error (MAE). (**C**) Detection and classification: diagnostic accuracy for aortic dissection or acute aortic syndrome at patient or examination level (median AUC on internal and external testing; median external sensitivity and specificity, Se/Sp) and overall risk of bias. In (B), the schematic longitudinal section shows the aortic wall (dark grey lines), lumen (light grey area), intraluminal thrombus (ochre areas), centreline (dashed grey line) and maximum diameter (Dmax, blue double-headed arrow). In (**C**), the left schematic shows a receiver operating characteristic curve (blue line) with the area under the curve shaded and the chance diagonal (dashed grey line); the right schematic shows an axial CT section with the aorta flagged by the model (dashed orange circle). In all panels, values in orange indicate the least favourable findings. Schematics are illustrative and not drawn from study data. All values are unweighted medians across the studies reporting the metric (Tables 1–3); the proportion of studies with any external validation (Table 4) is given for each task. FL: false lumen; ILT: intraluminal thrombus.

These findings are consistent with previous syntheses and extend them. Meta-analyses in aortic dissection reported high pooled detection accuracy and segmentation overlap. They nevertheless rated the certainty of evidence as low, because datasets were retrospective, enriched, and predominantly single-centre [7–9]. The present review shows that the same limitations apply beyond dissection and beyond detection, to segmentation and measurement in aneurysmal disease, where external validation was even less frequent. Organising the evidence by task rather than by pathology also shows where evidence is scarce. Only two studies segmented thoracic aneurysm or dilatation, only four measured dissection, and endoleak detection rested on four small case–control studies.

4.2. Segmentation: Performance Depends on the Compartment

Segmentation performance depended more on the anatomical compartment than on the AI architecture. Median overlap declined from DSC 0.96 for the outer wall and DSC 0.92 for the lumen to DSC 0.86 for ILT and DSC 0.63 for false lumen thrombus. The same gradient appeared within individual multistructure studies [25,35,46]. The most challenging targets were false lumen thrombus and the postoperative anatomy after EVAR, where the stent graft and the altered sac add further sources of difficulty.

Two mechanisms plausibly contribute. First, the boundary itself is uncertain, for the reference reader as well as for the model. Thrombus has attenuation close to that of the aortic wall and periaortic tissue, and false lumen thrombus can be confused with delayed opacification of a patent false lumen. Second, DSC depends on structure size: the same boundary error penalises a small or thin compartment more than a large one. Across the 23 classes of the AortaSeg24 challenge, DSC declined with structure size [24]. Differences between compartments therefore reflect both true difficulty and a property of the metric. A single DSC averaged over the aorta, or over many classes, conceals the compartment that is often clinically decisive. This pattern agrees with the lower performance for true and false lumen than for whole-aorta delineation reported in dissection [7], and with the difficulties described for ILT segmentation [3].

A high DSC also does not guarantee an accurate boundary, which is what determines diameters, volumes and short segments such as landing zones. For the outer surface in type-B dissection, two models with the same DSC 0.96 differed in normalised surface Dice. The authors judged neither accurate enough for unsupervised use [26]. Distance metrics could not be summarised in this review: they were absent from the main text of 15 studies; HD95 was reported in only nine; units were not stated in six. For segmentation intended to support measurement, overlap should be complemented by a boundary metric with stated units, such as HD95 or surface Dice at a clinically meaningful tolerance.

Performance also depended on case mix. Whole-aorta models showed reduced overlap in pathological subsets [56–58]. A general purpose model evaluated without adaptation, by reader-based scoring rather than against a reference segmentation, failed systematically in disease [55]. In aneurysms it followed the contrast-filled lumen and omitted thrombus and outer wall, and in dissection it did not recognise the intimal flap. Models trained on normal anatomy or on public datasets should therefore not be assumed to transfer to aneurysmal or dissected aortas. The definition of the label, outer wall or lumen, also needs to be stated, because it changes both the metric and any measurement derived from it.

External validation was the exception (six of 42 studies) and gave divergent results. In one study, luminal overlap was maintained on an external public dataset, while false lumen thrombus remained poorly segmented [25]. In another, overlap fell from IoU 0.98 to 0.88 at an external centre [31]. Acquisition may explain part of this variation. Slice thickness and spacing affect segmentation accuracy [6], and some development datasets included slices of up to 5 mm [25].

Finally, the segmentation evidence rests on small datasets. Of the 42 studies, 33 analysed 200 or fewer patients or scans. Internal test sets were sometimes limited to 10–30 scans, and the reference was often drawn by a single reader. Annotation cost limits the size of datasets, and three approaches in the included studies address it: weak supervision matched fully supervised performance while reducing annotation time from 84.6 to 15.2 h [45]; training on synthetic images approached the performance of training on real scans [41]; image synthesis extended lumen segmentation to non-contrast CT [34]. These strategies may facilitate the development of larger multicentre training datasets, while not eliminating the need for a well-documented reference standard in the test set.

4.3. Measurement: Small Bias, Variable Individual Error

For maximum diameter, automated measurement was accurate on average but less precise for the individual patient. The median absolute bias was 0.26 mm, whereas the median half-width of the 95% limits of agreement was 5.2 mm (range 2.4–8.5 mm), with a median MAE of 1.5 mm. A negligible bias supports the use of automated measurement for cohorts and population-level analyses. For the individual patient, however, an error of this size is clinically relevant. In AAA, current guidelines recommend repair at a diameter of

55 mm in men and 50 mm in women, while a growth rate of ≥10 mm per year is considered an indication for repair irrespective of aneurysm diameter [135]. An error of ±5 mm is therefore enough to change the indication for a patient near a threshold and corresponds to half a year of rapid growth.

Where inter-reader variability was measured within the same study, the automated error was generally similar to the disagreement between readers [68,69,87,93]. The exception was complex aneurysms treated with fenestrated repair, where the automated error exceeded inter-reader variability [73]. The appropriate benchmark is therefore human variability, not an error-free reference. Automated measurement can replace a reader only where its error falls within that variability.

Deviations were not uniformly distributed but concentrated in predictable locations and acquisitions: in the thoracic aorta, it was largest at the sinus of Valsalva, the sinotubular junction, and the arch [77,81,91]; on non-contrast CT, the limits of agreement exceeded ±5 mm at every level in one study [76]; an unadapted general purpose model overestimated diameters more on non-contrast than on contrast-enhanced CT, consistent with inclusion of the wall [89]; measurements from venous and non-contrast phases differed from the arterial-phase output by more than 5 mm in 16–17% of cases [65]. These locations and acquisitions are where manual measurement is also most difficult; automated output there should be flagged for review rather than accepted by default.

Processing failures were reported inconsistently but were not negligible where stated. Between 10% and 23% of cases were excluded, flagged as outliers, or incompletely measured in several evaluations [66,72,73,81]. Agreement statistics computed only on successfully processed cases overestimate performance in practice.

For surveillance, reproducibility may matter more than agreement with a single reader, because management depends on change between examinations. On serial CTA of stable aortas, automated measurements showed approximately half the variability of manual measurements [85]. In untreated AAA, a fully automated pipeline gave identical repeated measurements and complete agreement on the indication for repair [71]. AI assistance also reduced inter-reader variability from 1.16 to 0.42 mm and reading time from 13:01 to 4:46 min [82]. The first two studies, however, assessed reproducibility without agreement against a reference, so a consistent but biased measurement cannot be excluded. Taken together, these findings support a role for automated diameter as a standardising tool within a supervised workflow, rather than as a replacement for expert measurement.

Volumetry was the most distinctive application, particularly after EVAR. Sac, lumen and thrombus volumes can capture changes that a single maximum diameter does not reflect. Automated sac volume agreed closely with the reference [64,74], including on external data. Agreement on a single volume, however, did not guarantee agreement on change. In one study, the ICC was 0.94 for sac volume but 0.85 for volume change at one year, and classification into regression, stability, or growth agreed with the reference in only 37 of 49 patients [72]. Volume and diameter also classified sac evolution differently, and most discordant cases had an endoleak [63]. Longitudinal change was assessed in only eight studies. Volumetry is therefore promising, but its clinical use requires agreement on change between examinations and validated thresholds for volume change, neither of which is yet established.

In aortic dissection measurement remains at an earlier stage: four studies addressed zonal or compartmental volume and the landmarks that define it. Relative volume error reached 136% in small or thrombosed false lumina [61], and zonal volumes were produced without a reference [60].

Conversely, automated measurement is feasible at population scale. More than 19,000 examinations were measured in a single study [92], and zone-specific reference values were

derived from healthy aortas [90]. This opens the way to normative data and to opportunistic detection, discussed below.

