Review

AI Decision Support for Emergency Triage of Large Vessel Occlusion Using Noncontrast Computed Tomography: Systematic Review and Bayesian Diagnostic Test Accuracy Network Meta-Analysis

Huiqing Zhou[1*], MM; Li Liu[2*], MM; Bolun Fu[3*], MD; Bin Cao[4], MM; Jiafu Ma[4], MD; Kun Yang[4], MM; Yuzhen Bao[4], BS; Fengyong Yang[4], MD; Xin Lian[4], MD

[1]Department of Radiology, People's Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

[2]Department of Neurology, Jimo District Hospital of Traditional Chinese Medicine, Qingdao, Shandong, China

[3]Department of Infectious Diseases, Qingdao Public Health Clinical Center, Qingdao, Shandong, China

[4]Department of Emergency Medicine, People's Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

[*]these authors contributed equally

Corresponding Author:
Xin Lian, MD
Department of Emergency Medicine
People's Hospital Affiliated to Shandong First Medical University
No. 1 Xuehu Avenue, North Changshao Road
Jinan, Shandong, 271199
China
Phone: 86 53176279570
Fax: 86 53176279570
Email: sdzylx7@163.com

Abstract

Background: Emergency triage of anterior circulation large vessel occlusion (LVO) is constrained by delays in vascular imaging, specialist interpretation, and transfer decision-making. Noncontrast computed tomography (NCCT) is often obtained first in suspected stroke, but visual recognition of LVO in NCCT images is difficult outside specialist settings. NCCT-based AI may provide an early human-in-the-loop escalation signal.

Objective: This study aimed to compare validated NCCT-based AI paradigms with human reader paradigms for detection of anterior circulation LVO and assess whether current evidence supports prospective evaluation of AI-assisted escalation pathways.

Methods: We searched the PubMed, Embase, Web of Science Core Collection, and Cochrane Library databases from inception to May 24, 2026. Eligible studies evaluated NCCT-based LVO detection in validation datasets independent of model development, included within-study head-to-head comparisons, and provided reconstructible 2 × 2 data. Four nodes were compared: expert readers, nonexpert readers, unimodal imaging AI, and clinically informed multimodal AI. A Bayesian bivariate hierarchical diagnostic test accuracy network meta-analysis estimated sensitivity, specificity, diagnostic odds ratio, and absolute differences. Risk of bias and applicability were assessed using the Prediction Model Risk of Bias Assessment Tool for AI and the complementary Quality Assessment of Diagnostic Accuracy Studies–3, and certainty was rated using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE).

Results: All 10 included studies were retrospective validation studies, comprising 11 validation datasets, 28 study node arms, and 3632 patients. Unimodal imaging AI had a sensitivity of 0.78 (95% credible interval 0.68-0.86), specificity of 0.88 (95% credible interval 0.81-0.94), and diagnostic odds ratio of 30.89 (95% credible interval 14.19-60.27). Expert readers and nonexpert readers had lower sensitivity estimates of 0.62 (95% credible interval 0.48-0.75) and 0.60 (95% credible interval 0.45-0.75), respectively, with similar specificity estimates of 0.86. In league table comparisons, unimodal imaging AI showed higher sensitivity than expert and nonexpert readers by 0.16 (95% credible interval 0.03-0.29) and 0.18 (95% credible interval 0.03-0.32), respectively, with no clear specificity separation. Clinically informed multimodal AI had a sensitivity of 0.81 (95% credible interval 0.64-0.92) and specificity of 0.92 (95% credible interval 0.80-0.98), but this sparse node was connected to human readers only through indirect evidence.

Conclusions: In retrospective validation cohorts, there was low-certainty comparative evidence suggesting that NCCT-based AI, most clearly unimodal imaging AI, had higher sensitivity than human reader paradigms, without a clear difference in specificity. These findings support prospective evaluation of NCCT-based AI as a bounded human-in-the-loop prompt for expert review, computed tomography angiography prioritization, tele-stroke consultation, or transfer discussion. The evidence remains accuracy based and does not establish workflow effectiveness, reperfusion acceleration, or functional outcome benefit.

Trial Registration: PROSPERO CRD420261362111; https://www.crd.york.ac.uk/PROSPERO/view/CRD420261362111

(J Med Internet Res 2026;28:e105068) doi: 10.2196/105068

KEYWORDS

artificial intelligence; AI; large vessel occlusion; noncontrast computed tomography; noncontrast CT; stroke triage; diagnostic test accuracy; network meta-analysis

Introduction

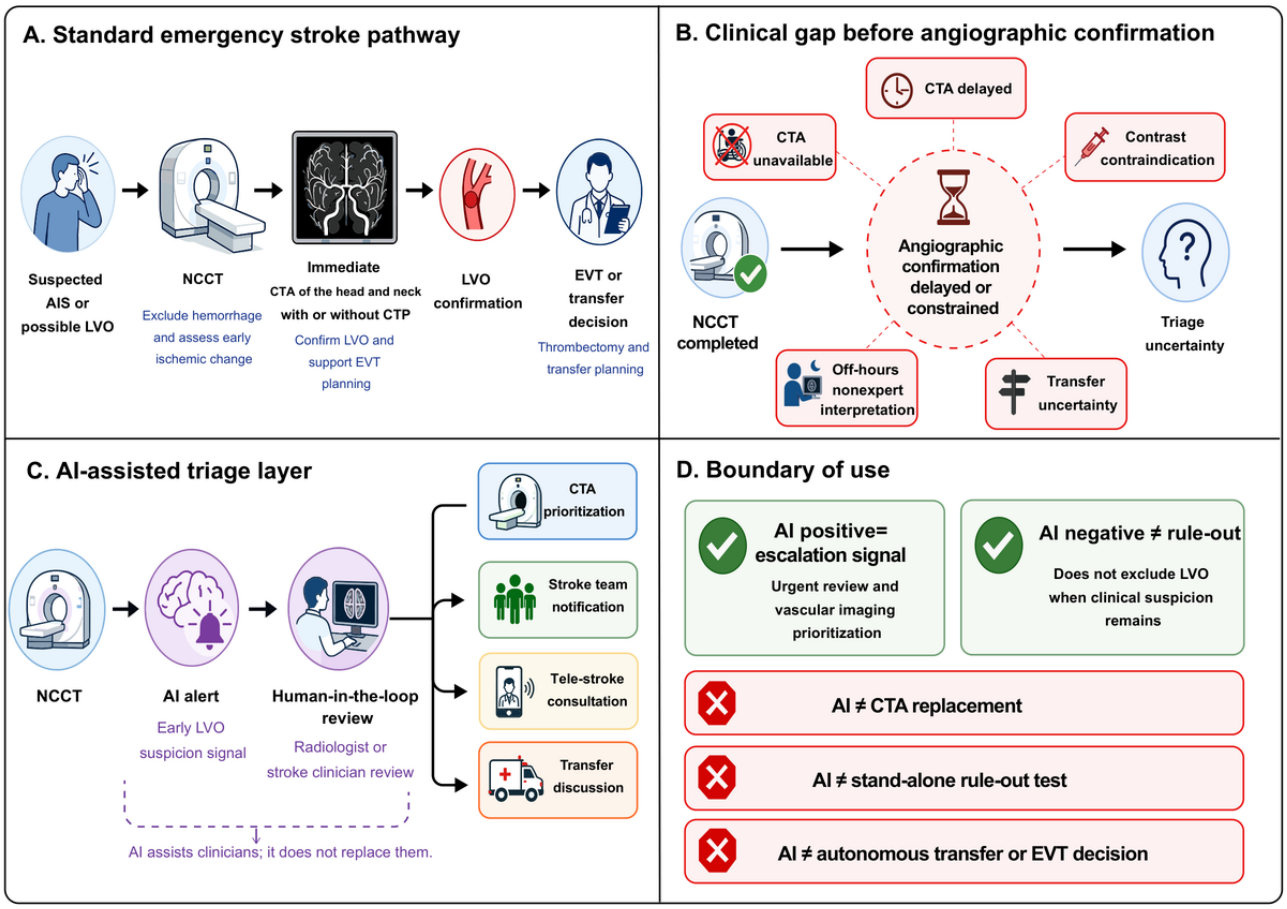
Anterior circulation large vessel occlusion (LVO) is one of the most disabling and time-sensitive subtypes of acute ischemic stroke (AIS) [1,2]. Endovascular thrombectomy (EVT) improves outcomes, but its benefit declines rapidly with treatment delay. In pooled individual patient data from the HERMES collaboration, each 1-hour delay between onset and reperfusion was associated with a 5.2% absolute reduction in functional independence [3]. Interhospital transfer can add substantial delay; the STRATIS (Systematic Evaluation of Patients Treated With Neurothrombectomy Devices for Acute Ischemic Stroke) registry reported a median treatment delay of 110 minutes and a 7.8–percentage point lower rate of functional independence among transferred patients [4]. Rapid LVO recognition at first medical contact is therefore central to emergency stroke triage.

Computed tomography angiography (CTA) is the first-line noninvasive test for confirming anterior circulation LVO in the emergency setting, whereas digital subtraction angiography (DSA) remains the definitive invasive reference standard [5]. Magnetic resonance angiography (MRA) is used less often in the hyperacute phase because of time and compatibility constraints. Immediate CTA, however, is not consistently available in all hospitals or at all hours, particularly in some resource-constrained or off-hours settings. Barriers include equipment limitations, contrast contraindications, gaps in overnight technologist availability, and logistical constraints in remote settings [6]. When CTA cannot be obtained promptly, noncontrast computed tomography (NCCT) often remains the earliest imaging input for triage. Human detection of anterior circulation LVO in NCCT images is limited, with reported sensitivities of 30% to 52% [7-9], and the hyperdense artery sign has a pooled sensitivity of approximately 52% [10].

