Preprints.org

Article

Diagnostic Performance of a Deep Learning-Powered Application for Aortic Dissection Triage Prioritization and Classification

Vladimir Laletin [*] , Angela Ayobi , Peter D. Chang , Daniel Chow , Jennifer Soun , Jacqueline C. Junn , Marlene Scudeler , Sarah Quenet , Maxime Tassy , Christophe Avare , Mar Roca-Sogorb , Yasmina Chaibi

Article

Diagnostic Performance of a Deep Learning-Powered Application for Aortic Dissection Triage Prioritization and Classification

Vladimir Laletin [1],*, Angela Ayobi [1], Peter Chang [2,3], Daniel Chow [2,3], Jennifer Soun [2,3], Jacqueline C. Junn [4], Marlene Scudeler [1], Sarah Quenet [1], Maxime Tassy [1], Christophe Avare [1], Mar Roca-Sogorb [1] and Yasmina Chaibi [1]

[1]   Avicenna.AI, 375 Avenue du Mistral, 13600 La Ciotat, France.
[2]   Department of Radiological Sciences, University of California Irvine, Irvine, CA 92697, USA.
[3]   Center for Artificial Intelligence in Diagnostic Medicine, University of California Irvine, Irvine, CA 92697, USA.
[4]   Emory University Hospital, Department of Radiology and Imaging Science, Atlanta, GA 30322, USA.
*   Correspondence: vladimir.laletin@avicenna.ai

Abstract: This multi-center retrospective study evaluated the diagnostic performance of a deep learning (DL)-based application for detecting, classifying, and highlighting suspected aortic dissections (ADs) on chest and thoraco-abdominal CT angiography (CTA) scans. CTA scans from over 200 U.S. and European cities acquired on 52 scanner models from 6 manufacturers were retrospectively collected and processed by CINA-CHEST (AD) (Avicenna.AI, La Ciotat, France) device. The diagnostic performance of the device was compared with the ground truth established by the majority agreement of three US board-certified radiologists. Furthermore, the DL-algorithm's time-to-notification was evaluated to demonstrate clinical effectiveness. The study included 1,303 CTAs (mean age 58.8 ± 16.4 years old, 46.7% male, 10.5% positive). The device demonstrated a sensitivity of 94.2% [95% CI: 88.8% – 97.5%] and a specificity of 97.3% [95% CI: 96.2% - 98.1%]. The application classified positive cases by AD types with an accuracy of 99.5% [95% CI: 98.9% – 99.8%] for type A and 97.5 [95% CI: 96.4% – 98.3%] for type B. The application did not miss any type A cases. The device flagged 32 cases incorrectly, primarily due to acquisition artefacts and aortic pathologies mimicking AD.   The mean time to process and notify of potential AD cases was 27.9 ± 8.7 seconds. This deep learning-based application demonstrated strong performance in detecting and classifying aortic dissection cases, potentially enabling faster triage of these urgent cases in clinical settings.

Keywords: deep learning; medical and biomedical image processing; aortic dissection; AI-based solution for radiology; machine learning diagnostic performance; medical imaging automated analysis; emergency radiology

1. Introduction

Acute aortic syndrome (AAS) represents a spectrum of pathological conditions affecting the thoracic and abdominal aorta. Their prevalence ranges from 0.2% to 0.8%, equating to approximately 2.6-3.5 cases per 100,000 individuals in the general population, with a predilection towards males [1,2].   Aortic dissection (AD) constitutes the predominant subtype within the range of AAS, accounting for 85-95% of cases [3]. AD arises from a tear in the aorta's intimal and medial layers, facilitating blood ingress into the aortic wall and the formation of a secondary circulating pathway within the aorta known as the false lumen [4]. This condition presents as a critical, life-threatening emergency with mortality rates at approximately 1-2% per hour within the initial 48 hours [1].   In

cases where prompt surgical intervention is not undertaken, mortality rates escalate to 58% [5]. Therefore, expedited diagnosis is pivotal for effective patient management [6].

In response to the severity of AD, the Stanford classification system was developed, distinguishing between type A and type B dissections. Type A primarily involves the ascending aorta, whereas type B is limited to the descending aorta distal to the left subclavian artery [1,2,4]. Type A AD demands urgent surgical intervention to mitigate the risk of life-threatening complications. Conversely, type B AD can often be managed conservatively through pharmacological means, particularly focusing on hypertension control, although close monitoring and potential intervention remain essential for optimal patient outcomes [1,2].

The clinical presentation of aortic dissection exhibits a lack of specificity, which can result in symptoms resembling those of various other medical conditions. This diagnostic ambiguity poses a significant challenge in the accurate identification of aortic dissection, as its manifestations overlap with a diverse range of diagnoses [5]. Hence, ensuring accurate diagnostics and AD type classification is imperative for providing appropriate treatment.