4.4. Detection: High Performance in Selected Samples

Detection of aortic dissection on non-contrast CT was the most clinically relevant application in this group. It addresses patients in whom the diagnosis is not suspected and contrast-enhanced CT has not yet been performed. Among the 11 studies of dissection detection on non-contrast CT, four compared the model with radiologists reading the same images and reported comparable or better performance [98,99,102,110]. The only prospective evaluations measured clinical process rather than accuracy: model-positive studies were opened sooner in a teleradiology deployment [101]; the time to the correct diagnostic pathway fell from 219.7 to 61.6 min in patients with an initially incorrect suspicion [103]. These accuracy estimates, however, come from samples that do not represent intended use. Overall risk of bias was high in 33 of 38 studies, most often because of case–control, matched, or enriched sampling (22 studies). This design inflates prevalence and spectrum. It also explains an apparent paradox: median external AUC (0.97) was no lower than internal AUC (0.96). This more likely reflects enriched external samples than true robustness. Where the external setting differed in acquisition or case mix, performance fell: specificity dropped to 0.55 on thick-slice CT [116]; AUC dropped from 0.94 to 0.85 on external data [117].

Only three studies were at low risk of bias [103,104,112], and all three raised applicability concerns.

Errors were concentrated in the presentations where detection is most clinically relevant: sensitivity was lower for type-B than type-A dissection [104], and all false negatives of a commercial device were type-B dissections [106]; sensitivity fell to 0.83 for penetrating ulcer [97] and to 0.81 in atypical presentations [108]. These are also the cases most easily missed by readers. Image-level partitioning inflated estimates further. The highest accuracy for AAA was reported with image-level splits, and sensitivity fell on reanalysis of a selected set [119,126].

Opportunistic detection of AAA and thoracic dilatation raises a different problem: the balance between yield and downstream workload. Two examples illustrate it: on CT colonography, a screening pipeline reached high sensitivity and specificity but a positive predictive value of 13% [120]; in CT reported as normal, a commercial tool identified 452 unreported dilatations with a positive predictive value of 45% [132]. In both designs, and in a study verifying only discordant cases [121], negative results were not systematically verified, so sensitivity could not be estimated. The value of opportunistic detection depends on the prevalence of the finding in the population and on the cost of confirming false positives. These quantities should be reported alongside accuracy.

Endoleak detection after EVAR was supported by four small case–control studies. One study suggested added value: automated analysis detected low-flow type-II endoleaks missed in the original reports [127].

Across detection, standalone accuracy was the usual endpoint. Studies of AI assistance to readers, which reflect how such tools would be used, were rare.

4.5. Validation, Reporting, and Clinical Translation

Across tasks, several methodological choices favoured optimistic estimates. In at least seven studies, data were partitioned at the level of images, slices, patches, or scans, so that images from the same patient could contribute to both training and testing. In others, operating thresholds or models were selected on data later used for testing [95,101,102,110].

External validation was not a uniform category, and the three forms encountered do not provide equivalent evidence of transportability (Table 4). Testing a model developed

in the study on data from other institutions or on public datasets is the most informative, because it exposes the model to different scanners, protocols and case mix; it was reported in 20 studies, 14 of them in detection and only three in segmentation. A temporally separated or otherwise independent cohort from the same centre (eight studies) guards against overfitting to individual patients but not against site-specific acquisition and population characteristics. Independent evaluation of a pre-trained commercial or open-source tool (12 studies, 10 of them in measurement) does test transportability to a new site, but the independence of the test data rests on the vendor's statement about the training data, and the result describes a fixed product rather than a reproducible method. In 13 further studies, 11 of them evaluating commercial or manufacturer-linked software, independence of the test data from training could not be established, and their results should be read as internal performance. The high headline rate of external validation in detection and the low rate in segmentation therefore reflect, in part, different research cultures: detection studies are mostly clinical diagnostic-accuracy studies with multicentre designs, whereas segmentation studies are mostly methodological. Conclusions on readiness for deployment across heterogeneous clinical environments rest, for each task, on the small number of studies in the first category.

The reference standard was a second recurrent weakness: segmentation and measurement references were often drawn by a single reader; in several studies, the reference was generated or corrected with the authors' own or proprietary software [34,41,47,48,65,90], so that agreement partly measures self-consistency; in opportunistic and triage studies, the reference was often the original radiology report, with verification limited to model-positive or discordant cases [121,132] or with the report incorporated into the reference itself [134].

The evidence reviewed here should be read at three levels, which the included studies address very unequally. The first is technical accuracy: overlap with a reference segmentation, agreement with a reference measurement, or diagnostic accuracy against a reference diagnosis. Almost all studies report at this level, and the values are high; but a DSC of 0.95 or an AUC of 0.97 does not by itself establish that diagnoses, decisions or outcomes improve. The second is clinical effectiveness, which requires endpoints that involve the reader and the patient. Direct comparison with radiologists was reported in 13 detection studies and, as agreement of the AI error with inter-reader variability, in a minority of measurement studies [68,69,83,87,93]; a change in reader performance or in inter-reader variability with AI assistance was measured in three studies [80,82,134]; time to diagnosis or to the correct diagnostic pathway was measured in the two prospective evaluations [101,103]; and the downstream consequences of opportunistic detection were quantified in two studies as positive predictive values of 13% and 45% [120,132]. Safety-relevant error patterns—lower sensitivity for type-B dissection, penetrating ulcer and atypical presentations [97,104,106,108]—belong to this level and are rarely reported as such. The third is implementation readiness: processing failures and exclusions, reported inconsistently but reaching 7–23% where stated [66,72,73,81,86,132]; workflow time against a comparator timed within the study, available in only 13 of the 27 studies reporting time; integration into reading and reporting; and prospective deployment, limited to two systems [101,103]. On this reading, the evidence for aortic CT AI is strong at the first level, sparse at the second, and almost absent at the third. Prospective studies with readers in the loop and clinical endpoints—time to diagnosis, agreement on management decisions, inter-reader agreement with and without AI, downstream workload and the handling of failures—are the evidence still required before integration into routine practice.

4.6. Strengths and Limitations

A key strength of this review is that it provides a comprehensive overview of a broad, fragmented, and rapidly evolving field, which may be particularly useful for readers seeking to navigate the current evidence. It is, to our knowledge, the first to synthesise AI applications across segmentation, measurement, and detection for the whole thoracoabdominal aorta, spanning aneurysmal disease and dissection, contrast-enhanced and non-contrast CT, and pre- and postoperative settings. Two databases were searched and supplemented by backward and forward citation searching. Extraction followed predefined conventions and closed term lists, which prevented aggregation of non-comparable quantities: the compartment segmented, the name of each metric, and the unit of analysis were recorded explicitly. Performance was summarised within comparable compartments and units rather than across all studies.

On the other hand, several limitations should be considered. Embase and Scopus were not included because of accessibility constraints, and IEEE Xplore was not searched; we deliberately focused on databases that are broadly accessible, so that the search can be reproduced by any reader. Studies published mainly in engineering venues may therefore be under-represented. Three observations suggest that the effect on the findings is limited. Citation searching of the included studies and of previous reviews [2–9], seven of which had searched Embase, Scopus or Web of Science [3–9], added only one eligible study; 14 of the 118 included studies were published in engineering or medical physics venues indexed in MEDLINE or Europe PMC; and the studies most likely to have been missed are technical reports on public datasets without a clinical reference standard, which would have fallen almost entirely into the segmentation task, where methodological studies are already concentrated. The most plausible consequence is an under-representation of segmentation methods from engineering venues rather than a distortion of the measurement and detection findings, which originate predominantly from clinical journals. Screening and full-text assessment were performed by two reviewers jointly rather than independently in duplicate. This does not generate two independent judgements and may reduce the detection of individual errors; the risk of omitting eligible records was mitigated by the structured re-examination of screened records and by citation searching, which identified one additional study. Data were extracted by one reviewer with AI assistance and verified by a second reviewer rather than extracted independently in duplicate, which may leave some extraction errors undetected. The included studies were heterogeneous in sample size, population, acquisition protocol, reference standard and validation design, and the summary medians do not adjust for these differences; performance values reported by different studies are therefore not directly comparable, and the subgroup summaries in Supplementary Table S9 are descriptive and based on small numbers of studies. Assigning multitask studies to a single primary task required judgement and may have introduced some degree of subjectivity.

5. Conclusions

AI for aortic CT and CTA has reached high technical accuracy across segmentation, measurement, and detection; evidence of clinical effectiveness is sparse, and evidence of implementation readiness is almost absent.

- Segmentation: Performance depends on the compartment. Luminal and outer wall targets are well addressed, whereas thrombus remains the principal difficulty, and overlap metrics alone do not establish the boundary accuracy on which measurement depends.
- Measurement: Automated diameter is unbiased on average and comparable to inter-reader variability, but its error for the individual patient is large enough to matter near

decision thresholds. Its most promising role is reproducible surveillance, provided that failure rates and agreement on change are established.