NCCT-based AI has therefore been proposed as a decision support layer for emergency LVO triage. Several commercial platforms have been evaluated, including Heuron [11], Brainomix [7,12], RapidAI [8], and Methinks AI [9]. The intended role of NCCT-based AI is not to substitute for CTA but to provide an early signal before vascular imaging confirmation or when CTA is not immediately available. That signal could prompt urgent vascular imaging escalation, tele-stroke consultation, or transfer discussion. Existing reviews have addressed related questions. A recent conventional meta-analysis pooled NCCT-based AI studies for LVO prediction but did not incorporate human reader comparators or a connected evidence network [13]. Another review summarized commercial stroke platforms without quantitative synthesis focused on NCCT-based diagnostic accuracy [14].

For digital stroke triage, the evidentiary question is not only whether NCCT-based AI can detect LVO. The more clinically relevant question is whether the diagnostic signal is robust enough to support escalation decisions in human-in-the-loop workflows before or alongside confirmatory vascular imaging. We therefore conducted a systematic review and Bayesian diagnostic test accuracy network meta-analysis (DTA-NMA) to evaluate whether validated NCCT-based AI provides a comparative diagnostic signal that could justify prospective workflow testing as a bounded escalation aid. This clinical positioning is summarized in Figure 1.

Figure 1. Clinical positioning of NCCT-based AI in emergency anterior circulation LVO triage. AI: artificial intelligence; AIS: acute ischemic stroke; CTA: computed tomography angiography; CTP: computed tomography perfusion; EVT: endovascular thrombectomy; LVO: large vessel occlusion; NCCT: noncontrast computed tomography.



Methods

Reporting Guidelines and Protocol Registration

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [15], the PRISMA extension for diagnostic test accuracy (PRISMA-DTA) [16], the PRISMA extension for network meta-analyses (PRISMA-NMA) [17], the PRISMA extension for Abstracts [18], and the PRISMA literature search extension (PRISMA-S) [19]. Completed reporting checklists are provided in Tables S1 to S4 in Multimedia Appendix 1 [15-19]. The protocol was prospectively registered with PROSPERO (CRD420261362111).

Literature Search Strategy

Two investigators (HZ and LL) systematically searched the PubMed, Embase, Web of Science Core Collection, and Cochrane Library databases from inception to May 24, 2026. The search combined controlled vocabulary and free-text terms across 4 concept blocks: ischemic stroke, occlusion location, NCCT, and AI or automated detection. No restrictions on publication date, language, or study type were applied. Citation tracking of included studies and relevant reviews was also performed. The full reproducible search strategy and final database yields are provided in Table S5 in Multimedia Appendix 1. Two reviewers (HZ and LL) independently screened titles, abstracts, and full texts against the eligibility criteria. Disagreements were resolved through discussion, with adjudication by XL when consensus could not be reached.

Eligibility Criteria, Validation Standards, and Node Taxonomy

Eligibility criteria were structured using the participants, index tests, target condition, reference standard, outcomes, and study design framework for diagnostic test accuracy reviews [16]. Participants were adults with suspected AIS who underwent baseline NCCT. Index tests comprised 4 mutually exclusive nodes. Expert readers were senior neuroradiologists, stroke specialists, or experienced radiologists explicitly identified as expert or senior readers or reported as having substantial NCCT interpretation experience in the source publications. Nonexpert readers were nonneuroradiology or non–stroke specialist readers, including general radiologists, emergency physicians, junior radiology or neurology residents, or trainees as specified in the source publications. Unimodal imaging AI comprised fully automated algorithms receiving only NCCT images without clinical information. Clinically informed multimodal AI comprised fully automated algorithms integrating NCCT images with immediately available bedside clinical variables such as the National Institutes of Health Stroke Scale (NIHSS) score and time from symptom onset. Variables requiring laboratory turnaround time, such as blood test results, were excluded. The target condition was predominantly anterior circulation LVO, defined as occlusion of the internal carotid artery or middle

cerebral artery M1 segment with or without M2 involvement; anterior circulation cases were required to constitute 90% or more of the study cohort.

The reference standard was CTA in the vast majority of included studies; DSA, the definitive angiographic gold standard, and MRA were accepted as alternative reference standards when available. One included cohort used mixed vascular reference standards and a broader reference standard timing window; its influence was evaluated through prespecified sensitivity analysis. Outcomes required sufficient data for complete 2 × 2 contingency table reconstruction. Study design required within-study head-to-head comparisons involving at least 2 of the 4 prespecified diagnostic nodes to ensure clinically meaningful network connectivity. This head-to-head requirement restricted the primary evidence base to studies capable of informing comparative network inference; otherwise relevant single-node validation studies excluded under this criterion are summarized in Table S6 [20-22] in Multimedia Appendix 1. The study by Olive-Gadea et al [23] was retained because anterior circulation occlusions constituted 95% of the validation cohort, satisfying the predefined anterior circulation threshold; its influence was evaluated in sensitivity analysis.

Eligible AI studies had to evaluate diagnostic performance in validation cohorts independent of model development. Acceptable validation designs included independent external validation, independently assembled validation cohorts, and temporal internal validation. Studies relying solely on training set performance, random hold-out splits, cross-validation, or bootstrapping were excluded from the main analysis. Studies were also excluded if they were limited to posterior circulation occlusions; evaluated CTA-only, computed tomography perfusion–only, or magnetic resonance imaging–based AI; lacked sufficient data for 2 × 2 table reconstruction; were single-arm diagnostic studies; were nonoriginal reports; or used overlapping cohorts. These criteria were intended to preserve clinical comparability and support the plausibility of the transitivity assumption; statistical consistency was evaluated separately.

Data Extraction

Two investigators (JM and FY) independently extracted data using a standardized form; XL adjudicated discrepancies. Extracted information included study design, country, center type, reference standard, validation design, cohort size, LVO prevalence, target vessel segments, NCCT acquisition parameters, AI software or model type, reader category, and 2 × 2 diagnostic data. For studies reporting both internal and external validation, external validation data were preferentially extracted. For multi-reader studies, source-reported group-level results were used when available. Otherwise, reader-level results were averaged within the prespecified expertise category and converted to an integer 2 × 2 table; no noninteger cell counts were entered into the binomial model. Study-specific derivation is detailed in Table S7 in Multimedia Appendix 1 [7-9,11,12,23-27]. Rai et al [7] reported 2 nonoverlapping external validation subsets with different node compositions; these subsets were entered separately because merging them would have disrupted the node-level data structure (Tables S8 and S9 in Multimedia Appendix 1 [7-9,11,12,23-27]). Source-reported reader expertise, comparator structure, and classification basis for expert reader and nonexpert reader nodes are summarized in Table S7 in Multimedia Appendix 1 [7-9,11,12,23-27].

Quality Assessment and Certainty of Evidence

Two independent reviewers (BF and BC) assessed risk of bias and applicability using the Prediction Model Risk of Bias Assessment Tool for AI (PROBAST+AI) [28]. Diagnostic accuracy risk of bias and applicability were additionally assessed using the Quality Assessment of Diagnostic Accuracy Studies–3 (QUADAS-3) across the participants, index test, target condition, and analysis domains [29]. Discrepancies were resolved by XL. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework was used to rate certainty for node-specific and major comparative sensitivity and specificity estimates [30]. Risk-of-bias judgments within the GRADE framework were informed by the PROBAST+AI and QUADAS-3 assessments. Domain-level judgments and evidence profiles are provided in Tables S10 to S12 [7-9,11,12,23-29], S13 [30], and S14 [7-9,11,12,23-30] in Multimedia Appendix 1.

Diagnostic Accuracy Measures

Sensitivity quantified correct identification of anterior circulation LVO, and specificity quantified correct exclusion of LVO. Because emergency LVO triage is primarily intended to reduce missed occlusions, sensitivity and false negative consequences were prespecified as the primary clinically relevant dimensions. Specificity and false positive transfer burden were interpreted as secondary clinical dimensions. The diagnostic odds ratio (DOR) and surface under the cumulative ranking curve (SUCRA) were treated as global, exploratory summaries rather than the primary basis for clinical ranking [31].

Statistical Analysis

The Bayesian DTA-NMA was performed in R (version 4.5.2; R Foundation for Statistical Computing) using *R2jags*. A bivariate hierarchical random-effects model jointly estimated sensitivity and specificity while accounting for potential threshold effects [32,33]. For cohorts contributing more than one diagnostic node, rows were indexed using a shared cohort identifier and modeled using common cohort-level random effects together with cohort-by-node relative random effects, thereby accounting for clustering of diagnostic arms within the same validation cohort. Weakly informative normal priors (mean 0; precision=0.01 [variance=100]) were assigned to logit-transformed sensitivity and specificity parameters. In the primary model, uniform(0, 2) priors were assigned to the between-study SDs, and a uniform(−0.99, 0.99) prior was assigned to the correlation coefficient ρ. Posterior distributions were estimated using Markov chain Monte Carlo (MCMC) simulation with 4 chains, 10,000 burn-in iterations, and 200,000 sampling iterations per chain; chains were thinned every 10 iterations, yielding 80,000 retained posterior draws for inference. Convergence was assessed via trace plot inspection and the

potential scale reduction factor, with values below 1.05 considered acceptable [34].

Between-study heterogeneity was quantified using τ on the logit scale [35]. Local inconsistency was assessed through node splitting when an independent indirect comparison was identifiable. Global inconsistency was assessed using the design-by-treatment interaction model, with model fit compared using the deviance information criterion (DIC); ΔDIC was defined as DIC for the inconsistency model minus DIC for the consistency model [36]. SUCRA values were calculated as descriptive summaries of diagnostic hierarchy rather than definitive evidence of clinical superiority [37]. League tables of relative DORs (rDORs) and absolute differences in sensitivity and specificity were generated; the rDOR league table is provided in Table S15 in Multimedia Appendix 1. Clinical utility was evaluated using absolute effects per 1000 patients and Fagan nomograms at pretest probabilities of 20%, 30%, and 49.8% [38]. The 20% scenario was used as the primary clinical illustration for suspected AIS or emergency triage populations, 30% represented confirmed or high-suspicion AIS populations [39,40], and 49.8% represented the enriched median prevalence of the included validation cohorts.