In recent years, CTA has emerged as the gold standard for assessing aortic dissection, owing to its exceptional precision and effectiveness in identifying this condition [2,3,7]. Despite the high sensitivity and specificity exhibited by CT diagnostics in detecting aortic dissection (AD), the escalating demands within the emergency environment and the onset of fatigue among radiologists could potentially elevate error rates and prolong the time required for AD diagnosis. Thus, deep learning (DL) tools for AD detection and prioritization have been shown to expedite the evaluation process for radiologists, consequently reducing the time required for clinical decision-making [6]. The current investigation sought to scrutinise the efficacy and applicability of a recently introduced and commercially available DL-based tool for AD case prioritization. Specifically, the objectives of the study encompassed the evaluation of the diagnostic accuracy of the evaluated device, its capacity for detection of aortic dissection types, and its potential to reduce the notification time of AD occurrences. This study aims to contribute valuable insights into the utility of DL technologies in enhancing the efficiency and precision of AD diagnosis and clinical patient management.

2. Materials and Methods

2.1. DL-Powered Algorithm for AD Detection: Architecture and Training

Commercially available application (FDA-approved and CE-marked), CINA-CHEST for AD was provided by Avicenna.AI (La Ciotat, France). In the development of the AD DL-powered application, a two-stage approach was employed. The first algorithm was utilized for the segmentation of the aorta, while the second algorithm was employed for the localization of the AD within it, specifically focusing on the visible intimal flap between the true and false lumens. This methodology was selected due to its effectiveness in maximizing the detection of AD in anatomically consistent regions. Both algorithms were based on convolutional neural networks (CNN). A hybrid 3D/2D U-Net variant, known for its robust performance in 2D and 3D segmentation tasks and previously published by Chang et al. [8], was used.

Regarding the ground truth used to train the algorithms, the segmentation of the aorta, encompassing both the true and false lumens while excluding necrosis or intramural hematoma, was performed per slice by two expert radiologists. Per slice segmentation of the visible intimal flap between the false and true lumens was conducted, so that the algorithm can target the localization of the AD.

The training dataset consisted of 649 3D CTA studies sourced from 40 different scanner models, representing various manufacturers (55% GE, 20% Siemens, 10% Philips and 15% Canon), spanning from January 2019 to December 2022. Of these studies, 25% constituted positive cases of aortic dissection, categorized as type A or type B according to the Stanford classification. The age distribution among positive cases was as follows: 8% in the [18-40] age range, 56% in the [41-70] age bracket, and 36% over 70 years old. The cases with confounding conditions such as thoracic or abdominal aneurysms, intramural hematoma, calcifications, and post-surgery instances (e.g., presence of stents) were sought out to enrich the training dataset.

2.2. Data Selection

The retrospective and observational CTA data acquisition for this study spanned from July 2017 to March 2022 and was provided from multiple clinical sources. All data were anonymized and supplied by two teleradiology networks located in the USA and France, comprising more than 200 cities and 6 CT makers. The data were acquired on 52 different scanner models. Informed consent was waived when mandatory, aligning with both national legislation and institutional protocols. Among the received data, CTA cases were consecutively preselected according to the recommended requirements (Table 1). The final dataset consisted of 1,303 cases.

Table 1. CINA-CHEST (AD) inclusion and exclusion criteria.

The inclusion criteria for CINA-CHEST (AD)
Chest or thoraco-abdominal CTA scans
Age ≥ 18 yo
Matrix size ≥ 512 x 512 (rectangular matrix accepted),
Axial acquisition only
Slice thickness ≤ 3 mm with no gap between successive slices
Radiation dose parameters: 60 kVp to 160 kVp
Reconstruction diameter above 200 mm
Density threshold in the Aorta ≥ 140 HU
Soft tissue reconstruction kernel
Field of view including the aortic arch and thoracic aorta

The exclusion criteria for CINA-CHEST (AD)
Parameters not compatible with acquisition protocol
Thoracic aorta out of the field of view
Significant motion artefacts (Uninterpretable images)
Significant streak artefacts (Uninterpretable images)
Significant noise (Uninterpretable images)
Bad bolus timing (Uninterpretable images)

2.3. The Ground Truth

Two U.S. board-certified expert radiologists, with 7- and 6-year-experience in radiology clinical practice, independently visually annotated Chest and Thoraco-Abdominal CTAs and determined the cases with suspected ADs. For positive cases, the ground-truthers defined the AD type according to Sanford classification (type A or type B). A third U.S. board-certified expert radiologist, with 8-year-experience in radiology clinical practice, settled any disagreements. Presence or absence of AD, and AD classification by type were determined by majority agreement. Hyperacute, acute, subacute or chronic ADs, were all considered positive. The radiologists also reported the observed confounding factors such as thoracic or abdominal aneurysms, intramural hematoma, calcifications, and post-surgery instances (e.g., presence of stents).