- Detection: Accuracy for aortic dissection, including on non-contrast CT, is high but was estimated mostly in enriched samples, and performance is weakest in the presentations that are hardest for readers.

The next step is not further evidence of technical accuracy but evidence at the two levels that remain thin: independent multicentre validation in consecutive cohorts with transparent reporting of failures, and prospective studies with readers in the loop measuring time to diagnosis, agreement on management decisions, inter-reader agreement with and without AI, and downstream workload. Until such evidence is available, these tools are best regarded as aids to measurement and triage under radiologist supervision rather than as autonomous systems.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/tomography12110166/s1. Table S1: Search strategies; Table S2: Reports excluded after full-text assessment, with reasons (n = 68); Table S3: Data extraction: segmentation studies (n = 42); Table S4: Data extraction: measurement studies (n = 38); Table S5: Data extraction: detection and classification studies (n = 38); Table S6: CLAIM-informed appraisal of segmentation studies (n = 42); Table S7: CLAIM-informed appraisal of measurement studies (n = 38); Table S8: QUADAS-2/QUADAS-AI appraisal of detection studies (n = 38); Table S9: Subgroup summaries of the headline performance statistics by task; Table S10: Decision rules for the CLAIM-informed limitation tiers (segmentation and measurement studies); PRISMA 2020 checklist [10].

Author Contributions: Conceptualisation, F.P., M.M., G.S., G.P. and L.M.; methodology, F.P., M.M., N.P., E.F.M. and G.C.; writing—original draft preparation, F.P. and G.S.; writing—review and editing, F.P., G.S., N.P., E.F.M., M.M., G.C., G.P. and L.M.; supervision, L.M., G.P. and G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created in this study. All data extracted from the included studies, the search strategies, the list of excluded reports and the quality appraisal tables are available in the Supplementary Materials (Tables S1–S8).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AAA	abdominal aortic aneurysm
AI	artificial intelligence
AUC	area under the receiver operating characteristic curve
CLAIM	Checklist for Artificial Intelligence in Medical Imaging
CNN	convolutional neural network
CT	computed tomography
CTA	computed tomography angiography
DL	deep learning
DSC	Dice similarity coefficient
ECG	electrocardiogram
EVAR	endovascular aneurysm repair
FEVAR	fenestrated endovascular aneurysm repair
GAN	generative adversarial network
HD95	95th-percentile Hausdorff distance
ICC	intraclass correlation coefficient

ILT	intraluminal thrombus
IoU	intersection over union
LoA	limits of agreement
MAE	mean absolute error
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
SVS/STS	Society for Vascular Surgery/Society of Thoracic Surgeons
SWiM	Synthesis Without Meta-analysis
TAA	thoracic aortic aneurysm
TEVAR	thoracic endovascular aortic repair

References

1. Li, C.; Liu, C. Multimodal imaging of acquired aortic diseases: Clinical efficacy, comparative analysis, and future perspectives. *Int. J. Cardiovasc. Imaging* **2026**, *42*, 561–577. [CrossRef] [PubMed]
2. Xu, J.; Liu, J.; Cao, G. Artificial intelligence in aortic CT angiography: Current applications and future perspectives. *Front. Cardiovasc. Med.* **2026**, *12*, 1674486. [CrossRef] [PubMed]
3. Guo, J.; Lareyre, F.; Goffart, S.; Chierici, A.; Delingette, H.; Raffort, J. Automatic segmentation of intraluminal thrombus in abdominal aortic aneurysms based on CT images: A comprehensive review of deep learning-based methods. *J. Clin. Med.* **2025**, *14*, 8497. [CrossRef] [PubMed]
4. Lo, E.M.K.; Chen, S.; Ng, K.H.L.; Wong, R.H.L. Artificial intelligence-powered solutions for automated aortic diameter measurement in computed tomography: A narrative review. *Ann. Transl. Med.* **2024**, *12*, 116. [CrossRef] [PubMed]
5. Antoun, I.; Layton, G.R.; Somani, R.; Ibrahim, M.; Ng, G.A.; Mariscalco, G.; Zakkar, M. Time is aorta: Can artificial intelligence improve surgical timelines in acute type A aortic dissection? A comprehensive review. *Health Sci. Rep.* **2026**, *9*, e72906. [CrossRef] [PubMed]
6. Spinella, E.; Magliocco, M.; Basso, C.; Fantazzini, A.; Pane, B.; Conti, M.; Pratesi, G.; Spinella, G. Impact of CT acquisition parameters on deep learning of aortic segmentation performance: Systematic review. *J. Imaging Inform. Med.* 2026, *in press*. [CrossRef]
7. Zhao, Y.; Zhang, Y.; Zhang, J.; Chen, C.; Li, P.; Li, J.; Fu, Q. Performance of image-based deep learning models for aortic dissection segmentation and diagnosis: A systematic review and meta-analysis. *Front. Cardiovasc. Med.* **2026**, *13*, 1734208. [CrossRef] [PubMed]
8. Wang, T.W.; Hong, J.S.; Liu, H.R.; Wu, K.T.; Fu, H.N.; Lee, Y.T.; Liao, P.-H.; Huang, G.-H.; Chen, I.-M.; Yin, W.-H.; et al. Deep learning versus radiologists for acute aortic dissection on CT: A systematic review and network meta-analysis. *Eur. J. Radiol.* **2026**, *205*, 113189. [CrossRef] [PubMed]
9. Asif, A.; Alsayyari, M.; Monekosso, D.; Remagnino, P.; Lakshminarayan, R. Role of artificial intelligence in detecting and classifying aortic dissection: Where are we? A systematic review and meta-analysis. *Radiol. Cardiothorac. Imaging* **2025**, *7*, e240353. [CrossRef] [PubMed]
10. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
11. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
12. Whiting, P.F.; Rutjes, A.W.S.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.G.; Sterne, J.A.C.; Bossuyt, P.M.M.; QUADAS-2 Group. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **2011**, *155*, 529–536. [CrossRef] [PubMed]
13. Sounderajah, V.; Ashrafian, H.; Rose, S.; Shah, N.H.; Ghassemi, M.; Golub, R.; Kahn, C.E.; Esteva, A.; Karthikesalingam, A.; Mateen, B.; et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nat. Med.* **2021**, *27*, 1663–1665. [CrossRef] [PubMed]
14. Guni, A.; Sounderajah, V.; Whiting, P.; Bossuyt, P.; Darzi, A.; Ashrafian, H. Revised tool for the quality assessment of diagnostic accuracy studies using AI (QUADAS-AI): Protocol for a qualitative study. *JMIR Res. Protoc.* **2024**, *13*, e58202. [CrossRef] [PubMed]
15. Tejani, A.S.; Klontzas, M.E.; Gatti, A.A.; Mongan, J.T.; Moy, L.; Park, S.H.; Kahn, C.E.; Panel, F.T.C.2.U.; Abbara, S.; Afat, S.; et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 update. *Radiol. Artif. Intell.* **2024**, *6*, e240300. [CrossRef] [PubMed]