Meta-regression explored 4 covariates: NCCT slice thickness (<3 vs ≥3 mm), M2 occlusion inclusion, validation strategy, and algorithm origin (commercial vs laboratory developed). The validation strategy and algorithm origin covariates were restricted to AI nodes because internal vs external model validation and algorithm development origin do not apply to human visual assessment. Five data structure sensitivity analyses were performed. Sensitivity analysis 1 restricted the network to independent external validation and anterior circulation cohorts and excluded the study by Olive-Gadea et al [23]. Sensitivity analysis 2 excluded the sole laboratory-developed AI model reported by Tolhuisen et al [24]. Sensitivity analysis 3 retained M2-including studies only. Sensitivity analysis 4 removed subset A in the study by Rai et al [7]. Sensitivity analysis 5 excluded the US cohort in the study by Sunwoo et al [25]. These analyses evaluated robustness to anterior circulation restriction, validation design, algorithm origin, M2 inclusion, dual subset handling, and mixed or nonconcurrent reference standard characteristics. Small-study effects were assessed using the Deeks funnel plot asymmetry test [41].

To assess robustness to heterogeneity prior assumptions, an additional prior sensitivity analysis (sensitivity analysis 6) refitted the model after replacing the primary uniform(0, 2) priors for the between-study SDs with weakly informative half-normal(0, 1) priors [42]. The likelihood, node definitions, correlation prior, MCMC settings, convergence criteria, and diagnostic data structure were otherwise unchanged. Sensitivity analysis 6 results were compared with those of the primary model for node-level sensitivity, specificity, and DOR and the exploratory diagnostic hierarchy.

Transitivity was assessed by comparing key potential effect modifiers across studies and nodes, including reader expertise, target vessel definition, M2 inclusion, NCCT section thickness, reference standard, LVO prevalence, validation design, algorithm type, and multimodal clinical inputs (Table S16 in Multimedia Appendix 1).

Ethical Considerations

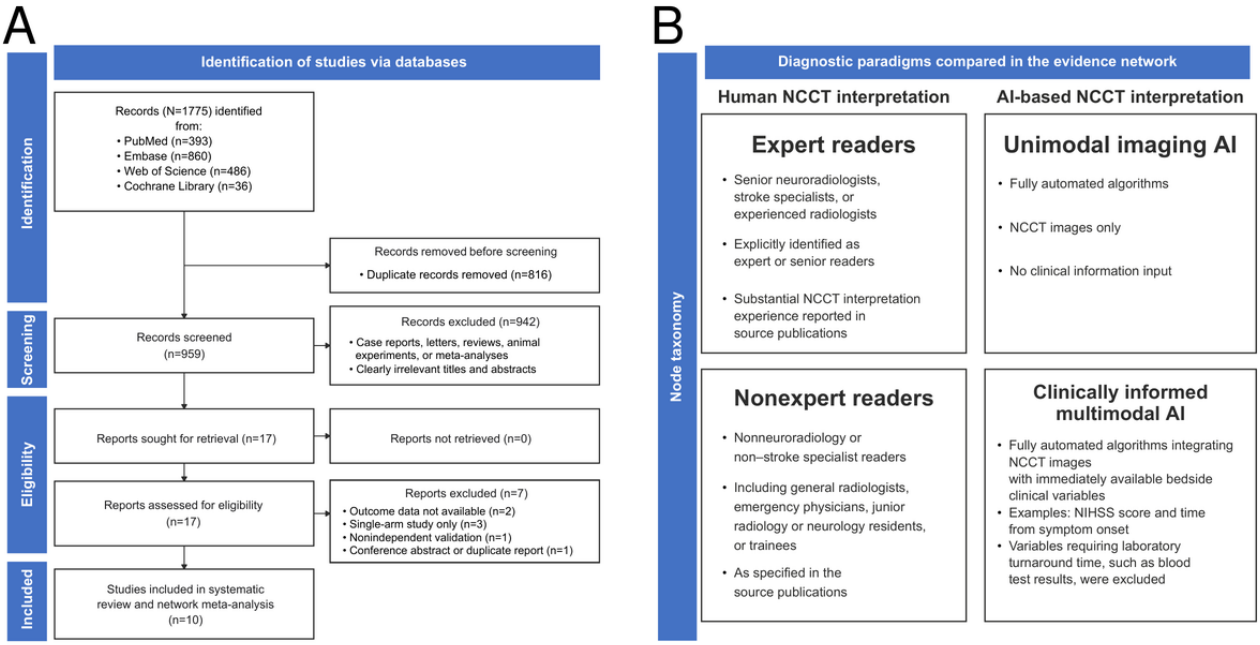
This study used published aggregate data and did not involve new individual-level patient data. Additional institutional review board approval or informed consent was therefore not required.

Results

Literature Search and Study Selection

Database searches up to May 24, 2026, identified 1775 records (PubMed: n=393, 22.1%; Embase: n=860, 48.5%; Web of Science: n=486, 27.4%; Cochrane Library: n=36, 2%). After deduplication and screening, 10 independent studies [7-9,11,12,23-27] encompassing 11 validation cohorts, 28 study node arms, and 3632 patients met the eligibility criteria (Figure 2A). Reasons for exclusion of the single-arm studies are detailed in Table S6 [20-22] in Multimedia Appendix 1.

Figure 2. Study selection and diagnostic node taxonomy: (A) PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing record identification, screening, eligibility assessment, and inclusion; and (B) definitions of the 4 mutually exclusive diagnostic nodes compared in the Bayesian diagnostic test accuracy network meta-analysis. NCCT: noncontrast computed tomography; NIHSS: National Institutes of Health Stroke Scale.



Study Characteristics and Validation Profile

The 10 included studies were published between 2020 and 2026 and were conducted in South Korea, the United States, Spain, the United Kingdom, Germany, Greece, Chile, Brazil, and the Netherlands. Baseline characteristics are summarized in Table 1, with additional validation cohort demographics and NCCT acquisition details provided in Tables S8 and S9 in Multimedia Appendix 1 [7-9,11,12,23-27]. All included studies were retrospective validation studies. In total, 90% (9/10) of the studies evaluated AI software developed by commercial companies, whereas 10% (1/10) of the studies evaluated a laboratory-developed AI model. A total of 90% (9/10) of the studies used independent external validation, and 10% (1/10) used temporal internal validation. CTA was the reference standard in 90% (9/10) of the studies, comprising 3392 participants; 10% (1/10) of the studies, comprising 240 participants, used mixed CTA, MRA, and DSA reference standards. Potential effect modifiers varied across studies, particularly in reader expertise, M2 inclusion, NCCT section thickness, LVO prevalence, and multimodal clinical inputs (Table S16 in Multimedia Appendix 1).

Table 1. Baseline characteristics of the included studies[a].

Studies	Nodes[b]	Design	Center	Source	Country	Reference standard	Training set (LVO[c]/total)	Internal validation (LVO/total)	External validation (LVO/total)
Lee et al [11], 2025	1, 2, and 3	Retrospective	Single center	Local hospital	South Korea	CTA[d]	NR[e]	NR	112/477
Rai et al [7], 2025	1, 2, 3, and 4	Retrospective	Multicenter	Local hospital	United States	CTA	NR	NR	304/612
Yedavalli et al [8], 2023	1, 2, and 3	Retrospective	Multicenter	CRISP[f] and DEFUSE 3[g]	United States, Chile, and Brazil	CTA	NR	NR	115/244
Weyland et al [12], 2022	2 and 3	Retrospective	Multicenter	Local hospital	Germany, Greece, and United Kingdom	CTA	275/500	NR	84/154
Chung et al [26], 2025	1 and 3	Retrospective	Multicenter	Local hospital	South Korea	CTA	NR	NR	35/65
Urra et al [9], 2025	1, 2, and 3	Retrospective	Multicenter	Local hospital	United States and Spain	CTA	NR	NR	71/189
Olive-Gadea et al [23], 2020	3 and 4	Retrospective	Multicenter	Local hospital	Spain	CTA	NR/24214	823/1453	NR
Kim et al [27], 2024	3 and 4	Retrospective	Multicenter	Local hospital	South Korea	CTA	419/2463	45/275	25/95
Tolhuisen et al [24], 2020	1, 2, and 3	Retrospective	Multicenter	MR CLEAN[h]	The Netherlands	CTA	86/86	43/43	58/103
Sunwoo et al [25], 2026 (US cohort)	1, 2, and 3	Retrospective	NR	Segmed	United States	CTA, MRA[i], and DSA[j]	NR	NR	120/240

[a]The study by Rai et al [7] included 2 nonoverlapping external validation subsets; this table reports them as a combined baseline validation cohort, whereas diagnostic analyses entered subsets A and B separately. The study by Olive-Gadea et al [23] contributed a temporal internal validation cohort independent of model development and was excluded in sensitivity analysis 1. Sunwoo et al [25] contributed a Segmed-derived US external validation cohort with mixed reference standards (computed tomography angiography, magnetic resonance angiography, and digital subtraction angiography) and a reference standard window within 24 hours; this cohort was excluded in sensitivity analysis 5. Only validation data independent of model development contributed to the Bayesian diagnostic test accuracy network meta-analysis.

[b]Node 1: expert readers; node 2: nonexpert readers; node 3: unimodal imaging AI; node 4: clinically informed multimodal AI.

[c]LVO: large vessel occlusion.

[d]CTA: computed tomography angiography.