2.4. Data Processing

The next step entailed processing the same anonymized dataset with CINA-CHEST (AD) v1.0.2 AI-powered application. The application automatically processed incoming CTA, displaying notifications of suspected findings (if any) alongside image series information. For cases flagged positive by the application, the AD type (A or B) was also displayed. All results were gathered for assessment. The evaluation was conducted blindly, without access to the results of the U.S. board-

certified radiologists. Notifications, AD types (for positive cases) and processing times were recorded for all CTA cases, measured from the end of DICOM reception to positive or negative identification.

2.5. Statistical Analysis

The results provided by U.S. board-certified radiologists and those automatically computed by CINA-CHEST (AD) were compared. Sensitivity, specificity, accuracy, and the area under the receiver operating characteristic curve (ROC AUC) were calculated for the complete dataset. 95% confidence intervals (95% CI) for sensitivity, specificity and accuracy were determined using the exact binomial distribution test (Clopper–Pearson). Matthews correlation coefficient (MCC) was also computed to assess binary classification quality. Positive and negative predictive values were derived using sensitivity and specificity accounting for prevalence in the current dataset (10.5%). Subset performance analyses based on imaging acquisition parameters (Manufacturer, Slice Thickness) and patient characteristics (Age and Sex) were conducted. Moreover, the clinical performance of Stanford classification for AD (type A and type B) was determined. Sensitivity, specificity and accuracy were computed for each type of AD. Additionally, AD prioritization and triage effectiveness were evaluated based on the standalone per-case processing time of the device (mean ± SD, 95%CI, and median values) for all cases in the database and for true positive cases only. Statistical analyses were performed using MedCalc Statistical Software (v20.015, MedCalc Software Ltd).

3. Results

3.1. Data Distribution

According to the ground truth, the final dataset included 137 (10.5%) positive and 1,166 (89.5%) negative cases. Among 137 positive cases, 63 and 74 aortic dissections were identified as type A and type B AD, respectively. The mean ± SD age of 1,303 patients included in the study was 58.8 ± 16.4 y/o (min = 18 y/o and max = 90 y/o). Male and female populations were almost equally distributed (46.7% and 53.3%, respectively). The CTA examinations were acquired on different scanner makers: GE Healthcare (259; 19.9%), Philips Healthcare (489; 37.5%), Siemens Healthinners (474; 36.4%), Canon Medical Systems Corporation (formerly Toshiba) (76; 5.8%), Hitachi Ltd. (4; 0.3%), and Philips-Neusoft Medical Systems (PNMS) (1; 0.08%). 456 (35%) scans had slice thickness values lower 1.5 mm, 629 (48%) lower 3mm and 218 (17%) equal to 3mm (Table 2).

Table 2. The dataset characteristics.

Characteristic	Parameters	AD Dataset (1,303 cases)	AD positive cases (137 cases)
Age	Mean ± SD	58.8 ± 16.4 y/o	59.0 ± 13.3 y/o
Sex	Male	609 (46.7%)	84 (61.3%)
	Female	692 (53.3%)	53 (38.7%)
Scanner makers	GE	259 (19.9%)	77 (56,2%)
	Philips	489 (37.5%)	14 (10.2%)
	Siemens	474 (36.4%)	33 (24.1%)
	Canon	76 (5.8%)	13 (9.5%)
	Hitachi	4 (0.3%)	0 (0.0%)
	PNMS	1 (0.08%)	0 (0.0%)
Slice thickness	<1.5mm	456 (35%)	53 (38.7%)
	1.5mm < ST < 3mm	629 (48%)	56 (40.9%)
	=3mm	218 (17%)	28 (20.4%)

3.2. Performance Statistical Results

The DL-based application correctly identified 129 AD cases (True Positives-TP; Figure 1, A and B). 8 cases were missed by the application (false negatives - FN), leading to a sensitivity of 94.2% [95%