16. Campbell, M.; McKenzie, J.E.; Sowden, A.; Katikireddi, S.V.; Brennan, S.E.; Ellis, S.; Hartmann-Boyce, J.; Ryan, R.; Shepperd, S.; Thomas, J.; et al. Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ* **2020**, *368*, l6890. [CrossRef] [PubMed]
17. Abaid, A.; Ilancheran, S.; Iqbal, T.; Hynes, N.; Ullah, I. Exploratory analysis of Type B Aortic Dissection (TBAD) segmentation in 2D CTA images using various kernels. *Comput. Med. Imaging Graph.* **2024**, *118*, 102460. [CrossRef] [PubMed]
18. Cao, L.; Shi, R.; Ge, Y.; Xing, L.; Zuo, P.; Jia, Y.; Liu, J.; He, Y.; Wang, X.; Luan, S.; et al. Fully automatic segmentation of type B aortic dissection from CTA images enabled by deep learning. *Eur. J. Radiol.* **2019**, *121*, 108713. [CrossRef] [PubMed]
19. Chen, D.; Zhang, X.; Mei, Y.; Liao, F.; Xu, H.; Li, Z.; Xiao, Q.; Guo, W.; Zhang, H.; Yan, T.; et al. Multi-stage learning for segmentation of aortic dissections using a prior aortic anatomy simplification. *Med. Image Anal.* **2021**, *69*, 101931. [CrossRef] [PubMed]
20. Feng, H.; Fu, Z.; Wang, Y.; Zhang, P.; Lai, H.; Zhao, J. Automatic segmentation of thrombosed aortic dissection in post-operative CT-angiography images. *Med. Phys.* **2023**, *50*, 3538–3548. [CrossRef] [PubMed]
21. Guo, X.; Liu, T.; Yang, Y.; Dai, J.; Wang, L.; Tang, D.; Sun, H. Automatic Segmentation of Type A Aortic Dissection on Computed Tomography Images Using Deep Learning Approach. *Diagnostics* **2024**, *14*, 1332. [CrossRef] [PubMed]
22. Hahn, L.D.; Mistelbauer, G.; Higashigaito, K.; Koci, M.; Willemink, M.J.; Sailer, A.M.; Fischbein, M.; Fleischmann, D. CT-based True- and False-Lumen Segmentation in Type B Aortic Dissection Using Machine Learning. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e190179. [CrossRef] [PubMed]
23. Imran, M.; Krebs, J.R.; Gopu, V.R.R.; Fazzone, B.; Sivaraman, V.B.; Kumar, A.; Viscardi, C.; Heithaus, R.E.; Shickel, B.; Zhou, Y.; et al. CIS-UNet: Multi-class segmentation of the aorta in computed tomography angiography via context-aware shifted window self-attention. *Comput. Med. Imaging Graph.* **2024**, *118*, 102470. [CrossRef] [PubMed]
24. Imran, M.; Krebs, J.R.; Sivaraman, V.B.; Zhang, T.; Kumar, A.; Ueland, W.R.; Fassler, M.J.; Wang, C.; Huang, J.; Sun, X.; et al. Multi-class segmentation of aortic branches and zones in computed tomography angiography: The AortaSeg24 challenge. *Med. Image Anal.* **2026**, *113*, 104188. [CrossRef] [PubMed]
25. Jung, J.H.; Oh, H.M.; Jeong, G.J.; Kim, T.W.; Koo, H.J.; Lee, J.G.; Yang, D.H. ZOZI-Seg: A transformer and UNet cascade network with Zoom-Out and Zoom-In scheme for aortic dissection segmentation in enhanced CT images. *Comput. Biol. Med.* **2024**, *175*, 108494. [CrossRef] [PubMed]
26. Kesavuori, 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. *Eur. Radiol. Exp.* **2023**, *7*, 35. [CrossRef] [PubMed]
27. Lyu, T.; Yang, G.; Zhao, X.; Shu, H.; Luo, L.; Chen, D.; Xiong, J.; Yang, J.; Li, S.; Coatrieux, J.L.; et al. Dissected aorta segmentation using convolutional neural networks. *Comput. Methods Programs Biomed.* **2021**, *211*, 106417. [CrossRef] [PubMed]
28. Wobben, L.D.; Codari, M.; Mistelbauer, G.; Pepe, A.; Higashigaito, K.; Hahn, L.D.; Mastrodicasa, D.; Turner, V.L.; Hinostroza, V.; Bäumler, K.; et al. Deep Learning-Based 3D Segmentation of True Lumen, False Lumen, and False Lumen Thrombosis in Type-B Aortic Dissection. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2021**, *2021*, 3912–3915. [CrossRef] [PubMed]
29. Yao, Z.; Xie, W.; Zhang, J.; Dong, Y.; Qiu, H.; Yuan, H.; Jia, Q.; Wang, T.; Shi, Y.; Zhuang, J.; et al. ImageTBAD: A 3D Computed Tomography Angiography Image Dataset for Automatic Segmentation of Type-B Aortic Dissection. *Front. Physiol.* **2021**, *12*, 732711. [CrossRef] [PubMed]
30. Yu, Y.; Gao, Y.; Wei, J.; Liao, F.; Xiao, Q.; Zhang, J.; Yin, W.; Lu, B. A Three-Dimensional Deep Convolutional Neural Network for Automatic Segmentation and Diameter Measurement of Type B Aortic Dissection. *Korean J. Radiol.* **2021**, *22*, 168–178. [CrossRef] [PubMed]
31. Abdolmanafi, A.; Forneris, A.; Moore, R.D.; Di Martino, E.S. Deep-learning method for fully automatic segmentation of the abdominal aortic aneurysm from computed tomography imaging. *Front. Cardiovasc. Med.* **2022**, *9*, 1040053. [CrossRef] [PubMed]
32. Brutti, F.; Fantazzini, A.; Finotello, A.; Muller, L.O.; Auricchio, F.; Pane, B.; Spinella, G.; Conti, M. Deep Learning to Automatically Segment and Analyze Abdominal Aortic Aneurysm from Computed Tomography Angiography. *Cardiovasc. Eng. Technol.* **2022**, *13*, 535–547. [CrossRef] [PubMed]
33. Caradu, C.; Spampinato, B.; Vrancianu, A.M.; Berard, X.; Ducasse, E. Fully automatic volume segmentation of infrarenal abdominal aortic aneurysm computed tomography images with deep learning approaches versus physician controlled manual segmentation. *J. Vasc. Surg.* **2021**, *74*, 246–256.e6. [CrossRef] [PubMed]
34. Chandrashekar, A.; Handa, A.; Lapolla, P.; Shivakumar, N.; Uberoi, R.; Grau, V.; Lee, R. A Deep Learning Approach to Visualize Aortic Aneurysm Morphology Without the Use of Intravenous Contrast Agents. *Ann. Surg.* **2023**, *277*, e449–e459. [CrossRef] [PubMed]
35. Derycke, L.; Doudouh, A.; Cotte, F.; Habi, A.; Albertini, J.N.; Millon, A. Validation of a Deep Learning U-Net Algorithm for Multistructure Segmentation of Infrarenal Abdominal Aortic Aneurysms including Lumen, Thrombus, and Calcifications. *EJVES Vasc. Forum* **2026**, *65*, 1–10. [CrossRef] [PubMed]