[e]NR: not reported.

[f]CRISP: Computed Tomographic Perfusion to Predict Response to Recanalization in Ischemic Stroke.

[g]DEFUSE 3: Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution 3.

[h]MR CLEAN: Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands.

[i]MRA: magnetic resonance angiography.

[j]DSA: digital subtraction angiography.

Full 2 × 2 Diagnostic Data and Node Composition

Complete 2 × 2 diagnostic data for all 28 study node arms are shown in Table 2. Expert readers were predominantly senior neuroradiologists, stroke specialists, or experienced radiologists, whereas nonexpert readers included general radiologists, emergency physicians, and junior radiology or neurology trainees, as reported in the source studies. These data formed the basis of the Bayesian DTA-NMA and enabled comparison across the 4 predefined diagnostic paradigms (Figure 2B). Node distribution was k=7 for expert readers, k=7 for nonexpert readers, k=11 for unimodal imaging AI, and k=3 for clinically informed multimodal AI. Individual study-level forest plots for sensitivity, specificity, and DOR are provided in Figures S1 to S4 in Multimedia Appendix 1.

Table 2. Diagnostic performance of the included models[a].

Studies and nodes[b]	Reader group or AI model	Input data modality	Clinical variables	Validation design	TPs[c], n	FPs[d], n	FNs[e], n	TNs[f], n	Sensitivity	Specificity
Lee et al [11], 2025										
1	Expert readers	Pure image (NCCT[g])	None	External	85	62	27	303	0.76	0.83
2	Nonexpert readers	Pure image (NCCT)	None	External	75	118	37	247	0.67	0.68
3	Heuron ELVO	Pure image (NCCT)	None	External	99	32	13	333	0.88	0.91
Rai et al [7], 2025										
1—subset A	Expert readers	Pure image (NCCT)	None	External	55	6	58	108	0.49	0.95
2—subset A	Nonexpert readers	Pure image (NCCT)	None	External	53	17	60	97	0.47	0.85
3—subset A	Brainomix 360 Stroke	Pure image (NCCT)	None	External	78	11	35	103	0.69	0.90
3—subset B	Brainomix 360 Stroke	Pure image (NCCT)	None	External	126	12	65	182	0.66	0.94
4—subset B	Brainomix 360 Stroke+NIHSS[h]	Multimodal (NCCT+clinical)	NIHSS	External	124	2	67	192	0.65	0.99
Yedavalli et al [8], 2023										
1	Expert readers	Pure image (NCCT)	None	External	59	17	56	112	0.51	0.87
2	Nonexpert readers	Pure image (NCCT)	None	External	47	4	68	125	0.41	0.97
3	Rapid NCCT Stroke (RapidAI)	Pure image (NCCT)	None	External	73	6	42	123	0.63	0.95
Weyland et al [12], 2022										
2	Nonexpert readers	Pure image (NCCT)	None	External	73	11	11	59	0.87	0.84
3	Brainomix 360 Stroke	Pure image (NCCT)	None	External	65	9	19	61	0.77	0.87
Chung et al [26], 2025										
1	Expert readers	Pure image (NCCT)	None	External	19	4	16	26	0.54	0.87
3	JLK CTL (JLK)	Pure image (NCCT)	None	External	27	2	8	28	0.77	0.93
Urra et al [9], 2025										
1	Expert readers	Pure image (NCCT)	None	External	30	8	41	110	0.42	0.93
2	Nonexpert readers	Pure image (NCCT)	None	External	21	13	50	105	0.30	0.89
3	Methinks LVO (Methinks AI)	Pure image (NCCT)	None	External	58	13	13	105	0.82	0.89
Olive-Gadea et al [23], 2020										
3	Methinks LVO (Methinks AI)	Pure image (NCCT)	None	Internal (temporal)	685	181	138	449	0.83	0.71
4	Methinks LVO+ (Methinks)	Multimodal (NCCT+clinical)	NIHSS and time from onset	Internal (temporal)	684	94	139	536	0.83	0.85

Studies and nodes[b]	Reader group or AI model	Input data modality	Clinical variables	Validation design	TPs[c], n	FPs[d], n	FNs[e], n	TNs[f], n	Sensitivity	Specificity
Kim et al [27], 2024										
3	JLK CTL (JLK)	Pure image (NC-CT)	None	External	20	8	5	62	0.80	0.89
4	JLK CTL+ (JLK)	Multimodal (NC-CT+clinical)	NIHSS	External	23	13	2	57	0.92	0.81
Tolhuisen et al [24], 2020										
1	Expert readers	Pure image (NC-CT)	None	External	54	18	4	27	0.93	0.60
2	Nonexpert readers	Pure image (NC-CT)	None	External	45	5	13	40	0.78	0.89
3	3D CNN[i] (laboratory developed)	Pure image (NC-CT)	None	External	50	14	8	31	0.86	0.69
Sunwoo et al [25], 2026 (US cohort)										
1	Expert readers	Pure image (NC-CT)	None	External	63	16	57	104	0.53	0.87
2	Nonexpert readers	Pure image (NC-CT)	None	External	76	24	44	96	0.63	0.80
3	JLK CTL (JLK)	Pure image (NC-CT)	None	External	95	8	25	112	0.79	0.93

[a]Study node arms represent unique combinations of validation cohort and diagnostic paradigm. For multi-reader studies, source-reported group-level performance was used when available; otherwise, reader-level results were averaged within the prespecified reader category and reconstructed as integer 2 × 2 counts. Detailed derivation is provided in Table S7 [7-9,11,12,23-27] in Multimedia Appendix 1. Laboratory-developed AI denotes a noncommercial algorithm developed by an academic or research group without commercial company involvement. In this review, only the study by Tolhuisen et al [24] met this criterion. All other AI algorithms were developed by commercial companies (Heuron, Brainomix, RapidAI, Methinks AI, and JLK) regardless of whether regulatory clearance had been obtained at the time of data collection.

[b]Node 1: expert readers; node 2: nonexpert readers; node 3: unimodal imaging AI (pure noncontrast computed tomography input); node 4: clinically informed multimodal AI (noncontrast computed tomography plus clinical variables).

[c]TP: true positive.

[d]FP: false positive.

[e]FN: false negative.

[f]TN: true negative.

[g]NCCT: noncontrast computed tomography.

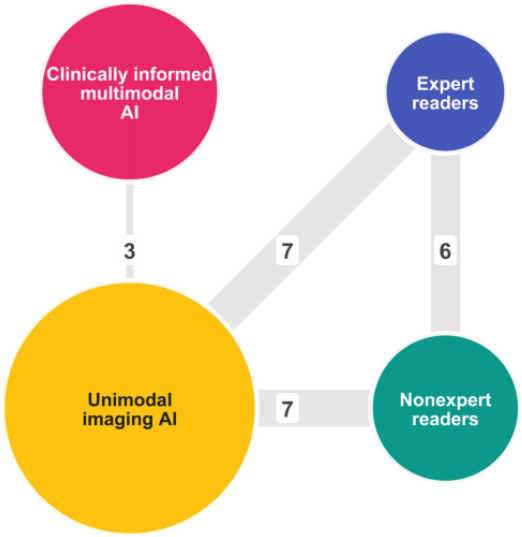
[h]NIHSS: National Institutes of Health Stroke Scale.

[i]CNN: convolutional neural network.

Network Geometry, Model Convergence, and Consistency

The evidence network was connected across the 4 diagnostic paradigms, with edge thickness proportional to the volume of within-study head-to-head comparisons (Figure 3). The clinically informed multimodal AI node was connected to human-reader nodes only through the unimodal imaging AI node; comparisons with human readers were therefore entirely indirect. MCMC diagnostics showed satisfactory convergence for node-level sensitivity, specificity, and DOR estimates, with adequate trace plot mixing, stable posterior density overlays, and potential scale reduction factor values below 1.05. The design-by-treatment interaction model yielded a ΔDIC of −0.19, representing negligible separation in model fit between the inconsistency and consistency models. No compelling global inconsistency signal was identified, although node splitting suggested potential local inconsistency for sensitivity between expert and nonexpert readers and the sparse network limited the power to detect inconsistency. Detailed direct and indirect evidence structures and posterior inconsistency factors are provided in Table S17 in Multimedia Appendix 1. Convergence and consistency diagnostics are shown in Figures S5 to S8 in Multimedia Appendix 1.