CI: 88.8% – 97.5%]. 1,134 cases were correctly labelled as negatives for AD (true negatives - TN) and 32 cases were wrongly flagged as positives (False Positives - FP; Figure 1, C), resulting in a specificity of 97.3% [95% CI: 96.2% - 98.1%]. Among 1,303 scans, no difference was found between the ground truth and AI-based application assessments in 1263 cases, which represents an accuracy of 96.9% [95% CI: 95.8% - 97.8%]. The area under the receiver operating characteristic curve (ROC AUC) was also calculated: 0.96 [95% CI: 0.95 - 0.97]. Matthews correlation coefficient (MCC) was equal to 0.85. Finally, positive and negative predictive values (PPV and NPV) were derived using sensitivity, specificity and prevalence values of the actual dataset (10,5%) and were 80.1% and 99.3%, respectively (Table 3).
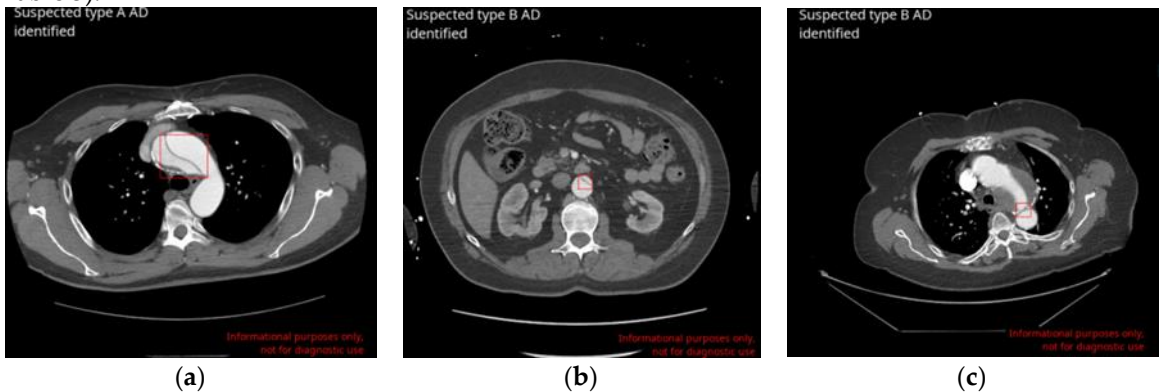

(**a**)   (**b**)   (**c**)

Figure 1. Examples of CINA-CHEST (AD) outputs upon true and false positive AD cases determined by CINA-CHEST (AD). (**a**) Correct detection of type A AD. (**b**) Correct detection of type B AD. (**c**) False-positive identification of a complicated case in the presence of intramural hematoma following aortic repair.

Table 3. CINA-CHEST (AD) performance data.

Characteristic	CINA-CHEST (AD)
Sensitivity [95% CI], %	94.2 [88.8 – 97.5]
Specificity [95% CI], %	97.3 [96.2 - 98.1]
Accuracy [95% CI], %	96.9 [95.8 - 97.8]
ROC AUC [95% CI]	0.96 [0.95 - 0.97]
MCC	0.85
PPV, %	80.1%
NPV, %	99.3%

Among the 8 missed ADs (FNs), 6 are complicated cases and were the subject of disagreements between both truthers. Thus, even visually, the detection of these aortic dissections was not obvious. 4 of these dissections were located within the abdominal infrarenal aorta or the distal descending thoracic aorta. They were all related to AD type B. Among the confounding factors that impacted the correct AD identification by the AI-based algorithm was the combination of acquisition artefacts (1 scan was noisy, 1 - presented streak artefacts, 1 - included motion artefacts) and aortic pathologies (intramural hematoma (IMH) - 4 cases, Penetrating Atherosclerotic Ulcer (PAU) - 2 cases, aortic calcifications - 2 cases, one aneurysm with large mural thrombus and one endoleak with active extravasation of contrast from the graft in the mid descending thoracic aorta). One additional case was missed due to bad contrast filling. The last scan presented AD only within the last two abdominal slices, since the acquisition stopped at the level of the right kidney, thus preventing visualisation of the entire dissection (Figure 2; for more details see Supplementary Table 1).
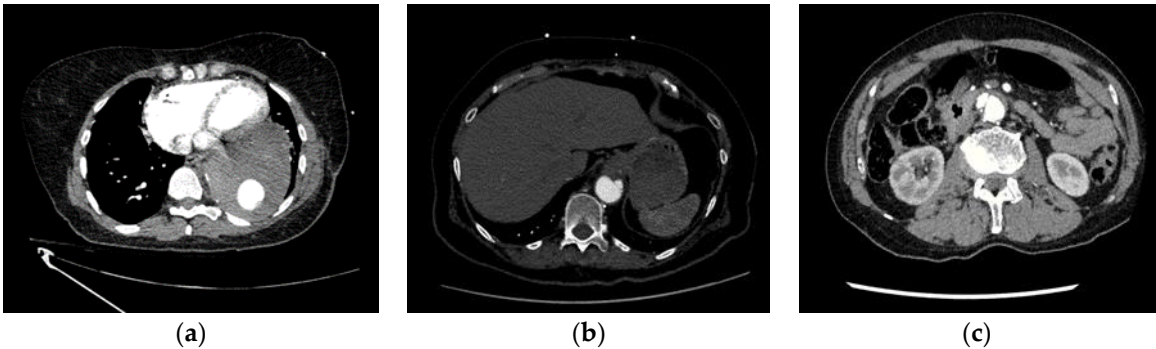
| (a) | (b) | (c) |