36. Fantazzini, A.; Esposito, M.; Finotello, A.; Auricchio, F.; Pane, B.; Basso, C.; Spinella, G.; Conti, M. 3D Automatic Segmentation of Aortic Computed Tomography Angiography Combining Multi-View 2D Convolutional Neural Networks. *Cardiovasc. Eng. Technol.* **2020**, *11*, 576–586. [CrossRef] [PubMed]
37. Hendricks, C.H.F.; van Veldhuizen, W.A.; Ranem, A.; Fioole, B.; Kropman, R.H.J.; Wolterink, J.M.; de Vries, J.P.P.M. Multi-centre Validation of Automated Outer-to-Outer Wall Segmentation of Infrarenal Abdominal Aortic Aneurysms Using a nnU-Net Based Model. *J. Imaging Inform. Med.* **2026**, *online ahead of print*. [CrossRef] [PubMed]
38. Hwang, B.; Kim, J.; Lee, S.; Kim, E.; Kim, J.; Jung, Y.; Hwang, H. Automatic Detection and Segmentation of Thrombi in Abdominal Aortic Aneurysms Using a Mask Region-Based Convolutional Neural Network with Optimized Loss Functions. *Sensors* **2022**, *22*, 3643. [CrossRef] [PubMed]
39. Jung, Y.; Kim, S.; Kim, J.; Hwang, B.; Lee, S.; Kim, E.Y.; Kim, J.H.; Hwang, H. Abdominal Aortic Thrombus Segmentation in Postoperative Computed Tomography Angiography Images Using Bi-Directional Convolutional Long Short-Term Memory Architecture. *Sensors* **2022**, *23*, 175. [CrossRef] [PubMed]
40. Lopez-Linares, K.; Aranjuelo, N.; Kabongo, L.; Maclair, G.; Lete, N.; Ceresa, M.; Garcia-Familiar, A.; Macia, I.; Gonzalez Ballester, M.A. Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using Deep Convolutional Neural Networks. *Med. Image Anal.* **2018**, *46*, 202–214. [CrossRef] [PubMed]
41. Lopez-Linares, K.; Stephens, M.; Garcia, I.; Macia, I.; Ballester, M.A.G.; Estepar, R.S.J. Abdominal Aortic Aneurysm Segmentation Using Convolutional Neural Networks Trained with Images Generated with a Synthetic Shape Model. In *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting (MLMECH/CVII-STENT 2019)*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2019; Volume 11794, pp. 167–174. [CrossRef] [PubMed]
42. Lyu, Z.; King, K.; Rezaeitaleshmahalleh, M.; Pienta, D.; Mu, N.; Zhao, C.; Zhou, W.; Jiang, J. Deep-learning-based image segmentation for image-based computational hemodynamic analysis of abdominal aortic aneurysms: A comparison study. *Biomed. Phys. Eng. Express* **2023**, *9*, 067001. [CrossRef] [PubMed]
43. Lyu, Z.; Mu, N.; Rezaeitaleshmahalleh, M.; Zhang, X.; McBane, R.; Jiang, J. Automatic segmentation of intraluminal thrombosis of abdominal aortic aneurysms from CT angiography using a mixed-scale-driven multiview perception network (M(2)Net) model. *Comput. Biol. Med.* **2024**, *179*, 108838. [CrossRef] [PubMed]
44. Ma, Q.; Lucas, A.; Hammami, H.; Shu, H.; Kaladji, A.; Haigron, P. Deep-learning approach to automate the segmentation of aorta in non-contrast CTs. *J. Med. Imaging* **2023**, *10*, 024001. [CrossRef] [PubMed]
45. Ma, Q.; Kaladji, A.; Shu, H.; Yang, G.; Lucas, A.; Haigron, P. Beyond strong labels: Weakly-supervised learning based on Gaussian pseudo labels for the segmentation of ellipse-like vascular structures in non-contrast CTs. *Med. Image Anal.* **2025**, *99*, 103378. [CrossRef] [PubMed]
46. Mu, N.; Lyu, Z.; Rezaeitaleshmahalleh, M.; Zhang, X.; Rasmussen, T.; McBane, R.; Jiang, J. Automatic segmentation of abdominal aortic aneurysms from CT angiography using a context-aware cascaded U-Net. *Comput. Biol. Med.* **2023**, *158*, 106569. [CrossRef] [PubMed]
47. Roby, M.; Restrepo, J.C.; Park, H.; Muluk, S.C.; Eskandari, M.K.; Baek, S.; Finol, E.A. Automatic Segmentation of Abdominal Aortic Aneurysm From Computed Tomography Angiography Using a Patch-Based Dilated UNet Model. *IEEE Access* **2025**, *13*, 24544–24554. [CrossRef] [PubMed]
48. Roby, M.; Sakib, A.N.M.; Zhang, Z.; Muluk, S.C.; Eskandari, M.K.; Finol, E.A. Automatic Explainable Segmentation of Abdominal Aortic Aneurysm From Computed Tomography Angiography. *IEEE Access* **2025**, *13*, 178021–178033. [CrossRef] [PubMed]
49. Salvi, A.; Finol, E.; Menon, P.G. Convolutional Neural Network based Segmentation of Abdominal Aortic Aneurysms. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2021**, *2021*, 2629–2632. [CrossRef] [PubMed]
50. Siriapisith, T.; Kusakunniran, W.; Haddawy, P. A retrospective study of 3D deep learning approach incorporating coordinate information to improve the segmentation of pre- and post-operative abdominal aortic aneurysm. *PeerJ Comput. Sci.* **2022**, *8*, e1033. [CrossRef] [PubMed]
51. Wang, Y.; Zhou, M.; Ding, Y.; Li, X.; Zhou, Z.; Xie, T.; Shi, Z.; Fu, W. Fully automatic segmentation of abdominal aortic thrombus in pre-operative CTA images using deep convolutional neural networks. *Technol. Health Care* **2022**, *30*, 1257–1266. [CrossRef] [PubMed]
52. Aghayeva, F.; Abdi, Y.; Uzair, A.; Aghayev, A. Automated Detection and Segmentation of Ascending Aorta Dilation on a Non-ECG-Gated Chest CT Using Deep Learning. *Diagnostics* **2025**, *15*, 2336. [CrossRef] [PubMed]
53. Comelli, A.; Dahiya, N.; Stefano, A.; Benfante, V.; Gentile, G.; Agnese, V.; Raffa, G.M.; Pilato, M.; Yezzi, A.; Petrucci, G.; et al. Deep learning approach for the segmentation of aneurysmal ascending aorta. *Biomed. Eng. Lett.* **2021**, *11*, 15–24. [CrossRef] [PubMed]
54. Chen, H.; Yan, S.; Xie, M.; Huang, J. Application of cascaded GAN based on CT scan in the diagnosis of aortic dissection. *Comput. Methods Programs Biomed.* **2022**, *226*, 107130. [CrossRef] [PubMed]
55. El Rahal, D.; Pozzessere, C.; Gulizia, M.; Rotzinger, D.C.; Fahrni, G. TotalSegmentator-AI for aortic segmentation in CT: Reliable performance in normal anatomy but limited utility in pathological aortic disease. *Eur. J. Radiol. Open* **2026**, *16*, 100772. [CrossRef] [PubMed]

56. Liu, J.; Huang, H.; Zhang, Y.; Wu, B. MAFR-UNet: Multi-scale adaptive feature reassembly network for aortic CTA segmentation. *Sci. Rep.* **2026**, *online ahead of print*. [CrossRef] [PubMed]
57. Sieren, M.M.; Widmann, C.; Weiss, N.; Moltz, J.H.; Link, F.; Wegner, F.; Stahlberg, E.; Horn, M.; Oecherting, T.H.; Goltz, J.P.; et al. Automated segmentation and quantification of the healthy and diseased aorta in CT angiographies using a dedicated deep learning approach. *Eur. Radiol.* **2022**, *32*, 690–701. [CrossRef] [PubMed]
58. Vagenas, T.P.; Vezakis, I.; Kakkos, I.; Mavridis, C.; Economopoulos, T.; Anastasiou, A.; Raptis, A.; Manopoulos, C.; Matsopoulos, G.K. Self-Supervised Pre-Training with Intensity Guided Masking for Enhanced Aorta Segmentation in CT. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2025**, *2025*, 1–7. [CrossRef] [PubMed]
59. Bäumler, K.; Codari, M.; Mastrodicasa, D.; Mistelbauer, G.; Willemink, M.J.; Walters, S.; Hinostroza, V.; Turner, V.; Chepelev, L.; Sriprachyakul, A.; et al. Deep reinforcement learning for automatic anatomic CT landmark localization in Stanford Type B aortic dissection. *Radiol. Adv.* **2026**, *3*, umag006. Correction in *Radiol. Adv.* **2026**, *3*, umag024. https://doi.org/10.1093/radadv/umag024. [CrossRef] [PubMed]
60. Krebs, J.R.; Imran, M.; Fazzone, B.; Viscardi, C.; Berwick, B.; Stinson, G.; Heithaus, E.; Upchurch, G.R.J.; Shao, W.; Cooper, M.A. Volumetric analysis of acute uncomplicated type B aortic dissection using an automated deep learning aortic zone segmentation model. *J. Vasc. Surg.* **2024**, *80*, 1025–1034.e4. [CrossRef] [PubMed]
61. Nakano, Y.; Nishi, S.; Kojima, I.; Shimahara, Y. AI-Based Segmental Volumetry of the Downstream Aorta in Aortic Dissection: End-to-End Versus Hybrid Strategies. *J. Imaging Inform. Med.* **2026**, *online ahead of print*. [CrossRef] [PubMed]
62. Pepe, A.; Egger, J.; Codari, M.; Willemink, M.J.; Gsaxner, C.; Li, J.; Roth, P.M.; Schmalstieg, D.; Mistelbauer, G.; Fleischmann, D. Automated cross-sectional view selection in CT angiography of aortic dissections with uncertainty awareness and retrospective clinical annotations. *Comput. Biol. Med.* **2023**, *165*, 107365. [CrossRef] [PubMed]
63. Broda, M.; Rossillon, A.; Le Houerou, T.; Ruppli, C.; Lorisson, E.; Al Osail, E.; Fabre, D.; Bartoli, M.; Haulon, S. The benefit of automated sac volume measurements in postoperative endovascular aortic repair surveillance. *J. Vasc. Surg.* **2025**, *81*, 623–629. [CrossRef] [PubMed]
64. Caradu, C.; Pouncey, A.L.; Lakhlifi, E.; Brunet, C.; Berard, X.; Ducasse, E. Fully automatic volume segmentation using deep learning approaches to assess aneurysmal sac evolution after infrarenal endovascular aortic repair. *J. Vasc. Surg.* **2022**, *76*, 620–630.e3. [CrossRef] [PubMed]
65. Coastaliou, Q.; Webster, C.; Bicknell, C.; Pouncey, A.; Ducasse, E.; Caradu, C. Artificial Intelligence With Deep Learning Enables Assessment of Aortic Aneurysm Diameter and Volume Through Different Computed Tomography Phases. *Eur. J. Vasc. Endovasc. Surg.* **2024**, *68*, 408–409. [CrossRef] [PubMed]
66. Hatzl, J.; Uhl, C.; Barb, A.; Henning, D.; Fiering, J.; El-Sanosy, E.; Cuypers, P.W.M.; Bockler, D.; Zephyr Study Group Collaborators. External Validation of Fully-Automated Infrarenal Maximum Aortic Aneurysm Diameter Measurements in Computed Tomography Angiography Scans Using Artificial Intelligence (PRAEVAorta 2). *J. Endovasc. Ther.* **2024**, *online ahead of print*. [CrossRef] [PubMed]
67. Kim, T.; On, S.; Gwon, J.G.; Kim, N. Computed tomography-based automated measurement of abdominal aortic aneurysm using semantic segmentation with active learning. *Sci. Rep.* **2024**, *14*, 8924. [CrossRef] [PubMed]
68. Lareyre, F.; Chaudhuri, A.; Flory, V.; Augene, E.; Adam, C.; Carrier, M.; Amrani, S.; Chikande, J.; Le, C.D.; Raffort, J. Automatic Measurement of Maximal Diameter of Abdominal Aortic Aneurysm on Computed Tomography Angiography Using Artificial Intelligence. *Ann. Vasc. Surg.* **2022**, *83*, 202–211. [CrossRef] [PubMed]
69. Pouncey, A.L.; Charles, E.; Bicknell, C.D.; Berard, X.; Ducasse, E.; Caradu, C. Editor's Choice—Fully Automatic Volume Segmentation Using Deep Learning Approaches to Assess the Thoracic Aorta, Visceral Abdominal Aorta, and Visceral Vasculature. *Eur. J. Vasc. Endovasc. Surg.* **2026**, *71*, 672–681. [CrossRef] [PubMed]
70. Robbi, E.; Ravanelli, D.; Allievi, S.; Raunig, I.; Bonvini, S.; Passerini, A.; Trianni, A. Automatic CTA analysis for blood vessels and aneurysm features extraction in EVAR planning. *Sci. Rep.* **2025**, *15*, 16431. [CrossRef] [PubMed]
71. Sharma, C.; Wu, M.; Enjeti, A.; Thanigaimani, S.; Jenkins, J.; Quigley, F.; Golledge, J. Use of artificial intelligence-assisted measurement of aortic diameter in clinical decision making about abdominal aortic aneurysm repair. *Abdom. Radiol.* **2026**, *51*, 3643–3652. [CrossRef] [PubMed]
72. van Tongeren, O.L.R.M.; Vanmaele, A.; Rastogi, V.; Hoeks, S.E.; Verhagen, H.J.M.; de Bruin, J.L. Volume Measurements for Surveillance after Endovascular Aneurysm Repair Using Artificial Intelligence. *Eur. J. Vasc. Endovasc. Surg.* **2025**, *69*, 61–70. [CrossRef] [PubMed]
73. Wegner, M.; Fontaine, V.; Nana, P.; Dieffenbach, B.V.; Fabre, D.; Haulon, S. Artificial Intelligence-Assisted Sac Diameter Assessment for Complex Endovascular Aortic Repair. *J. Endovasc. Ther.* **2025**, *32*, 1172–1181. [CrossRef] [PubMed]
74. Weiss, D.; Hager, T.; Lin, M.; Sritharan, D.; Bousabarah, K.; Renninghoff, D.; Holler, W.; Simmons, K.; Haubold, J.; Loh, S.; et al. Deep learning based volumetric analysis of infrarenal abdominal aortic aneurysms characterized on CTA. *npj Digit. Med.* **2026**, *9*, 87. [CrossRef] [PubMed]