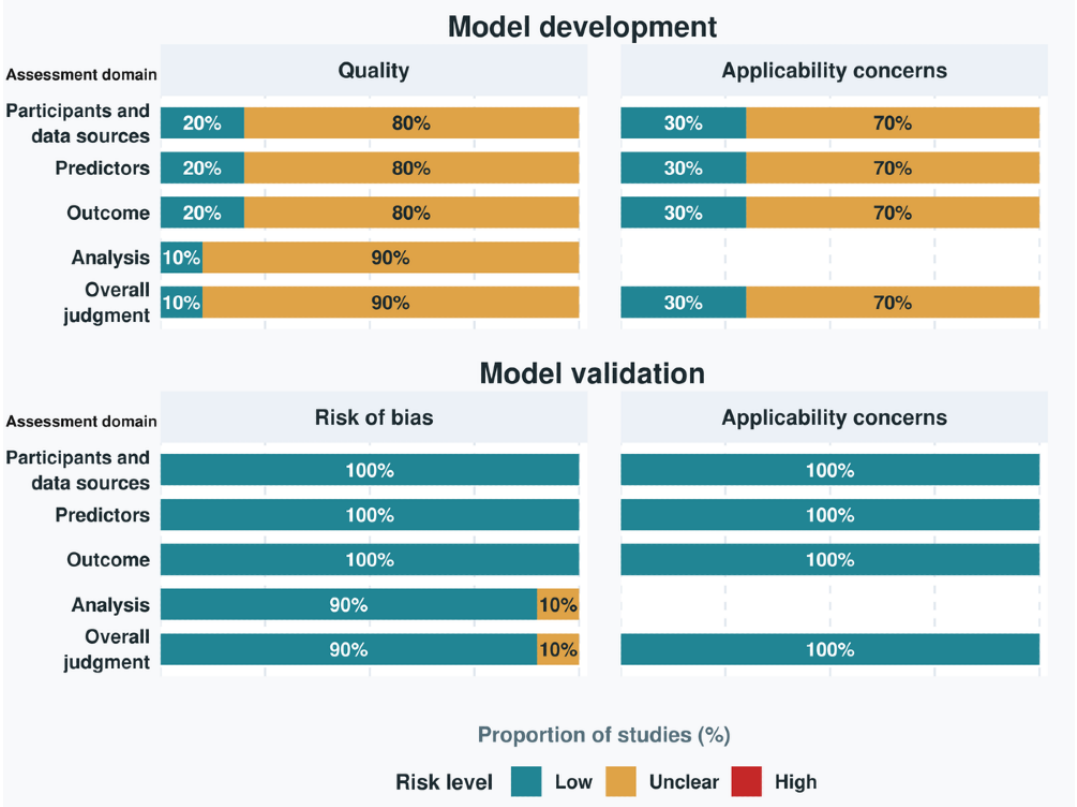
Figure 3. Network geometry of the 4 noncontrast computed tomography (NCCT)–based diagnostic paradigms for detecting anterior circulation large vessel occlusion. Node area is proportional to cumulative patient sample size; edge thickness and superimposed numbers indicate the volume of direct head-to-head comparisons. Node 1: expert readers; node 2: nonexpert readers; node 3: unimodal imaging AI (pure NCCT input); node 4: clinically informed multimodal AI (NCCT plus clinical variables).



Risk of Bias and Certainty of Evidence

Validation-stage PROBAST+AI assessments were low risk in 90% (9/10) of the studies and unclear in 10% (1/10), whereas complementary QUADAS-3 assessments identified high risk in the participants domain in all studies (10/10, 100%) and high participant applicability concerns in 90% (9/10) of the studies. Additional study-specific concerns were identified in the index test, target condition, and analysis domains; overall QUADAS-3 risk of bias and overall applicability concerns were high in all studies (10/10, 100%; Figure 4; Tables S10-S12 [7-9,11,12,23-29] and Figure S9 [29] in Multimedia Appendix 1). GRADE certainty was low for 7 of 8 node-specific estimates and moderate for unimodal imaging AI specificity. Comparative certainty was low for unimodal imaging AI vs either reader node and for clinically informed multimodal AI vs unimodal imaging AI and very low for clinically informed multimodal AI vs either reader node (Tables S13 [30] and S14 [7-9,11,12,23-30] in Multimedia Appendix 1).

Figure 4. Risk-of-bias and applicability assessment using the Prediction Model Risk of Bias Assessment Tool for AI for the 10 included studies.



Primary Diagnostic Performance: Sensitivity, Specificity, and Absolute Effects

Under the prespecified clinical interpretation framework, sensitivity and false negative burden were treated as the primary dimensions for emergency LVO triage, whereas specificity and false positive transfer burden were treated as secondary dimensions. Pooled sensitivity and specificity estimates are summarized in Figure 5 and Figure 6 and in Figures S10 and S11 in Multimedia Appendix 1. Clinically informed multimodal AI had a sensitivity of 0.81 with a 95% credible interval of 0.64 to 0.92 and specificity of 0.92 (95% credible interval 0.80-0.98). Unimodal imaging AI had a sensitivity of 0.78 (95% credible interval 0.68-0.86) and specificity of 0.88 (95% credible interval 0.81-0.94). Expert readers had a sensitivity of 0.62 (95% credible interval 0.48-0.75) and specificity of 0.86 (95% credible interval 0.77-0.93). Nonexpert readers had a sensitivity of 0.60 (95% credible interval 0.45-0.75) and specificity of 0.86 (95% credible interval 0.75-0.93). These estimates indicated a clearer signal for sensitivity than for specificity.

Figure 5. Pooled diagnostic accuracy, certainty of evidence, and absolute diagnostic outcomes per 1000 patients at a 20% pretest probability. DOR: diagnostic odds ratio; FN: false negative; FP: false positive; TN: true negative; TP: true positive.

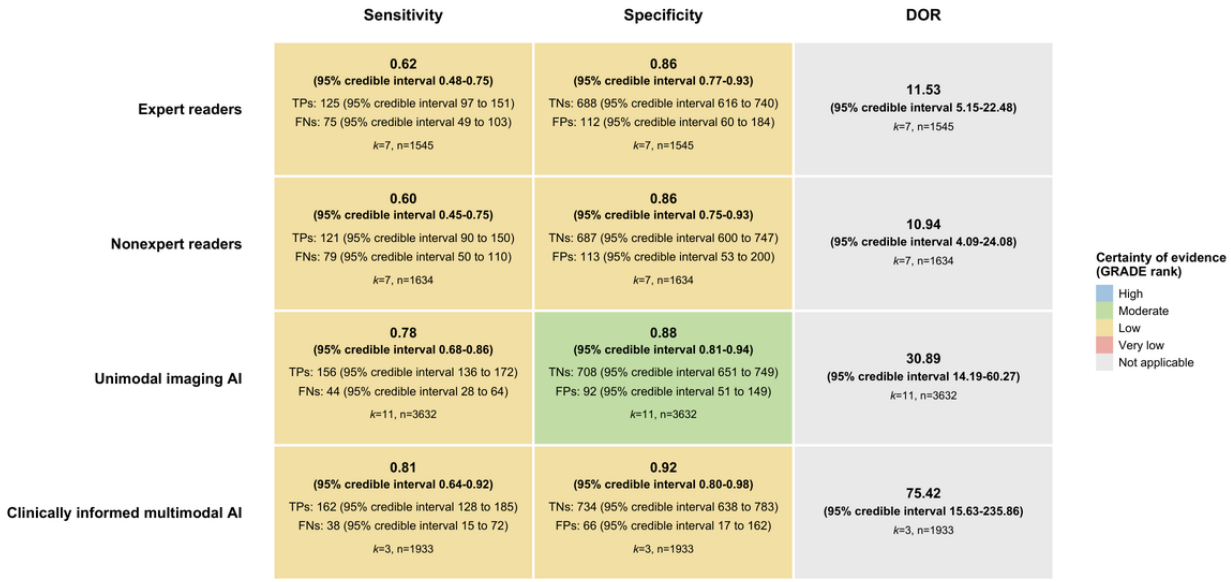
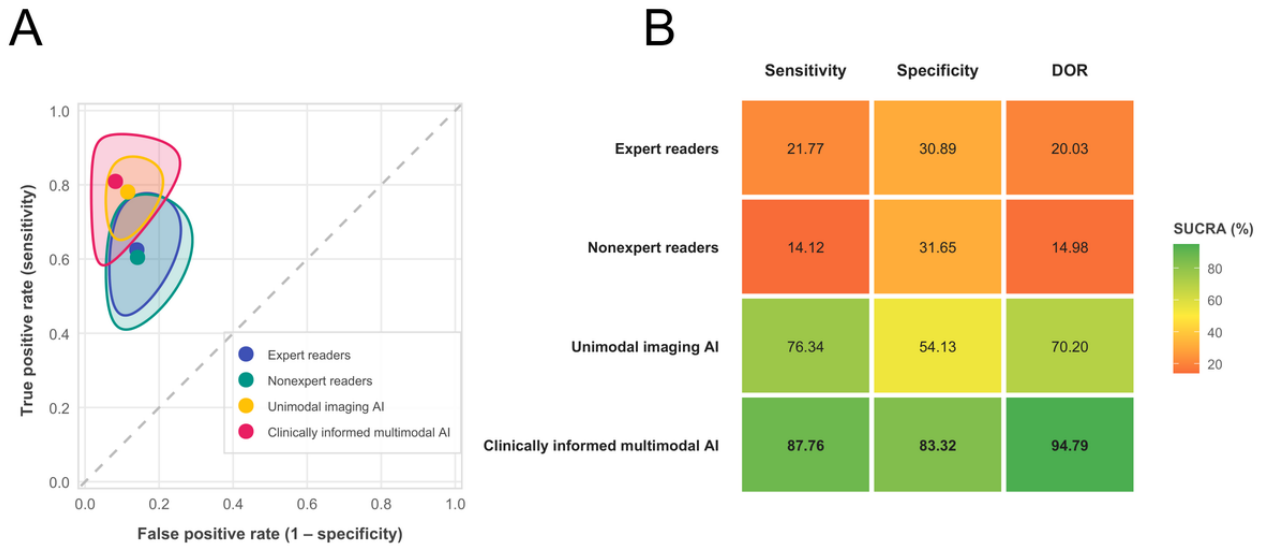


Figure 6. Diagnostic performance across four NCCT-based paradigms: (A) pooled sensitivity and specificity with 95% credible regions; (B) SUCRA values for sensitivity, specificity, and diagnostic odds ratio. DOR: diagnostic odds ratio; SUCRA: surface under the cumulative ranking curve.



Relative Diagnostic Accuracy and Exploratory Rankings

Exploratory DOR estimates followed the same broad pattern: clinically informed multimodal AI had a DOR of 75.42 (95% credible interval 15.63-235.86), unimodal imaging AI had a DOR of 30.89 (95% credible interval 14.19-60.27), expert readers had a DOR of 11.53 (95% credible interval 5.15-22.48), and nonexpert readers had a DOR of 10.94 (95% credible

interval 4.09-24.08; Figure 5; Figure S12 in Multimedia Appendix 1).