Figure 2. Examples of missed AD cases by CINA-CHEST (AD). (**a**) Missed subtle type B AD due to a streak artefact. (**b**) Missed type B AD in the presence of penetrating atherosclerotic ulcer. (**c**) Missed type B AD in the presence of large intramural hematoma and penetrating atherosclerotic ulcer.

There were 32 false positive cases. The inaccurate identification of the dissections, resulting in false positives, stemmed from various factors including inadequate contrast opacification (13 cases), motion artefacts (10 cases), instances of pathology mimicking dissection (i.e. Penetrating Atherosclerotic Ulcer (PAU), intramural hematoma (IMH) and Aneurysms (7 cases)), and interference from stent grafts (2 cases).

The reasons for misdiagnoses (FN and FP) are summarized in Table 4.

Table 4. Reasons for AD misdiagnoses by CINA-CHEST (AD).

Main reasons for false negatives (n=8)	Main reasons for false positives (n=32)
Intramural hematoma (IMH) (4)	Inadequate contrast opacification (13)
Penetrating atherosclerotic Ulcer (PAU) (2)	Motion artefacts (10)
Acquisition artefacts (2)	Instances of pathology mimicking dissection (7)
	Interference from stent grafts (2)

3.3. Stanford AD Type Classification

In the current dataset, the ground truth identified 63 positive AD cases as type A dissections and 74 as type B. Among the 137 positive cases, the AI-based application accurately classified all type A dissections and generated 7 false positives, resulting in a sensitivity of 100% [95% CI: 92.8%-100%] and a specificity of 99.4% [95% CI: 98.8%-99.8%]. For type B AD, the AI-based application produced 8 false negatives and 25 false positives, yielding a sensitivity of 89.2% [95% CI: 79.3%-94.9%] and a specificity of 97.9% [95% CI: 97.0%-98.7%] (Table 5). The application classified positive cases by AD types with an accuracy of 99.5% [95% CI: 98.9% – 99.8%] for type A and 97.5 [95% CI: 96.4% – 98.3%] for type B (Table 5).

Table 5. Stanford AD classification performances of CINA-CHEST (AD) application.

AD Type	Sensitivity [95% CI], %	Specificity [95% CI], %	Accuracy [95% CI], %
Type A	100 [92.8-100]	99.4 [98.8-99.8]	99.5 [98.9-99.8]
Type B	89.2 [79.3-94.9]	97.9 [97.0-98.7]	97.5 [96.4-98.3]

3.4. Stratified Statistical Analysis Results

The stratified statistical analyses are provided for the subgroups of CT-scan makers, image acquisition (slice thickness), age and sex groups (Table 6). The AI-based algorithm performances across these groups are presented in Table 6. The in-depth stratified statistical analysis revealed sensitivities ranging from 89.2% to 100% and specificities from 96.2% to 100%. This comprehensive examination demonstrated that across all categories and within each group, both sensitivities and specificities consistently surpassed the 89% threshold, and the accuracy for all groups was higher than 95%.

Table 6. Detailed stratified statistical analysis of CINA-CHEST (AD) application.

Parameter	Condition	Sensitivity [95% CI], %	Specificity [95% CI], %	Accuracy [95% CI], %
Age	18 ≤ Age < 40	100 [47.8-100]	97.7 [94.3-99.4]	97.9 [94.5-99.4]
	40 ≤ Age ≤ 60	97.1 [89.9-99.6]	98.2 [96.3-99.3]	98.0 [96.3-99.1]
	Age > 60	90.5 [80.4-96.4]	96.5 [94.7-97.8]	95.9 [94.1-97.3]
Sex	Male	96.4 [89.9-99.2]	96.9 [95.1-98.3]	96.9 [95.1-98.1]
	Female	90.6 [79.3-96.9]	97.5 [95.9-98.6]	96.9 [95.4-98.1]
Scanner makers*	GE	94.8 [87.2-98.6]	96.2 [92.2-98.4]	95.8 [92.5-97.9]
	Philips	92.2 [66.1-99.8]	96.8 [94.8-98.2]	96.7 [94.6-98.1]
	Siemens	93.9 [79.8-99.3]	97.7 [95.8-98.9]	97.5 [95.6-98.7]
	Canon	93.3 [62.1-99.6]	100 [92.8-100]	98.7 [92.9-99.9]
Slice thickness	<1.5mm	90.5 [79.3-96.9]	97.7 [95.8-98.9]	96.9 [94.9-98.3]
	1.5mm < ST < 3mm	98.2 [90.5-99.9]	96.7 [94.9-98.0]	96.8 [95.1-98.0]
	=3mm	92.9 [76.5-99.1]	97.8 [94.7-99.4]	97.2 [94.1-99.0]

* No stratification was done for PNMS and Hitachi Ltd. scanner makers as a small number of scans were included with these scanners, 1 and 4 CT scans respectively.

3.5. Time to Notification Evaluation Results

The application was run with the following hardware specifications: CPU: 8 threads/16 cores at 3.0+ GHz and RAM: 16 GB on Ubuntu 22.04.4 LTS. Time-to-notification (TTN) was calculated for all 1,303 cases. The mean TTN ± SD was 27.9 ± 8.7 seconds; 95% CI 27.4-28.3 seconds. The median TTN value was 26.7 seconds. The mean TTN ± SD for 129 true positive cases was 35.4 ± 15.4 seconds, 95% CI 30.8-36.3 seconds, with a median of 33.3 seconds.

4. Discussion

A retrospective, multinational, multicenter, multiscanner blinded study was conducted to evaluate the diagnostic accuracy of a Deep Learning (DL)-based application CINA-CHEST (Avicenna.AI, La Ciotat, France) for prioritisation and triage of aortic dissection (AD) on CTA cases. The database represented a large number of negative AD scans in order to approach the prevalence found in clinical routine. Among the included 1,303 CTA scans, 137 (10.5%) cases were positive for AD, as established by the majority agreement of three US-board certified expert radiologists. The DL-based application correctly labelled 129 cases as positives and 1,134 as negatives, yielding 8 false negatives and 32 false positives. Therefore, the specificity and sensitivity of the application were 94.2% [95%CI: 88.8% – 97.5%] and 97.3% [95%CI: 96.2% - 98.1%], respectively. The application correctly classified the type of dissection with the sensitivity and specificity of 100% and 99.4% for type A and 89.2% and 97.9% for type B, respectively. The main causes of misdiagnoses were mainly pathologies mimicking dissection and acquisition artefacts. Moreover, the mean time to notification for all cases in the current dataset was evaluated (27.9 seconds) and was compatible with the practical use in emergency radiology.

Therefore, the assessed deep learning application successfully conducted screening (triage), classification (type A and type B dissection), and prioritization (operator notification) for the presence of suspected aortic dissections (AD). Several studies in the literature have explored DL-based algorithms aimed at detecting AD and assessed the same parameters (Table 7).

Table 7. Summary of previously published studies pertaining to automated detection solutions for AD.

Parameter	Harris et al. 2019[6]	Hata et al. 2020[9]	Huang et al. 2022[10]	Yi et al. 2022[11]	Current study
Image type	CTA	non-enhanced CT	CTA	non-enhanced CT	CTA

Architecture	5-layer CNN	CNN Xception	2-step network: attention U-net and ResNeXt	deep integrated model: 2.5D U-net, ResNet34	2-step 2.5D U-Net: aorta isolation and dissection detection
Model	2D	2D	3D	3D	3D
Population	34,577 cases (112 AD pos)	170 cases (85 AD pos)	298 cases (51 pos: 22 type A; 29 type B)	452 cases (internal cohort (341): 139 AD pos. external cohort (111): 46 AD pos.)	1,303 cases (137 AD pos)
Enrolment	retrospective	retrospective	retrospective	retrospective	retrospective
Samples	Multicenter, multiscanner	One center	One center	Internal center and external center	Multicenter, multiscanner, multinational
Sensitivity	87.8%	91.8%	Type A: 95.5% Type B: 79.3%	Internal: 86.2% External: 97.8%	All: 94.2% Type A: 100% Type B: 89.2%
Specificity	96.0%	88.2%	Type A: 98.5% Type B: 94.0%	Internal: 92.3% External: 55.4%	All: 97.3% Type A: 99.4% Type B: 97.9%
Features	Triage Mean time to notification: 23.5 ± 21.0 [SD] seconds	Triage Comparison with experts (5 readers): Sensitivity: 90.6% Specificity: 94.1%	Type A/B classification	Triage Comparison with experts (3 readers): Internal experts: Mean Sensitivity: 72.7% Mean Specificity: 98.3% External experts: Mean Sensitivity: 40.6% Mean Specificity: 94.0%	Triage Type A/B classification Mean time to notification: 27.9 ± 8.2 [SD] seconds

*pos - positive cases.

Enhancing the diagnostic capacity of radiologists' AD screening using non-enhanced CT scans was the objective of two studies [9,11]. The diagnostic performance of used DL-based algorithms was similar to or slightly outperformed those of trained radiologists. Hata et al. [9] showed that the sensitivity and specificity of their DL-based application were 91.8% and 88.2%, whereas trained radiologists performed at 90.6% and 94.1%, respectively. Yi et al. [11] thought to improve an DL-based model and demonstrated a diagnostic performance of their internal cohort of 86.2% and 92.3%. However, passing their DL-based algorithm on cases obtained from an external clinical centre drastically dropped the specificity to 55.4%. These findings emphasise the importance of cases originating from different clinical sources, CT scan makers and acquisition parameters for a proper diagnostic performance evaluation. Moreover, both mentioned above studies were conducted with a high prevalence of positive AD cases that do not occur in clinical routine.

Harris et al. [6] evaluated their DL-based tool for AD detection using a multi-centre, multi-scanner approach, using CTA images, with low AD prevalence. The sensitivity and specificity of this application were 87.8% and 96.0%, respectively. In comparison, CINA-CHEST (AD) (Avicenna.AI, La Ciotat) underwent evaluation using scans sourced from multiple clinical sources, various scanner makers and models, diverse acquisition parameters and approaches to real-world disease prevalence.

Therefore, the current performance evaluation, conducted under conditions closely resembling real-world clinical practice, showcased superior diagnostic performance than previously published solutions for AD triage and prioritization. Similar to CINA-CHEST (AD), another deep learning solution for the triage and prioritization of AD, Briefcase for AD (Aidoc Medical, Tel Aviv, ISRAEL), has received regulatory clearance and certification. Briefcase for AD demonstrated a sensitivity of 93.23% (95% CI: 88.70%-96.35%) and a specificity of 92.83% (95% CI: 89.35%-95.45%) on 499 CTAs cases, including 192 positives. Therefore, CINA-CHEST (AD) outperformed this similar solution[12]. Additionally, CINA-CHEST (AD) is unique among certified and commercially available applications in its ability to classify aortic dissection by Stanford classification types.

AD type identification is a crucial feature of AD screening DL-based applications, as it allows clinicians to promptly sort patients in the emergent surgery clinical pipeline (for type A) and conservative treatment pipeline (for type B). Deploying a two-step neural network Huang et al. [10] demonstrated the capacity of a DL-based application to classify AD by type with sensitivity and specificity of 95.5% and 98.5% for type A, and 79.3% and 94.0% for type B. In line with previously published articles, CINA-CHEST (AD) successfully identified all type A AD cases. All 8 missed by the application ADs were type B, demonstrating more complicated automated diagnostics for these cases. In fact, as stated by Yi et al. [11], the diagnostic performance for type B is shown to be lower than for type A. This is due to a wider range of dissections as a larger relative aorta volume is found in the descending aorta than in ascending aorta. However, this does not adversely affect clinical outcomes because type A ADs are life-threatening and require immediate intervention, making accurate detection of these cases significantly more critical.

Finally, Harris et al. [6] measured the notification time of the application from image download into the platform to visible notification of the application results. This time was equal to 23.5 ± 21 seconds. Images flagged as positive were prioritized for readers' evaluation. This prioritization impacted the time of delay (time from the image receiving and image opening for analysis). Cases flagged as positives had significantly reduced median delay time (265 seconds against 660 seconds). Considering that emergent cases must be addressed within 30 minutes, the delay improvement has cut this period by 20% [6]. CINA-CHEST (AD) had a similar notification time of 27.9 ± 8.9 seconds. A similar solution Briefcase for AD (Aidoc Medical, Tel Aviv, ISRAEL) demonstrated a mean notification time of 38 seconds [12].   Even if the evaluation of the application in clinical settings has not been performed yet, we can hypothesise that the clinical benefit in reducing the delay time would be the same for all applications.

The evaluated application CINA-CHEST (AD) generated a few numbers of false positive and false negative cases. The causes of these misdiagnoses were mainly the pathologies mimicking the aortic dissection, like penetrating atherosclerotic ulcer (PAU), intramural hematoma (IMH), aneurysms and acquisition artefacts. For all false positive cases, an alert will be dispatched to the on-call clinician, granting them the opportunity to examine the images and offer the appropriate diagnosis. For all misdiagnosed positive cases, as no notification will be issued to the clinician, these cases will instead undergo review via the standard care workflow, ensuring accurate diagnosis by the physician.

Our study had some limitations. Primarily, it did not include a direct comparison between the performance of the deep learning algorithm and that of a panel of independent radiologists. Hata et al. [9] and Yi et al. [11] assessed the radiologist's performance for AD detection, however, as this evaluation was performed on non-enhanced CT scans this information is not relevant for CTA images. Second, the retrospective selection of CTA cases may introduce selection bias. Prospective studies would not only demonstrate a real-world performance, regarding optimal and sub-optimal CTA scans but also will reveal the supposed clinical benefits regarding time-saving for diagnosis in such emergent conditions as AD. Moreover, radiologists' time saving by automated triage applications might become a crucial benefit in the next few years as the clinician workforce is limited under an ever-increasing workload [13]. Finally, we did not include any clinical parameter in the diagnostic pipeline like Yi et al. [11], who presented an integrated model where aorta morphology was taken into account for AD triage.

To sum up, CINA-CHEST (AD) (Avicenna.AI, La Ciotat) is a DL-based application performing triage, classification and prioritization for aortic dissection. Our multi-national, multi-center, multi-scanner study demonstrated the highest diagnostic performance reported in the literature for this class of devices. The study was performed with a prevalence that approaches to real-world clinical data and the dataset presented a significant distribution among clinical sites, scan vendors and acquisition parameters. This illustrates the device's robustness for extensive use across varied datasets and patient demographics. Moreover, clinical use of this application is associated with prompt time-to-notification that may improve the diagnostic speed and accuracy of clinicians in exigent emergency settings.

References

1. Coady, M.A.; Rizzo, J.A.; Goldstein, L.J.; Elefteriades, J.A. NATURAL HISTORY, PATHOGENESIS, AND ETIOLOGY OF THORACIC AORTIC ANEURYSMS AND DISSECTIONS. Cardiol. Clin. 1999, 17, 615–635, doi:10.1016/S0733-8651(05)70105-3.

2. Awal, S.S.; Prasad, N.; Biswas, S. CT Evaluation of Aortic Dissection and Other Acute Aortic Syndromes: An Update. Int. J. Radiol. Radiat. Ther. 2022, 159–165, doi:10.15406/ijrrt.2022.09.00343.

3. Bossone, E.; LaBounty, T.M.; Eagle, K.A. Acute Aortic Syndromes: Diagnosis and Management, an Update. Eur. Heart J. 2018, 39, 739–749d, doi:10.1093/eurheartj/ehx319.

4. Criado, F.J. Aortic Dissection: A 250-Year Perspective. Tex. Heart Inst. J. 2011, 38, 694–700.

5. Gawinecka, J.; Schönrath, F.; Eckardstein, A. von Acute Aortic Dissection: Pathogenesis, Risk Factors and Diagnosis. Swiss Med. Wkly. 2017, 147, w14489–w14489, doi:10.4414/smw.2017.14489.

6. 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, doi:10.1007/s10278-019-00281-5.

7. Isselbacher, E.M.; Preventza, O.; Hamilton Black, J.; Augoustides, J.G.; Beck, A.W.; Bolen, M.A.; Braverman, A.C.; Bray, B.E.; Brown-Zimmerman, M.M.; Chen, E.P.; et al. 2022 ACC/AHA Guideline for the Diagnosis and Management of Aortic Disease: A Report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. Circulation 2022, 146, e334–e482, doi:10.1161/CIR.0000000000001106.

8.   Chang, P.D.; Kuoy, E.; Grinband, J.; Weinberg, B.D.; Thompson, M.; Homo, R.; Chen, J.; Abcede, H.; Shafie, M.; Sugrue, L.; et al. Hybrid 3D/2D Convolutional Neural Network for Hemorrhage Evaluation on Head CT. Am. J. Neuroradiol. 2018, 39, 1609–1616, doi:10.3174/ajnr.A5742.

9.   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, doi:10.1007/s00330-020-07213-w.

10.   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, doi:10.1007/s00330-021-08370-2.

11.   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. 2022, 8, doi:10.3389/fcvm.2021.762958.

12.   U.S FOOD & DRUGS ADMINISTRATION. BriefCase for AD - K222329. 510(k) Premarket Notification. 2022.

13.   Bharadwaj, P.; Nicola, L.; Breau-Brunel, M.; Sensini, F.; Tanova-Yotova, N.; Atanasov, P.; Lobig, F.; Blankenburg, M. Unlocking the Value: Quantifying the Return on Investment of Hospital Artificial Intelligence. J. Am. Coll. Radiol. JACR 2024, S1546-1440(24)00292-8, doi:10.1016/j.jacr.2024.02.034.