75. Fink, N.; Yacoub, B.; Schoepf, U.J.; Zsarnoczay, E.; Pinos, D.; Vecsey-Nagy, M.; Rapaka, S.; Sharma, P.; O'Doherty, J.; Ricke, J.; et al. Artificial Intelligence Provides Accurate Quantification of Thoracic Aortic Enlargement and Dissection in Chest CT. *Diagnostics* **2024**, *14*, 866. [CrossRef] [PubMed]
76. Graby, J.; Harris, M.; Jones, C.; Waring, H.; Lyen, S.; Hudson, B.J.; Rodrigues, J.C.L. Assessing the role of an artificial intelligence assessment tool for thoracic aorta diameter on routine chest CT. *Br. J. Radiol.* **2023**, *96*, 20220853. [CrossRef] [PubMed]
77. Katakol, S.; Baker, T.J.; Bian, Z.; Lu, Y.; Spahlinger, G.; Hatt, C.R.; Burris, N.S. Fully automated pipeline for measurement of the thoracic aorta using joint segmentation and localization neural network. *J. Med. Imaging* **2023**, *10*, 051810. [CrossRef] [PubMed]
78. Livolsi, C.; Nioi, L.; Castelbuono, S.; Gentile, G.; Raffa, G.M.; Pilato, M.; Agnese, V.; Bellavia, D.; Pasta, S.; Cuscino, N. AI-based Framework for Automated Segmentation and Surgical Decision Support in Ascending Thoracic Aortic Aneurysm. *IEEE J. Biomed. Health Inform.* **2026**, *online ahead of print*. [CrossRef] [PubMed]
79. Lo Piccolo, F.; Hinck, D.; Segeroth, M.; Sperl, J.; Cyriac, J.; Yang, S.; Rapaka, S.; Bremerich, J.; Sauter, A.W.; Pradella, M. Impact of retraining a deep learning algorithm for improving guideline-compliant aortic diameter measurements on non-gated chest CT. *Eur. J. Radiol.* **2023**, *168*, 111093. [CrossRef] [PubMed]
80. Marway, P.S.; Campello Jorge, C.A.; Baker, T.; Tjahjadi, N.; Spahlinger, G.; Burris, N.S. Comparison of Human-in-the-Loop Neural Network and Manual Methods for Aortic Diameter Measurement at CT Angiography. *Radiol. Cardiothorac. Imaging* **2025**, *7*, e240490. [CrossRef] [PubMed]
81. Pradella, M.; Weikert, T.; Sperl, J.I.; Kargel, R.; Cyriac, J.; Achermann, R.; Sauter, A.W.; Bremerich, J.; Stieltjes, B.; Brantner, P.; et al. Fully automated guideline-compliant diameter measurements of the thoracic aorta on ECG-gated CT angiography using deep learning. *Quant. Imaging Med. Surg.* **2021**, *11*, 4245–4257. [CrossRef] [PubMed]
82. Rueckel, J.; Reidler, P.; Fink, N.; Sperl, J.; Geyer, T.; Fabritius, M.P.; Ricke, J.; Ingrisch, M.; Sabel, B.O. Artificial intelligence assistance improves reporting efficiency of thoracic aortic aneurysm CT follow-up. *Eur. J. Radiol.* **2021**, *134*, 109424. [CrossRef] [PubMed]
83. Adam, C.; Fabre, D.; Mougin, J.; Zins, M.; Azarine, A.; Ardon, R.; d'Assignies, G.; Haulon, S. Pre-surgical and Post-surgical Aortic Aneurysm Maximum Diameter Measurement: Full Automation by Artificial Intelligence. *Eur. J. Vasc. Endovasc. Surg.* **2021**, *62*, 869–877. [CrossRef] [PubMed]
84. Artzner, C.; Bongers, M.N.; Kargel, R.; Faby, S.; Hefferman, G.; Herrmann, J.; Nopper, S.L.; Perl, R.M.; Walter, S.S. Assessing the Accuracy of an Artificial Intelligence-Based Segmentation Algorithm for the Thoracic Aorta in Computed Tomography Applications. *Diagnostics* **2022**, *12*, 1790. [CrossRef] [PubMed]
85. Bratt, A.; Blezek, D.J.; Ryan, W.J.; Philbrick, K.A.; Rajiah, P.; Tandon, Y.K.; Walkoff, L.A.; Cai, J.C.; Sheedy, E.N.; Korfiatis, P.; et al. Deep Learning Improves the Temporal Reproducibility of Aortic Measurement. *J. Digit. Imaging* **2021**, *34*, 1183–1189. [CrossRef] [PubMed]
86. Gatinot, A.; Caradu, C.; Stephan, L.; Foret, T.; Rinckenbach, S. Validation Study Comparing Artificial Intelligence for Fully Automatic Aortic Aneurysms Segmentation and Diameter Measurements on Contrast and Noncontrast Enhanced Computed Tomography: A la memoire de Lucie Salomon du Mont (In memory of Lucie Salomon du Mont). *Ann. Vasc. Surg.* **2026**, *122*, 53–65. [CrossRef] [PubMed]
87. Hamelink, I.I.; de Heide, E.E.J.; Pelgrim, G.J.G.J.; Kwee, T.C.T.; van Ooijen, P.M.A.P.; de Bock, G.H.T.; Vliegenthart, R.R. Validation of an AI-based algorithm for measurement of the thoracic aortic diameter in low-dose chest CT. *Eur. J. Radiol.* **2023**, *167*, 111067. [CrossRef] [PubMed]
88. Heo, Y.; Lee, G.E.; Cho, J.; Choi, S.I. Deep Learning-Based Aortic Diameter Measurement in Traumatic Hemorrhage Using Shallow Attention Network: A Path Forward. *Diagnostics* **2025**, *15*, 1312. [CrossRef] [PubMed]
89. Hong, J.S.; Tzeng, Y.H.; Wu, K.T.; Huang, S.Y.; Wang, T.W.; Li, G.Y.; Lin, C.Y.; Liu, H.R.; Fu, H.N.; Lee, Y.T.; et al. Automated Aortic Quantification Based on Artificial Intelligence: Validation Using Contrast-Enhanced and Non-Contrast CT Scans from the Same Session. *Bioengineering* **2026**, *13*, 471. [CrossRef] [PubMed]
90. Koo, H.J.; Lee, J.G.; Lee, J.B.; Kang, J.W.; Yang, D.H. Deep Learning Based Automatic Segmentation of the Thoracic Aorta from Chest Computed Tomography in Healthy Korean Adults. *Eur. J. Vasc. Endovasc. Surg.* **2025**, *69*, 48–58. [CrossRef] [PubMed]
91. Park, C.; Lee, H.J.; Jeong, W.G. Artificial intelligence-assisted automated measurement of maximum aortic diameter in the Korean national lung cancer screening CT program. *Br. J. Radiol.* **2026**, *99*, 550–560. [CrossRef] [PubMed]
92. Paulsen, R.R.; Juul, L.H.; Pham, M.H.C.; Kuhl, J.T.; Kofoed, K.F.; Sorensen, K.A.; Sundgaard, J.V. AortaExplorer: AI-driven analysis of the aorta in CT images. *Med. Biol. Eng. Comput.* **2026**, *64*, 1745–1760. [CrossRef] [PubMed]
93. Postiglione, T.J.; Guillo, E.; Heraud, A.; Rossillon, A.; Bartoli, M.; Herpe, G.; Adam, C.; Fabre, D.; Ardon, R.; Azarine, A.; et al. Multicentric clinical evaluation of a computed tomography-based fully automated deep neural network for aortic maximum diameter and volumetric measurements. *J. Vasc. Surg.* **2024**, *79*, 1390–1400.e8. [CrossRef] [PubMed]
94. Saitta, S.; Sturla, F.; Caimi, A.; Riva, A.; Palumbo, M.C.; Nano, G.; Votta, E.; Corte, A.D.; Glauber, M.; Chiappino, D.; et al. A Deep Learning-Based and Fully Automated Pipeline for Thoracic Aorta Geometric Analysis and Planning for Endovascular Repair from Computed Tomography. *J. Digit. Imaging* **2022**, *35*, 226–239. [CrossRef] [PubMed]
95. Tahoces, P.G.; Varela, R.; Carreira, J.M. Deep learning method for aortic root detection. *Comput. Biol. Med.* **2021**, *135*, 104533. [CrossRef] [PubMed]