The rDOR league table showed higher DOR point estimates for AI diagnostic nodes than for human reader nodes, but these comparisons were interpreted as exploratory summaries rather than direct clinical rankings (Table S15 in Multimedia Appendix 1). The clearest relative sensitivity signal was observed for unimodal imaging AI, which showed higher sensitivity than expert readers and nonexpert readers by 0.16 (95% credible interval 0.03-0.29) and 0.18 (95% credible interval 0.03-0.32), respectively (Figure 7). Clinically informed multimodal AI had higher sensitivity point estimates than human reader nodes, but the corresponding credible intervals were wide and crossed or touched 0. Its rDOR comparison with unimodal imaging AI was also inconclusive at 2.04 (95% credible interval 0.55-7.83). All specificity differences had 95% credible intervals crossing 0 (Figure 7).

Figure 7. League table of absolute differences (the difference on the absolute probability scale, not the mathematical absolute value) in sensitivity and specificity between diagnostic paradigms. Values are posterior means with 95% credible intervals. Upper-right cells show sensitivity differences and lower-left cells show specificity differences, calculated as the column-defining paradigm minus the row-defining paradigm. Accordingly, a value greater than 0.00 indicates higher sensitivity (upper-right cells) or specificity (lower-left cells) for the column-defining diagnostic paradigm. Positive values indicate higher sensitivity or specificity for the column-defining paradigm. Estimates with 95% credible intervals excluding 0 are shown in italics. These comparisons should not be interpreted as product rankings or evidence of clinical superiority. AI: artificial intelligence.

Expert readers	-0.02 (-0.18 to 0.14)	*0.16 (0.03 to 0.29)*	0.18 (-0.01 to 0.36)
0.00 (-0.10 to 0.11)	**Nonexpert readers**	*0.18 (0.03 to 0.32)*	0.20 (-0.01 to 0.39)
-0.02 (-0.12 to 0.06)	-0.03 (-0.14 to 0.07)	**Unimodal imaging AI**	0.03 (-0.13 to 0.16)
-0.06 (-0.17 to 0.08)	-0.06 (-0.19 to 0.08)	-0.03 (-0.12 to 0.09)	**Clinically informed multimodal AI**

SUCRA rankings were interpreted as descriptive summaries of diagnostic hierarchy rather than definitive evidence of clinical superiority. Clinically informed multimodal AI had the highest SUCRA values for sensitivity (87.8%), specificity (83.3%), and DOR (94.8%), followed by unimodal imaging AI for sensitivity (76.3%) and DOR (70.2%). Because clinically informed multimodal AI was supported by only 30% (3/10) of the studies and relied entirely on indirect evidence for comparisons with human readers, its highest point estimates should be treated as exploratory rather than as proof of clinical superiority.

Heterogeneity, Threshold Effect, and Meta-Regression

Between-study heterogeneity (τ) on the logit scale was 0.49 (95% credible interval 0.10-0.91) for sensitivity, 0.61 (95% credible interval 0.12-1.06) for specificity, and 0.51 (95% credible interval 0.17-0.95) for log-DOR, with relatively wide credible intervals indicating limited precision. The sensitivity-specificity correlation was also imprecisely estimated (ρ=−0.58, 95% credible interval −0.97 to 0.34), precluding a firm conclusion regarding a clinically meaningful threshold-related effect. All 4 meta-regression covariates (NCCT slice thickness, M2 segment inclusion, validation strategy, and algorithm origin) had 95% credible intervals crossing 0 (Table S18 in Multimedia Appendix 1). Given the limited statistical power of 10 studies and the single laboratory-developed AI row for the commercial vs laboratory-developed AI comparison, these analyses should be considered exploratory and should not be interpreted as evidence that these factors have no influence on diagnostic performance.

Sensitivity Analyses

Across 6 sensitivity analyses (Table S19 in Multimedia Appendix 1), the exploratory DOR hierarchy remained broadly stable. Sensitivity analyses 1 to 5 addressed validation design, anterior circulation restriction, algorithm origin, M2 inclusion, dual subset handling, and mixed or nonconcurrent reference standard characteristics. Sensitivity analysis 6, which replaced the primary uniform(0, 2) between-study SD priors with half-normal(0, 1) priors, produced no material change in node-level sensitivity, specificity, or DOR estimates or the

exploratory diagnostic hierarchy. AI diagnostic nodes remained directionally favored over human reader nodes for sensitivity and DOR, whereas specificity differences remained unclear.

Clinical Utility and Publication Bias

Clinical utility was primarily illustrated at a 20% pretest probability, approximating suspected AIS or emergency triage populations. Per 1000 patients, unimodal imaging AI yielded approximately 156 true positives, 44 false negatives, 92 false positives, and 708 true negatives compared with 125, 75, 112, and 688 for expert readers, respectively, and 121, 79, 113, and 687 for nonexpert readers, respectively. Corresponding absolute outcomes at 30% and 49.8% pretest probabilities are provided in Table S20 in Multimedia Appendix 1, with Fagan nomograms for all 3 scenarios shown in Figures S13 to S15 in Multimedia Appendix 1. The 49.8% scenario represents the enriched median prevalence of the included validation cohorts rather than routine emergency triage prevalence. The Deeks test was uninformative because of the small number of studies (P=.76; Figure S16 [41] in Multimedia Appendix 1).

Discussion

Principal Findings

In this systematic review and Bayesian DTA-NMA of 10 studies and 3632 patients, NCCT-based AI showed a comparative sensitivity signal for emergency anterior circulation LVO triage. Sensitivity and false negative reduction rather than specificity were the clearest dimensions of potential clinical value. For context, these pooled sensitivities compare favorably with the sensitivity of approximately 52% for the hyperdense artery sign [10], suggesting that AI-assisted NCCT assessment may improve LVO recognition beyond reliance on this conventional visual marker alone. The relative evidence was strongest for NCCT-only AI, whereas the higher point estimates for clinically informed multimodal AI remained exploratory because that node was sparse and relied entirely on indirect evidence for comparisons with human readers. This review extends prior syntheses of NCCT-based AI for LVO and commercial platforms by isolating the NCCT-before-CTA decision point and embedding human reader comparators within the same evidence network [13,14]. Overall, the evidence supports potential triage assistance, but diagnostic certainty should not be conflated with implementation certainty. The GRADE ratings in this review apply to diagnostic accuracy estimates in validation cohorts, not to workflow effectiveness, transfer appropriateness, EVT delivery, or patient outcomes.

Clinical Interpretation

CTA remains the first-line noninvasive confirmatory test for anterior circulation LVO, and DSA remains the definitive invasive angiographic reference standard. NCCT-based AI should therefore not be framed as a replacement for CTA, a rule-out test, or an autonomous transfer decision system. Its clinically plausible role is earlier in the pathway: a human-in-the-loop signal that may prompt urgent image review, CTA prioritization, tele-stroke consultation, stroke team notification, or transfer discussion when vascular imaging is delayed or operationally constrained. Any reduction in missed LVOs must be balanced against false positive escalation, which may increase unnecessary CTA and contrast exposure, avoidable clinical activations or transfer discussions, workload, alert fatigue, and workflow disruption. AI-positive signals should therefore accelerate appropriate vascular confirmation rather than substitute for it.

Workflow and Node Interpretation

The network nodes should be interpreted as workflow functions rather than product rankings. Expert readers provide a specialist benchmark, whereas nonexpert readers approximate first-contact or off-hours interpretation. These are the settings in which escalation is most likely to be delayed. Such settings may also be particularly susceptible to automation bias if AI output is given disproportionate weight during subsequent image interpretation or clinical decision-making, potentially reducing independent critical assessment of the NCCT. This concern may be especially relevant for junior or nonspecialist readers, who had the lowest pooled sensitivity in our network. However, the included retrospective validation studies did not evaluate clinician reliance on AI output, override behavior, or changes in unaided interpretation after AI exposure; therefore, the magnitude of automation bias cannot be quantified from the current evidence. NCCT-based AI should consequently remain an adjunctive human-in-the-loop prompt rather than a substitute for independent clinical assessment and confirmatory vascular imaging. From an operational perspective, unimodal imaging AI can run immediately after image acquisition and does not depend on bedside data entry. The higher point estimates observed for clinically informed multimodal AI may partly reflect the incorporation of additional clinical information, including the NIHSS in all 3 included models and symptom onset time in 1. However, the current network cannot isolate the incremental contribution of these variables from other between-study differences, and this apparent advantage remains uncertain and hypothesis generating because the node was supported by only 3 studies and relied entirely on indirect evidence for comparisons with human readers. The relevant clinical question is therefore not whether AI outperforms ideal specialist interpretation under controlled conditions but whether an early signal can accelerate vascular imaging, specialist review, or transfer discussion in time-sensitive workflows.

Strengths and Limitations

The restrictive eligibility criteria made the network smaller but more interpretable. By requiring validation independent of model development, NCCT input, anterior circulation LVO, within-study head-to-head comparisons, and reconstructible 2 × 2 data, the review prioritized methodological comparability and interpretability over breadth. Accordingly, the included evidence represents comparative NCCT-based validation studies rather than the complete validation literature for NCCT-based LVO detection. The resulting evidence can frame a plausible escalation role, but it cannot estimate product-specific performance, workflow effectiveness, or patient benefit with precision. Generalizability and indirect comparison validity remain constrained by retrospective enriched validation cohorts and by residual variation in transitivity-related modifiers, including reader expertise, target vessel composition, NCCT

acquisition parameters, reference standard characteristics, validation design, and AI input variables. Statistical network connectivity does not necessarily establish clinical exchangeability; accordingly, indirect comparisons should be interpreted as directional comparative signals rather than definitive evidence that AI intrinsically outperforms human readers. The sparse network also limited the power to detect inconsistency, particularly for comparisons involving the multimodal AI node. Shared cohort-level effects were modeled for studies evaluating AI and human readers in the same patients, but aggregate 2 × 2 data precluded explicit modeling of within-patient AI-reader correlation. In addition, category-level averaging of multiple readers may have attenuated interreader variability, and the available data did not support a consistent reader-level sensitivity analysis. The robustness of the diagnostic hierarchy to an alternative half-normal heterogeneity prior supports statistical stability of the Bayesian synthesis, but it does not overcome the need for prospective workflow validation. Together, these constraints define where prospective evaluation must occur: lower-prevalence emergency populations; CTA-constrained or off-hours pathways; sites with different scanner protocols; and predefined action pathways linking an AI-positive NCCT alert to vascular imaging, tele-stroke consultation, or transfer discussion.

Future Directions

Future studies should move from stand-alone diagnostic accuracy to pathway effects. Deployment evaluations should predefine how an AI-positive NCCT alert is acted on and measure picture archiving and communication system integration, alert latency, clinician response, override behavior, false positive transfer burden, alert fatigue, NIHSS data entry burden, site-level generalizability, and equity of access [43]. Clinical deployment also requires appropriate IT infrastructure; integration with existing imaging and stroke workflows; continued clinician oversight; and, for multimodal systems, structured electronic health record interfaces or other standardized real-time access to relevant clinical variables. Early real-world studies have begun to evaluate NCCT-based AI implementation, whereas a cluster randomized clinical trial of a related CTA-based LVO workflow suggested that AI-assisted notification may shorten treatment times. However, prospective evidence demonstrating improved patient outcomes specifically from NCCT-based AI remains limited [44,45]. Prospective multicenter studies should also prespecify subgroup analyses by age, sex, stroke severity, scanner manufacturer, imaging protocol, hospital type, geographic setting, and relevant demographic groups to assess whether diagnostic performance and clinical utility are consistent across patient populations and implementation environments. Stroke-specific end points should include door-to-CTA time, door-in-door-out time, door-to-puncture time, transfer appropriateness, and 90-day modified Rankin Scale outcomes. Without predefined action pathways, improved diagnostic sensitivity may not translate into faster reperfusion or better functional outcomes. Future reports should also follow current AI-specific prediction model and diagnostic accuracy reporting guidelines [46,47].

Conclusions

Taken together, the current evidence supports prospective evaluation of NCCT-based AI as a bounded escalation and prioritization signal before or alongside confirmatory vascular imaging, with the clearest comparative signal observed for unimodal imaging AI sensitivity. The current evidence does not support CTA replacement, LVO rule-out, autonomous transfer decisions, or EVT decisions. Prospective workflow-integrated validation is required before accuracy evidence can justify deployment claims, reimbursement decisions, or protocolized changes in stroke triage pathways.

Acknowledgments

OpenAI GPT-5.6 Sol was used to assist with language polishing, wording consistency, and editorial clarity. AI tools were not used to search, screen, and select studies; extract data; assess risk of bias; run statistical analyses; generate study data; or interpret Bayesian diagnostic test accuracy network meta-analysis outputs. The authors reviewed and edited all AI-assisted outputs, verified the accuracy of all manuscript content, and take full responsibility for the final manuscript.

Funding

This work was supported by the Jinan Clinical Medical Specialty Construction Project – Qilu “Summit” Plan (Department of Emergency Medicine, People’s Hospital Affiliated to Shandong First Medical University), the Science and Technology Development Program of Jinan Municipal Health Commission (2023-2-56), the Shandong Provincial Key Discipline of Medical and Health Sciences (Lu Wei Ke Jiao Zi [2022] No. 3), and the Clinical Medical Research Center special fund (202101002). The funders had no role in the study design; data extraction, analysis, and interpretation; manuscript preparation; or decision to submit the work for publication.

Data Availability

All extracted aggregate data required to reproduce this review are provided in the main text, tables, figures, Multimedia Appendix 1, and the public code repository. The R and Just Another Gibbs Sampler scripts used for Bayesian diagnostic test accuracy network meta-analysis, convergence diagnostics, and sensitivity analyses are available at GitHub [48].

Authors' Contributions

XL contributed to conceptualization, methodology, formal analysis, visualization, supervision, and writing—original draft. HZ and LL contributed to literature search and investigation. BF and BC contributed to risk-of-bias assessment and validation. JM and FY contributed to data curation and data extraction. KY and YB contributed to project administration and logistical support. XL adjudicated screening, extraction, and risk-of-bias discrepancies. XL and HZ accessed and verified the aggregate data reported in the manuscript. All authors contributed to writing—review and editing, approved the final manuscript, and agreed to be accountable for the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Reporting checklists, search strategies, study characteristics, risk-of-bias and certainty assessments, and supplementary analyses for the Bayesian diagnostic test accuracy network meta-analysis.
[DOCX File , 3682 KB-Multimedia Appendix 1]

References

1. Ospel JM, Holodinsky JK, Goyal M. Management of acute ischemic stroke due to large-vessel occlusion: JACC focus seminar. J Am Coll Cardiol. Apr 21, 2020;75(15):1832-1843. [FREE Full text] [doi: 10.1016/j.jacc.2019.10.034] [Medline: 32299595]
2. Prabhakaran S, Gonzalez NR, Zachrison KS, Adeoye O, Alexandrov AW, Ansari SA, et al. 2026 guideline for the early management of patients with acute ischemic stroke: a guideline from the American Heart Association/American Stroke Association. Stroke. Aug 2026;57(8):e316-e436. [FREE Full text] [doi: 10.1161/STR.0000000000000513] [Medline: 41582814]
3. Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CB, Dippel DW, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. JAMA. Sep 27, 2016;316(12):1279-1288. [doi: 10.1001/jama.2016.13647] [Medline: 27673305]
4. Froehler MT, Saver JL, Zaidat OO, Jahan R, Aziz-Sultan MA, Klucznik RP, et al. Interhospital transfer before thrombectomy is associated with delayed treatment and worse outcome in the STRATIS registry (systematic evaluation of patients treated with neurothrombectomy devices for acute ischemic stroke). Circulation. Dec 12, 2017;136(24):2311-2321. [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.028920] [Medline: 28943516]
5. Al-Salahat A, Pirahanchi Y, Dhasakeerthi T, Nayar D, Almasri S, Verma K, et al. Current state and advancements of imaging in acute ischemic stroke: a practical review. Neurol Sci. Dec 22, 2025;47(1):29. [doi: 10.1007/s10072-025-08685-8] [Medline: 41428102]
6. Kim J, Olaiya MT, De Silva DA, Norrving B, Bosch J, De Sousa DA, et al. Global stroke statistics 2023: availability of reperfusion services around the world. Int J Stroke. Mar 2024;19(3):253-270. [FREE Full text] [doi: 10.1177/17474930231210448] [Medline: 37853529]
7. Rai AT, Al Halak A, Abdalkader M, Kaliaev A, Nguyen TN, Kallmes DF, et al. Artificial intelligence–driven detection of large vessel occlusions on NCCT: a multi-institutional study. Am J Neuroradiol. Dec 1, 2025;46(12):2528-2534. [doi: 10.3174/ajnr.a8923]
8. Yedavalli V, Heit JJ, Dehkharghani S, Haerian H, Mcmenamy J, Honce J, et al. Performance of RAPID noncontrast CT stroke platform in large vessel occlusion and intracranial hemorrhage detection. Front Neurol. Nov 24, 2023;14:1324088. [FREE Full text] [doi: 10.3389/fneur.2023.1324088] [Medline: 38156093]
9. Urra X, Rai A, Hernandez M, Lopes D, Oleaga L, Jovin T, et al. Evaluation of a deep learning tool for detecting large vessel occlusion and intracranial hemorrhage on noncontrast computed tomography scans. Stroke Vasc Interv Neurol. Nov 14, 2025;5(6):e001872. [doi: 10.1161/SVIN.125.001872] [Medline: 41608726]
10. Mair G, Boyd EV, Chappell FM, von Kummer R, Lindley RI, Sandercock P, et al. Sensitivity and specificity of the hyperdense artery sign for arterial obstruction in acute ischemic stroke. Stroke. Jan 2015;46(1):102-107. [FREE Full text] [doi: 10.1161/STROKEAHA.114.007036] [Medline: 25477225]
11. Lee SJ, Kim D, Choi DH, Lim YS, Park G, Jung S, et al. Using a deep learning-based decision support system to predict emergent large vessel occlusion using non-contrast computed tomography. J Clin Med. Jun 30, 2025;14(13):4635. [FREE Full text] [doi: 10.3390/jcm14134635] [Medline: 40649010]
12. Weyland CS, Papanagiotou P, Schmitt N, Joly O, Bellot P, Mokli Y, et al. Hyperdense artery sign in patients with acute ischemic stroke-automated detection with artificial intelligence-driven software. Front Neurol. Apr 5, 2022;13:807145. [FREE Full text] [doi: 10.3389/fneur.2022.807145] [Medline: 35449516]
13. Umam HF, Mustofa A, Umam DN, Zidny SN, Pranani D, Retnaningsih. Automated prediction of large vessel occlusion using artificial intelligence in non-contrast computed tomography: a systematic review and meta-analysis. Magna Neurol. 2025;3(2):132-137. [doi: 10.20961/magnaneurologica.v3i2.1704]