96. Yang, J.; Li, X.; Cheng, J.Z.; Xue, Z.; Shi, F.; Ji, Y.; Wang, X.; Yang, F. Segment aorta and localize landmarks simultaneously on noncontrast CT using a multitask learning framework for patients without severe vascular disease. *Comput. Biol. Med.* **2023**, *160*, 107002. [CrossRef] [PubMed]
97. Chang, J.; Lee, K.J.; Wang, T.H.; Chen, C.M. Multi-Stage Cascaded Deep Learning-Based Model for Acute Aortic Syndrome Detection: A Multisite Validation Study. *J. Clin. Med.* **2025**, *14*, 4797. [CrossRef] [PubMed]
98. Cheng, Z.; Zhao, L.; Yan, J.; Zhang, H.; Lin, S.; Yin, L.; Peng, C.; Ma, X.; Xie, G.; Sun, L. A deep learning algorithm for the detection of aortic dissection on non-contrast-enhanced computed tomography via the identification and segmentation of the true and false lumens of the aorta. *Quant. Imaging Med. Surg.* **2024**, *14*, 7365–7378. [CrossRef] [PubMed]
99. Dong, F.; Song, J.; Chen, B.; Xie, X.; Cheng, J.; Song, J.; Huang, Q. Improved detection of aortic dissection in non-contrast-enhanced chest CT using an attention-based deep learning model. *Heliyon* **2024**, *10*, e24547. [CrossRef] [PubMed]
100. Gao, Y.; Siyu, L.; Jiang, Y.; Hu, X.; Wu, Y.; Li, K.; Guo, Z.; Sun, H. Training an Artificial Intelligence Model for Aortic Dissection Detection Using Non-Contrast Computed Tomography Images from Human Patients. *J. Vis. Exp.* **2026**, *online ahead of print*. [CrossRef] [PubMed]
101. Harris, R.J.; Kim, S.; Lohr, J.; Towey, S.; Velichkovich, Z.; Kabachenko, T.; Driscoll, I.; Baker, B. Classification of Aortic Dissection and Rupture on Post-contrast CT Images Using a Convolutional Neural Network. *J. Digit. Imaging* **2019**, *32*, 939–946. [CrossRef] [PubMed]
102. Hata, A.; Yanagawa, M.; Yamagata, K.; Suzuki, Y.; Kido, S.; Kawata, A.; Doi, S.; Yoshida, Y.; Miyata, T.; Tsubamoto, M.; et al. Deep learning algorithm for detection of aortic dissection on non-contrast-enhanced CT. *Eur. Radiol.* **2021**, *31*, 1151–1159. [CrossRef] [PubMed]
103. Hu, Y.; Xiang, Y.; Zhou, Y.J.; He, Y.; Lang, D.; Yang, S.; Du, X.; Den, C.; Xu, Y.; Wang, G.; et al. AI-based diagnosis of acute aortic syndrome from noncontrast CT. *Nat. Med.* **2025**, *31*, 3832–3844. [CrossRef] [PubMed]
104. Huang, L.T.; Tsai, Y.S.; Liou, C.F.; Lee, T.H.; Kuo, P.T.P.; Huang, H.S.; Wang, C.K. Automated Stanford classification of aortic dissection using a 2-step hierarchical neural network at computed tomography angiography. *Eur. Radiol.* **2022**, *32*, 2277–2285. [CrossRef] [PubMed]
105. Kim, Y.S.; Kim, J.G.; Choi, H.Y.; Lee, D.; Kong, J.W.; Kang, G.H.; Jang, Y.S.; Kim, W.; Lee, Y.; Kim, J.; et al. Detection of Aortic Dissection and Intramural Hematoma in Non-Contrast Chest Computed Tomography Using a You Only Look Once-Based Deep Learning Model. *J. Clin. Med.* **2024**, *13*, 6868. [CrossRef] [PubMed]
106. Laletin, V.; Ayobi, A.; Chang, P.D.; Chow, D.S.; Soun, J.E.; Junn, J.C.; Scudeler, M.; Quenet, S.; Tassy, M.; Avare, C.; et al. Diagnostic Performance of a Deep Learning-Powered Application for Aortic Dissection Triage Prioritization and Classification. *Diagnostics* **2024**, *14*, 1877. [CrossRef] [PubMed]
107. Liu, H.H.; Chang, C.B.; Chen, Y.S.; Kuo, C.F.; Lin, C.Y.; Ma, C.Y.; Wang, L.J. Automated Detection and Differentiation of Stanford Type A and Type B Aortic Dissections in CTA Scans Using Deep Learning. *Diagnostics* **2024**, *15*, 12. [CrossRef] [PubMed]
108. Norajitra, T.; Baumgartner, M.A.; Cusumano, L.R.; Ulloa, J.G.; Rizzo, C.S.; Haag, F.; Hertel, A.; Rathmann, N.A.; Diehl, S.J.; Schoenberg, S.O.; et al. Automated AI detection of thoracic aortic dissection on CT imaging. *Eur. Radiol. Exp.* **2025**, *9*, 102. [CrossRef] [PubMed]
109. Raj, A.; Allababidi, A.; Kayed, H.; Gerken, A.L.H.; Muller, J.; Schoenberg, S.O.; Zollner, F.G.; Rink, J.S. Streamlining Acute Abdominal Aortic Dissection Management-An AI-based CT Imaging Workflow. *J. Imaging Inform. Med.* **2024**, *37*, 2729–2739. [CrossRef] [PubMed]
110. Tang, Z.; Huang, Y.; Hu, S.; Shen, T.; Meng, M.; Xue, T.; Jia, Z. Deep Learning Application of YOLOv8 for Aortic Dissection Screening Using Non-Contrast Computed Tomography. *Eur. J. Vasc. Endovasc. Surg.* **2026**, *71*, 687–695. [CrossRef] [PubMed]
111. Wada, T.; Takahashi, M.; Matsunaga, H.; Kawai, G.; Kaneshima, R.; Machida, M.; Fujita, N.; Matsuoka, Y. An automated screening model for aortic emergencies using convolutional neural networks and cropped computed tomography angiography images of the aorta. *Int. J. Comput. Assist. Radiol. Surg.* **2023**, *18*, 2253–2260. [CrossRef] [PubMed]
112. Wang, Q.; Huang, S.; Pan, W.; Feng, Z.; Lv, L.; Guan, D.; Yang, Z.; Huang, Y.; Liu, W.; Shui, W.; et al. Segmentation-model-based framework to detect aortic dissection on non-contrast CT images: A retrospective study. *Insights Imaging* **2025**, *16*, 207. [CrossRef] [PubMed]
113. Wu, H.; Chen, F.; Xia, R.; Wang, L.; Zhang, Z.; Zhang, J.; Cheng, Q.; Guo, L. Real-time, high-precision detection of aortic dissection key components for emergency triage using an enhanced deep learning model. *Med. Phys.* **2026**, *53*, e70461. [CrossRef] [PubMed]
114. Xiong, X.; Ding, Y.; Sun, C.; Zhang, Z.; Guan, X.; Zhang, T.; Chen, H.; Liu, H.; Cheng, Z.; Zhao, L.; et al. A Cascaded Multi-Task Generative Framework for Detecting Aortic Dissection on 3-D Non-Contrast-Enhanced Computed Tomography. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5177–5188. [CrossRef] [PubMed]
115. Yang, Z.; Xu, S.; Wang, B.; Liu, H.; Chang, J. AI-driven diagnosis of acute aortic syndrome based on multi-modal information fusion. *Sci. Rep.* **2026**, *16*, 8332. [CrossRef] [PubMed]
116. Yi, Y.; Mao, L.; Wang, C.; Guo, Y.; Luo, X.; Jia, D.; Lei, Y.; Pan, J.; Li, J.; Li, S.; et al. Advanced Warning of Aortic Dissection on Non-Contrast CT: The Combination of Deep Learning and Morphological Characteristics. *Front. Cardiovasc. Med.* **2021**, *8*, 762958. [CrossRef] [PubMed]

117. Zhang, R.; Situ, Z.; Cheng, Z.; Luo, Y.; Qiu, X.; He, X.; Yuan, X.; Zhou, Z.; Rong, Z.; Lin, Y.; et al. Adaptive geometric-attention multi-task framework with knowledge distillation for aortic dissection detection in non-contrast CT. *Phys. Med. Biol.* **2026**, *71*, 035011. [CrossRef] [PubMed]
118. Zhuang, C.; Wu, Y.; Qi, Q.; Zhao, S.; Sun, Y.; Hou, J.; Qian, W.; Yang, B.; Qi, S. A Fully Automatic Pipeline of Identification, Segmentation, and Subtyping of Aortic Dissection from CT Angiography. *Cardiovasc. Eng. Technol.* **2025**, *16*, 465–480. [CrossRef] [PubMed]
119. Camara, J.R.; Tomihama, R.T.; Pop, A.; Shedd, M.P.; Dobrowski, B.S.; Knox, C.J.; Abou-Zamzam, A.M.J.; Kiang, S.C. Development of a convolutional neural network to detect abdominal aortic aneurysms. *J. Vasc. Surg. Cases Innov. Tech.* **2022**, *8*, 305–311. [CrossRef] [PubMed]
120. Chatterjee, D.; Shen, T.C.; Mukherjee, P.; Lee, S.; Garrett, J.W.; Zacharias, N.; Pickhardt, P.J.; Summers, R.M. Automated detection of incidental abdominal aortic aneurysms on computed tomography. *Abdom. Radiol.* **2024**, *49*, 642–650. [CrossRef] [PubMed]
121. Chen, D.; Krycia, M.; Avondo, J.; Cavallo, J. Performance assessment of an artificial intelligence algorithm for opportunistic screening of abdominal aortic aneurysms. *Clin. Imaging* **2025**, *125*, 110550. [CrossRef] [PubMed]
122. Golla, A.K.; Tonnes, C.; Russ, T.; Bauer, D.F.; Froelich, M.F.; Diehl, S.J.; Schoenberg, S.O.; Keese, M.; Schad, L.R.; Zollner, F.G.; et al. Automated Screening for Abdominal Aortic Aneurysm in CT Scans Under Clinical Conditions Using Deep Learning. *Diagnostics* **2021**, *11*, 2131. [CrossRef] [PubMed]
123. Mohammadi, S.; Mohammadi, M.; Dehlaghi, V.; Ahmadi, A. Automatic Segmentation, Detection, and Diagnosis of Abdominal Aortic Aneurysm (AAA) Using Convolutional Neural Networks and Hough Circles Algorithm. *Cardiovasc. Eng. Technol.* **2019**, *10*, 490–499. [CrossRef] [PubMed]
124. Rengarajan, B.; Wu, W.; Wiedner, C.; Ko, D.; Muluk, S.C.; Eskandari, M.K.; Menon, P.G.; Finol, E.A. A Comparative Classification Analysis of Abdominal Aortic Aneurysms by Machine Learning Algorithms. *Ann. Biomed. Eng.* **2020**, *48*, 1419–1429. [CrossRef] [PubMed]
125. Spinella, G.; Fantazzini, A.; Finotello, A.; Vincenzi, E.; Boschetti, G.A.; Brutti, F.; Magliocco, M.; Pane, B.; Basso, C.; Conti, M. Artificial Intelligence Application to Screen Abdominal Aortic Aneurysm Using Computed Tomography Angiography. *J. Digit. Imaging* **2023**, *36*, 2125–2137. [CrossRef] [PubMed]
126. Tomihama, R.T.; Camara, J.R.; Kiang, S.C. Machine learning analysis of confounding variables of a convolutional neural network specific for abdominal aortic aneurysms. *JVS Vasc. Sci.* **2023**, *4*, 100096. [CrossRef] [PubMed]
127. Coatsaliou, Q.; Lareyre, F.; Raffort, J.; Webster, C.; Bicknell, C.; Pouncey, A.; Ducasse, E.; Caradu, C. Use of Artificial Intelligence With Deep Learning Approaches for the Follow-Up of Infrarenal Endovascular Aortic Repair. *J. Endovasc. Ther.* **2024**, *online ahead of print*. [CrossRef] [PubMed]
128. Hahn, S.; Perry, M.; Morris, C.S.; Wshah, S.; Bertges, D.J. Machine deep learning accurately detects endoleak after endovascular abdominal aortic aneurysm repair. *JVS Vasc. Sci.* **2020**, *1*, 5–12. [CrossRef] [PubMed]
129. Talebi, S.; Madani, M.H.; Madani, A.; Chien, A.; Shen, J.; Mastrodicasa, D.; Fleischmann, D.; Chan, F.P.; Mofrad, M.R.K. Machine learning for endoleak detection after endovascular aortic repair. *Sci. Rep.* **2020**, *10*, 18343. [CrossRef] [PubMed]
130. Yang, Q.; Hu, J.; Luo, Y.; Jia, D.; Chen, N.; Yao, C.; Wu, R. Detection of Endoleak after Endovascular Aortic Repair Through Deep Learning Based on Non-Contrast CT. *Cardiovasc. Interv. Radiol.* **2024**, *47*, 1267–1275. [CrossRef] [PubMed]
131. Macruz, F.B.D.C.; Lu, C.; Strout, J.; Takigami, A.; Brooks, R.; Doyle, S.; Yun, M.; Buch, V.; Hedgire, S.; Ghoshhajra, B. Quantification of the Thoracic Aorta and Detection of Aneurysm at CT: Development and Validation of a Fully Automatic Methodology. *Radiol. Artif. Intell.* **2022**, *4*, e210076. [CrossRef] [PubMed]
132. Pradella, M.; Achermann, R.; Sperl, J.I.; Kargel, R.; Rapaka, S.; Cyriac, J.; Yang, S.; Sommer, G.; Stieltjes, B.; Bremerich, J.; et al. Performance of a deep learning tool to detect missed aortic dilatation in a large chest CT cohort. *Front. Cardiovasc. Med.* **2022**, *9*, 972512. [CrossRef] [PubMed]
133. Baltaci, H.; Yalcin, S.; Yildirim, M.; Bingol, H. Automatic Diagnosis, Classification, and Segmentation of Abdominal Aortic Aneurysm and Dissection from Computed Tomography Images. *Diagnostics* **2025**, *15*, 2476. [CrossRef] [PubMed]
134. Fujiwara, J.; Orii, M.; Oyamada, K.; Araki, K.; Ogura, M.; Ito, T.; Bunu, D.; Ando, R.; Okada, T.; Yoshioka, K. Automatic detection and measurement system for aortic aneurysms using deep learning-based artificial intelligence. *Int. J. Cardiovasc. Imaging* **2026**, *42*, 845–854. [CrossRef] [PubMed]
135. Wanhainen, A.; Van Herzeele, I.; Bastos Goncalves, F.; Bellmunt Montoya, S.; Berard, X.; Boyle, J.R.; D'oRia, M.; Prendes, C.F.; Karkos, C.D.; Kazimierczak, A.; et al. Editor's Choice—European Society for Vascular Surgery (ESVS) 2024 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. *Eur. J. Vasc. Endovasc. Surg.* **2024**, *67*, 192–331. [CrossRef] [PubMed]

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.