14. Dorochowicz M, Kacała A, Tołkacz A, Kosikowska A, Gewald M, Guziński M. Transforming stroke diagnosis with artificial intelligence: a scoping review of Brainomix e-Stroke, Aidoc, RapidAI, and Viz.ai. Medicina (Kaunas). Mar 19, 2026;62(3):582. [FREE Full text] [doi: 10.3390/medicina62030582] [Medline: 41901663]
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. Mar 29, 2021;372:n71. [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
16. Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ. Aug 14, 2020;370:m2632. [FREE Full text] [doi: 10.1136/bmj.m2632] [Medline: 32816740]
17. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. Jun 02, 2015;162(11):777-784. [FREE Full text] [doi: 10.7326/M14-2385] [Medline: 26030634]
18. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, et al. PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. PLoS Med. 2013;10(4):e1001419. [FREE Full text] [doi: 10.1371/journal.pmed.1001419] [Medline: 23585737]
19. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. Jan 26, 2021;10(1):39. [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
20. Sanders JV, Keigher K, Oliver M, Joshi K, Lopes D. Methinks AI software for identifying large vessel occlusion in non-contrast head CT: a pilot retrospective study in American population. Interv Neuroradiol. Jul 25, 2025:15910199251362073. [doi: 10.1177/15910199251362073] [Medline: 40708398]
21. Fussell DA, Lopez JL, Chang PD. A deep learning model to detect acute MCA occlusion on high-resolution noncontrast head CT. AJNR Am J Neuroradiol. Feb 03, 2026;47(2):386-393. [doi: 10.3174/ajnr.A8954] [Medline: 40780878]
22. You J, Tsang AC, Yu PL, Tsui EL, Woo PP, Lui CS, et al. Automated hierarchy evaluation system of large vessel occlusion in acute ischemia stroke. Front Neuroinform. 2020;14:13. [FREE Full text] [doi: 10.3389/fninf.2020.00013] [Medline: 32265682]
23. Olive-Gadea M, Crespo C, Granes C, Hernandez-Perez M, Pérez de la Ossa N, Laredo C, et al. Deep learning based software to identify large vessel occlusion on noncontrast computed tomography. Stroke. Oct 2020;51(10):3133-3137. [doi: 10.1161/STROKEAHA.120.030326] [Medline: 32842922]
24. Tolhuisen ML, Ponomareva E, Boers AM, Jansen IG, Koopman MS, Sales Barros R, et al. A convolutional neural network for anterior intra-arterial thrombus detection and segmentation on non-contrast computed tomography of patients with acute ischemic stroke. Appl Sci. Jul 15, 2020;10(14):4861. [doi: 10.3390/app10144861]
25. Sunwoo L, Ryu WS, Buch K, Conklin J, Mehan W, Desalvo MN, et al. AI assisted detection of large vessel occlusion on non-contrast CT: multinational validation and reader study. J Neurointerv Surg (Forthcoming). Jun 02, 2026:jnis-2026-025339. [doi: 10.1136/jnis-2026-025339] [Medline: 42173676]
26. Chung JW, Lee M, Ha SY, Kim PE, Sunwoo L, Kim N, et al. Multicenter validation of artificial intelligence predicting anterior circulation large vessel occlusion using noncontrast head CT. Stroke Vasc Interv Neurol. Jul 30, 2025;5(5):e001788. [doi: 10.1161/SVIN.125.001788] [Medline: 41573318]
27. Kim PE, Yang H, Kim D, Sunwoo L, Kim CK, Kim BJ, et al. Automated prediction of proximal middle cerebral artery occlusions in noncontrast brain computed tomography. Stroke. Jun 2024;55(6):1609-1618. [FREE Full text] [doi: 10.1161/STROKEAHA.123.045772] [Medline: 38787932]
28. Moons KG, Damen JA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. Mar 24, 2025;388:e082505. [doi: 10.1136/bmj-2024-082505] [Medline: 40127903]
29. Whiting PF, Tomlinson E, Rutjes AW, Davenport CF, Yang B, Westwood ME, et al. QUADAS-3: a revised tool for the quality assessment of diagnostic test accuracy studies. Ann Intern Med. Apr 2026;179(4):548-555. [FREE Full text] [doi: 10.7326/ANNALS-25-02104] [Medline: 41698208]
30. Gopalakrishna G, Mustafa RA, Davenport C, Scholten RJ, Hyde C, Brozek J, et al. Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) to diagnostic tests was challenging but doable. J Clin Epidemiol. Jul 2014;67(7):760-768. [FREE Full text] [doi: 10.1016/j.jclinepi.2014.01.006] [Medline: 24725643]
31. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. Nov 2003;56(11):1129-1135. [doi: 10.1016/s0895-4356(03)00177-x] [Medline: 14615004]
32. Ma X, Lian Q, Chu H, Ibrahim JG, Chen Y. A Bayesian hierarchical model for network meta-analysis of multiple diagnostic tests. Biostatistics. Jan 01, 2018;19(1):87-102. [FREE Full text] [doi: 10.1093/biostatistics/kxx025] [Medline: 28586407]
33. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. Oct 2005;58(10):982-990. [doi: 10.1016/j.jclinepi.2005.02.022] [Medline: 16168343]
34. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat. 1998;7(4):434-455. [doi: 10.1080/10618600.1998.10474787]

35. Owen RK, Cooper NJ, Quinn TJ, Lees R, Sutton AJ. Network meta-analysis of diagnostic test accuracy studies identifies and ranks the optimal diagnostic tests and thresholds for health care policy and decision-making. J Clin Epidemiol. Jul 2018;99:64-74. [FREE Full text] [doi: 10.1016/j.jclinepi.2018.03.005] [Medline: 29548843]
36. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. Mar 30, 2010;29(7-8):932-944. [doi: 10.1002/sim.3767] [Medline: 20213715]
37. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. Feb 2011;64(2):163-171. [doi: 10.1016/j.jclinepi.2010.03.016] [Medline: 20688472]
38. Fagan TJ. Letter: nomogram for Bayes's theorem. N Engl J Med. Jul 31, 1975;293(5):257. [doi: 10.1056/NEJM197507312930513] [Medline: 1143310]
39. Waqas M, Rai AT, Vakharia K, Chin F, Siddiqui AH. Effect of definition and methods on estimates of prevalence of large vessel occlusion in acute ischemic stroke: a systematic review and meta-analysis. J Neurointerv Surg. Mar 2020;12(3):260-265. [doi: 10.1136/neurintsurg-2019-015172] [Medline: 31444289]
40. Lakomkin N, Dhamoon M, Carroll K, Singh IP, Tuhrim S, Lee J, et al. Prevalence of large vessel occlusion in patients presenting with acute ischemic stroke: a 10-year systematic review of the literature. J Neurointerv Surg. Mar 2019;11(3):241-245. [doi: 10.1136/neurintsurg-2018-014239] [Medline: 30415226]
41. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol. Sep 2005;58(9):882-893. [doi: 10.1016/j.jclinepi.2005.01.016] [Medline: 16085191]
42. Röver C, Bender R, Dias S, Schmid CH, Schmidli H, Sturtz S, et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. Res Synth Methods. Jul 2021;12(4):448-474. [doi: 10.1002/jrsm.1475] [Medline: 33486828]
43. Bosch J, Lotlikar R, Melifonwu R, Roushdy T, Sebastian I, Abraham SV, et al. Prehospital stroke care in low- and middle-income countries: a World Stroke Organization (WSO) scientific statement. Int J Stroke. Oct 2025;20(8):918-927. [FREE Full text] [doi: 10.1177/17474930251351867] [Medline: 40495741]
44. Lim YS, Kim E, Choi WS, Yang HJ, Moon JY, Jang JH, et al. Non-contrast computed tomography-based triage and notification for large vessel occlusion stroke: a before and after study utilizing artificial intelligence on treatment times and outcomes. J Clin Med. Feb 15, 2025;14(4):1281. [FREE Full text] [doi: 10.3390/jcm14041281] [Medline: 40004811]
45. Martinez-Gutierrez JC, Kim Y, Salazar-Marioni S, Tariq MB, Abdelkhaleq R, Niktabe A, et al. Automated large vessel occlusion detection software and thrombectomy treatment times: a cluster randomized clinical trial. JAMA Neurol. Nov 01, 2023;80(11):1182-1190. [FREE Full text] [doi: 10.1001/jamaneurol.2023.3206] [Medline: 37721738]
46. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. Apr 16, 2024;385:e078378. [FREE Full text] [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
47. Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. Oct 2025;31(10):3283-3289. [doi: 10.1038/s41591-025-03953-8] [Medline: 40954311]
48. XinL1an/LVO-DTA-NMA. GitHub. URL: https://github.com/XinL1an/LVO-DTA-NMA [accessed 2026-09-21]

Abbreviations

AIS: acute ischemic stroke
CTA: computed tomography angiography
DIC: deviance information criterion
DOR: diagnostic odds ratio
DSA: digital subtraction angiography
DTA-NMA: diagnostic test accuracy network meta-analysis
EVT: endovascular thrombectomy
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
LVO: large vessel occlusion
MCMC: Markov chain Monte Carlo
MRA: magnetic resonance angiography
NCCT: noncontrast computed tomography
NIHSS: National Institutes of Health Stroke Scale
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-DTA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for diagnostic test accuracy
PRISMA-NMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for network meta-analyses

PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
PROBAST+AI: Prediction Model Risk of Bias Assessment Tool for Artificial Intelligence
QUADAS-3: Quality Assessment of Diagnostic Accuracy Studies–3
rDOR: relative diagnostic odds ratio
STRATIS: Systematic Evaluation of Patients Treated With Neurothrombectomy Devices for Acute Ischemic Stroke
SUCRA: surface under the cumulative ranking curve

Edited by M Balcarras; submitted 19.Jun.2026; peer-reviewed by T Yusuff, NMI Sihombing; comments to author 19.Aug.2026; revised version received 22.Aug.2026; accepted 09.Sep.2026; published 24.Sep.2026

Please cite as:
Zhou H, Liu L, Fu B, Cao B, Ma J, Yang K, Bao Y, Yang F, Lian X
AI Decision Support for Emergency Triage of Large Vessel Occlusion Using Noncontrast Computed Tomography: Systematic Review and Bayesian Diagnostic Test Accuracy Network Meta-Analysis
J Med Internet Res 2026;28:e105068
URL: https://www.jmir.org/2026/1/e105068
doi: 10.2196/105068
PMID: