



Original Research

Agreement of Cercare artificial intelligence with neuroradiologist interpretation and Vitrea CT perfusion outputs in acute ischaemic stroke: A dual-centre retrospective study

Kausik Chatterjee^{a,b,c,f}, Allam Harfoush^{b,c,e,*}, Saranya Kalyanasundaram Lakshmi^c, Kumar Pulapula^c, Anastasia Liaretidou^d, Parthasarathy Karunakaran^c, Ciprian Rusu^d, Patricia O'keefe^c, Fathalla Elnagi^d

^a School of Medicine & Dentistry, University of Lancashire, Preston, United Kingdom

^b The University of Chester, Chester, United Kingdom

^c Countess of Chester Hospital, Chester, United Kingdom

^d St Helens and Knowsley Teaching Hospitals NHS Trust, St Helens and Knowsley, United Kingdom

^e North Bristol Hospital, Bristol, United Kingdom

^f Aintree Hospital, Liverpool, United Kingdom



ARTICLE INFO

Keywords:

Acute ischaemic stroke
Artificial intelligence
Computed tomography angiography
Computed tomography perfusion
Large vessel occlusion
Cerebral metabolic rate of oxygen

ABSTRACT

Purpose: To assess agreement between Cercare artificial intelligence (AI) and expert neuroradiologist interpretation for CT angiography (CTA) large vessel occlusion (LVO) detection, and to compare Cercare and Vitrea CT perfusion (CTP) outputs in acute ischaemic stroke.

Methods: This dual-centre retrospective study included 214 adults undergoing multimodal CT for suspected acute ischaemic stroke. LVO agreement was assessed in 162 CTA datasets using consultant neuroradiologist interpretation as the comparator; no angiographic or consensus reference standard was used. The quality-controlled primary CTP analysis included 196 paired datasets after exclusion of five studies with documented technical compromise; an all-output sensitivity analysis included 201 datasets. Treatment classification was assessed in 65 quality-controlled neuroradiologist-confirmed LVO cases with paired perfusion outputs, with an all-output sensitivity analysis in 67 cases. Cercare-derived cerebral metabolic rate of oxygen (CMRO₂) volumes were examined in exploratory, hypothesis-generating analyses against perfusion core volumes and discharge modified Rankin Scale (mRS).

Results: Cercare showed positive percent agreement of 82.3% (95% CI 71.0–89.8) and negative percent agreement of 83.0% (74.5–89.1) with neuroradiologist interpretation, with substantial agreement (kappa = 0.641). In the quality-controlled CTP cohort, core limits of agreement were – 34.7 to 30.1 mL (ICC 0.930) and penumbra limits were – 94.6 to 104.7 mL (ICC 0.812). Inclusion of five technically compromised studies widened the corresponding all-output limits to – 90.5 to 86.4 mL and – 191.4 to 219.5 mL. In 65 quality-controlled LVO patients, agreement was 93.8% for EXTEND-IA-style imaging mismatch and 87.7% for DAWN-style clinical-core mismatch classification. In exploratory, hypothesis-generating analyses, CMRO₂ correlated with Vitrea core ($\rho = 0.758–0.787$) and discharge mRS ($\rho = 0.390–0.401$).

Conclusion: Cercare showed substantial agreement with expert CTA interpretation. Quality control materially influenced CTP agreement estimates, and residual individual-patient variability, particularly for penumbra and treatment-threshold classification, indicates that outputs should not be regarded as interchangeable. The DAWN-derived analysis represents a threshold classification rather than full trial eligibility. CMRO₂ findings are exploratory and hypothesis-generating and require prospective validation.

* Corresponding author at: North Bristol Hospital, Bristol, United Kingdom.

E-mail address: a.harfoush@chester.ac.uk (A. Harfoush).

<https://doi.org/10.1016/j.jocn.2026.112299>

Received 9 June 2026; Accepted 8 September 2026

0967-5868/Crown Copyright © 2026 Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Acute ischaemic stroke remains a major cause of death and disability, and treatment selection depends on rapid neuroimaging assessment [1]. Multimodal CT, including non-contrast CT, CT angiography (CTA) and CT perfusion (CTP), is widely used to identify large vessel occlusion (LVO), estimate infarct core and assess potentially salvageable tissue. These tasks are time-critical and may be challenging where specialist neuroradiology support is not continuously available [2–4].

AI-based stroke imaging platforms aim to accelerate LVO detection and standardise perfusion quantification [5–8]. However, inter-software variability remains clinically important, particularly when CTP-derived core and penumbra volumes are used for late-window treatment selection. Differences in post-processing algorithms, arterial input function selection, delay correction and threshold definitions may change estimated volumes and affect eligibility near decision thresholds [9–13].

This variability is particularly relevant in pragmatic stroke pathways, where automated outputs may be reviewed rapidly by clinicians and incorporated into transfer, thrombolysis or thrombectomy discussions. A platform may show good cohort-level correlation with another package yet still generate clinically important differences for individual patients close to decision thresholds. Perfusion-based imaging selection was central to major late-window thrombectomy trials [14–16].

Cercare AI is a commercial platform for automated multimodal CT analysis. In addition to LVO detection and CTP quantification, it provides cerebral metabolic rate of oxygen (CMRO₂) estimates intended to reflect tissue viability. The aims of this study were: (1) to assess agreement between Cercare and a consultant neuroradiologist for CTA LVO detection, framed as inter-reader agreement rather than diagnostic accuracy against a definitive reference standard; (2) to compare Cercare and Vitrea perfusion outputs using routine clinical outputs, framed as a pragmatic workflow comparison rather than algorithmic harmonisation; (3) to assess agreement for EXTEND-IA-style imaging mismatch and DAWN-style clinical-core mismatch classifications; and (4) to explore CMRO₂ associations with perfusion-derived core and discharge functional status.

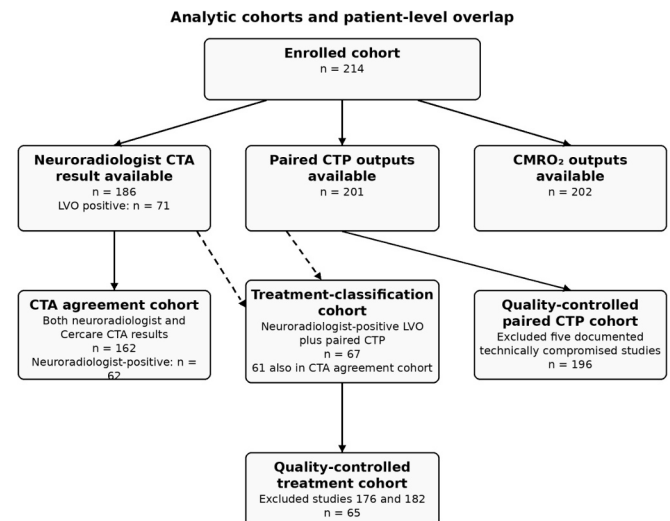
2. Materials and methods

2.1. Study design and population

This retrospective dual-centre study included consecutive adults with suspected acute ischaemic stroke who underwent multimodal CT imaging at two UK comprehensive stroke centres between January 2021 and December 2023. Institutional approval was obtained and the requirement for informed consent was waived because of the retrospective design.

Reporting was structured with reference to STARD-AI and CLAIM 2024 guidance for AI-centred diagnostic and medical imaging studies where applicable [17,18]. Because this was a retrospective agreement study using routine clinical outputs rather than model development or retraining, items relating to algorithm training data, source code and calibration were not applicable or were not available from the commercial vendor.

Inclusion criteria were clinical suspicion of acute ischaemic stroke, multimodal CT including CTA and CTP, age ≥ 18 years and imaging suitable for analysis. Exclusion criteria at cohort assembly were intracranial haemorrhage, failure to generate an interpretable study output or incomplete clinical data. A total of 214 patients were included. No formal sample-size calculation was performed; the sample reflects all eligible patients meeting inclusion criteria over the study period. During the reviewer-requested audit, five studies with generated outputs but documented technical compromise were excluded from the quality-controlled primary perfusion analysis and retained in an all-output sensitivity analysis. The analytic flow and the overlap between cohorts are shown in Fig. 1 and Supplementary Table S9.



Solid arrows show direct cohort derivation; dashed arrows show an intersection of partially overlapping cohorts.

Fig. 1. Analytic flow diagram showing the partially overlapping cohorts. Seventy-one patients were neuroradiologist-positive for LVO in the master cohort; 62 contributed to the CTA agreement analysis and 67 had paired CTP outputs for treatment classification. Five documented technically compromised CTP studies were excluded from the quality-controlled perfusion analysis, including two treatment-classification cases.

2.2. Image acquisition and processing

Patients underwent standard clinical multimodal CT protocols. At Whiston Hospital, imaging was performed on a GE Revolution EVO scanner. At the Countess of Chester Hospital, imaging was performed on a Toshiba Aquilion One scanner.

CTP acquisition parameters included coverage of two slab positions (basal ganglia and centrum semiovale), temporal sampling of 20 time points at 1.5-second intervals, acquisition duration of 30 s, contrast volume of 50 mL iodinated contrast (300–350 mgI/mL), injection rate of 5 mL/s, and reconstruction with 5 mm slice thickness using a standard soft-tissue kernel. CTA was acquired from aortic arch to vertex following a bolus-tracking technique with 60–80 mL contrast at 4–5 mL/s.

Datasets were anonymised using institutional de-identification protocols before analysis. Cercare processing (version 16) was performed retrospectively off-site after image transfer by a Cercare Medical ApS employee. The employee did not have access to clinical data or other assessments, including the neuroradiologist's interpretation. Cercare processing was fully automated without manual adjustment. Vitrea processing (version 7.11; Vital Images, USA) was performed on-site using routine clinical workflow. In 41 of 214 cases (19.2%), the initial Vitrea arterial input function and/or venous output function curve was considered unacceptable and was manually adjusted. A centre-specific breakdown of these manual adjustments was not available. Both platforms processed identical raw CTP source data.

Vitrea output was available in 213 patients and Cercare output in 202; complete paired core and penumbra outputs were available in 201 patients. Existing case-review notes identified five of these studies as technically compromised: three involved failure of Vitrea venous output function construction and two involved motion artefact that could not be corrected in either platform. These five studies, all from the Toshiba centre, were excluded from the quality-controlled primary CTP analysis (n = 196) and retained in an all-output sensitivity analysis (n = 201). CMRO₂ data were available in 202 patients. Neuroradiologist CTA results were available in 186 patients; both neuroradiologist and Cercare CTA results were available in 162 patients for the CTA agreement analysis. Categorisation relied on existing case-review notes; no new blinded source-image quality audit was undertaken.

2.3. LVO and perfusion assessment

CTA images were analysed by Cercare AI and by an experienced consultant neuroradiologist blinded to AI output and to clinical information other than suspected acute stroke. The neuroradiologist interpretation was used as the comparator for the agreement analysis. LVO was defined as complete occlusion or high-grade stenosis (>90%) involving the internal carotid artery, M1, M2 or basilar artery. High-grade stenosis was defined visually as near-complete luminal narrowing with a thread-like residual channel or contrast trickle. High-grade stenosis was included in the operational definition used for the original analysis; because stricter definitions may be limited to complete occlusion, complete occlusions were also examined separately in a sensitivity analysis.

For perfusion analysis, the two platforms used different core definitions matching their clinical default settings. Vitrea defined ischaemic core using relative cerebral blood volume (rCBV) <40% of the contralateral hemisphere, whereas Cercare defined ischaemic core using relative cerebral blood flow (rCBF) <30%. For Cercare, penumbra was defined as $T_{max} > 6$ s; for Vitrea, penumbra was defined as $T_{max} > 2.2$ s minus core volume. These different definitions were retained because they represent the outputs available in routine clinical workflow; the comparison therefore evaluates real-world interchangeability rather than algorithmic harmonisation.

Cercare estimates CMRO₂ from CTP-derived parameters using a model relating cerebral blood flow, oxygen extraction fraction and arterial oxygen content. The computation depends on deconvolution assumptions and perfusion modelling steps shared with the core and penumbra pipeline. The thresholds used (<0.20, <0.26, <0.32, <0.38) are vendor-specified defaults without prior external validation; the analysis is therefore exploratory and hypothesis-generating. Because CMRO₂ and Cercare core are generated by the same platform and may share processing steps, within-platform correlations were interpreted as internal consistency metrics; cross-platform correlations between CMRO₂ and Vitrea core were considered more independent.

2.4. Treatment-classification analysis

Treatment-classification analyses included 67 of the 71 neuroradiologist-positive LVO patients in the overall cohort who had complete paired Vitrea and Cercare perfusion outputs. This cohort was not restricted to patients with an available Cercare CTA result and therefore was not a subset of the 162-patient CTA agreement cohort. Sixty-one of these patients also belonged to the CTA agreement cohort, whereas six lacked a Cercare CTA result. Studies 176 and 182 had documented technical compromise and were excluded from the quality-controlled primary treatment analysis, leaving 65 patients; all 67 patients were retained in a sensitivity analysis. NIHSS was ascertained from admission clinical records and age was calculated from date of birth to scan date. Premorbid mRS was not consistently available, and the complete DAWN time-window, occlusion-location and premorbid-function criteria were not applied. The DAWN-derived result therefore represents clinical-core mismatch threshold classification rather than full DAWN trial eligibility.

EXTEND-IA-style imaging mismatch classification used core <70 mL, penumbra >10 mL and mismatch ratio >1.2 [14]. Mismatch ratio was defined as (core + penumbra)/core. When core was 0 mL and penumbra was >0 mL, the ratio was treated as infinite and classified above each tested ratio threshold; when both core and penumbra were 0 mL, the ratio was treated as undefined and classified as no mismatch for threshold tabulation. DAWN-style clinical-core mismatch classification applied only the age-, NIHSS- and core-volume thresholds derived from the original trial: core <21 mL for age ≥80 years with NIHSS ≥10, core <31 mL for age <80 years with NIHSS 10–19, and core <51 mL for age <80 years with NIHSS ≥20 [15].

2.5. Clinical data and statistical analysis

Clinical variables included age, sex, admission National Institutes of Health Stroke Scale (NIHSS), treatment with intravenous thrombolysis or endovascular thrombectomy, final diagnosis and discharge modified Rankin Scale (mRS). Denominators are reported where data were missing.

Agreement for LVO detection was summarised using positive percent agreement, negative percent agreement, overall percent agreement and Cohen's kappa with 95% confidence intervals. These metrics describe agreement with a single expert comparator rather than diagnostic accuracy against a definitive angiographic or consensus reference standard.

Paired CTP volumes were compared using paired t-tests for mean differences and Wilcoxon signed-rank tests as sensitivity analyses. Median and interquartile range (IQR) values are reported alongside means. Pearson and Spearman correlations, two-way mixed-effects absolute-agreement intraclass correlation coefficients (ICC) and Bland-Altman limits of agreement were calculated. Proportional bias was assessed by regressing the difference on the mean of paired measurements. The quality-controlled cohort excluding five documented technically compromised studies was the primary perfusion analysis; all 201 paired outputs were examined as a sensitivity analysis. Centre/scanner heterogeneity was assessed descriptively and using Brown-Forsythe tests for differences in variance, Welch tests for mean difference and Mann-Whitney tests for distributional differences. Scanner and centre effects could not be separated because each scanner was used at only one centre. Sensitivity analyses used $\log_{10}(\text{volume} + 1)$ to accommodate zero-valued measurements. Bland-Altman differences on the transformed scale were calculated as $\log_{10}(\text{Cercare volume} + 1) - \log_{10}(\text{Vitrea volume} + 1)$; the mean difference and 95% limits of agreement were then back-transformed and reported as ratios of (Cercare volume + 1) to (Vitrea volume + 1), rather than as simple raw-volume ratios.

For CMRO₂, Spearman correlation was the primary measure because threshold volumes were non-normally distributed; Pearson r is reported only descriptively where relevant. Associations with discharge mRS were assessed using Spearman correlation. The CMRO₂ analyses were exploratory and hypothesis-generating; no primary CMRO₂ endpoint was prespecified and no adjustment for multiple comparisons was applied. Reported CMRO₂ p values are therefore unadjusted and should be interpreted descriptively. Kappa and ICC confidence intervals were estimated by bootstrap resampling (1000 iterations).

Statistical analyses were performed using R version 4.3.2 (R Foundation for Statistical Computing, Vienna, Austria) with packages *irr*, *BlandAltmanLeh* and *boot*. Treatment classifications and threshold cross-tabulations were regenerated from the raw age, NIHSS, core and penumbra variables in the locked dataset rather than from legacy derived flags. Two-tailed $p < 0.05$ was considered significant.

3. Results

3.1. Patient characteristics

Baseline characteristics are summarised in Table 1. Mean age was 70.8 ± 13.5 years and 116/214 patients (54.2%) were male. Median admission NIHSS was 11 (IQR 6–18; $n = 207$), and median onset-to-scan time was 4.4 h (IQR 2.2–10.4; $n = 212$). Final diagnosis was ischaemic stroke in 179/213 patients (84.0%) and stroke mimic in 34/213 (16.0%).

Intravenous thrombolysis was recorded in 95/183 patients (51.9%) and endovascular thrombectomy in 36/173 (20.8%). Discharge mRS was available in 177 patients; 80/177 (45.2%) had mRS 0–2 and 106/177 (59.9%) had mRS 0–3.

Table 1
Baseline patient characteristics (n = 214).

Characteristic	Value
Age, years, mean ± SD	70.8 ± 13.5 (n = 214)
Male sex	116/214 (54.2%)
Scanner make	GE 108/214 (50.5%); Toshiba 106/214 (49.5%)
Wake-up stroke	62/202 (30.7%)
Onset-to-scan time, hours, median (IQR)	4.4 (2.2–10.4), n = 212
Scan within 6 h	131/214 (61.2%)
Admission NIHSS, median (IQR)	11 (6–18), n = 207
Final diagnosis of ischaemic stroke	179/213 (84.0%)
Received intravenous thrombolysis	95/183 (51.9%)
Received endovascular thrombectomy	36/173 (20.8%)
Discharge mRS 0–2	80/177 (45.2%)
Discharge mRS 0–3	106/177 (59.9%)

NIHSS, National Institutes of Health Stroke Scale; IQR, interquartile range; mRS, modified Rankin Scale; SD, standard deviation. Denominators reflect available data.

3.2. LVO detection

In 162 CTA datasets, the neuroradiologist identified LVO in 62 (38.3%) and Cercare in 68 (42.0%). Cercare agreed with 51/62 neuroradiologist-positive cases and 83/100 neuroradiologist-negative cases, giving positive percent agreement of 82.3% (95% CI 71.0–89.8), negative percent agreement of 83.0% (74.5–89.1), overall agreement of 82.7% and substantial agreement (kappa = 0.641; 95% CI 0.510–0.755; Table 2).

Discordant cases comprised 11 false-negative and 17 false-positive Cercare outputs relative to neuroradiologist interpretation. A descriptive analysis of discordant cases by vessel segment and centre/scanner is provided in Supplementary Table S1. Among neuroradiologist-positive LVO cases, occlusion location was M1 in 44, internal carotid artery in 10 and M2 in 8.

In a sensitivity analysis restricted to complete occlusions (excluding 7 cases classified as high-grade stenosis rather than complete occlusion), positive percent agreement was 83.6% (46/55), negative percent agreement remained 83.0% (83/100), and kappa was 0.654 (Supplementary Table S2).

3.3. Perfusion agreement

The quality-controlled primary CTP analysis included 196 patients after excluding five studies with documented technical compromise. Mean core volumes were 22.1 ± 45.9 mL for Vitrea and 19.7 ± 43.3 mL for Cercare (paired *t*-test *p* = 0.051; Wilcoxon *p* = 0.917). Median core volumes were 3.88 mL (IQR 0.08–27.45) and 4.70 mL (IQR 0.01–21.62),

Table 2
Agreement of Cercare AI with expert neuroradiologist for LVO detection (n = 162).

Metric	Value (95% CI)
Positive percent agreement	82.3% (71.0–89.8)
Negative percent agreement	83.0% (74.5–89.1)
Overall agreement	82.7% (76.2–87.8)
Cohen's kappa	0.641 (0.510–0.755)
Cross-tabulation	Neuroradiologist positive/Cercare positive = 51; neuroradiologist negative/Cercare positive = 17; neuroradiologist positive/Cercare negative = 11; neuroradiologist negative/Cercare negative = 83

CI, confidence interval; LVO, large vessel occlusion. Metrics reflect agreement with a single consultant neuroradiologist comparator rather than diagnostic accuracy against angiographic confirmation.

respectively. Correlation was strong by Pearson analysis (*r* = 0.933) and Spearman analysis remained *ρ* = 0.811. Absolute agreement was high (ICC = 0.930; bootstrap 95% CI 0.818–0.967). The mean difference was –2.3 mL (Cercare minus Vitrea), with limits of agreement from –34.7 to 30.1 mL (Figs. 2 and 3). Proportional bias was present (*p* = 0.029).

Mean penumbra volumes were 59.6 ± 73.8 mL for Vitrea and 64.6 ± 91.6 mL for Cercare (paired *t*-test *p* = 0.168; Wilcoxon *p* = 0.067); medians were 32.73 mL (IQR 3.74–95.07) and 40.10 mL (IQR 4.73–88.76), respectively. Pearson correlation was *r* = 0.832, Spearman correlation was *ρ* = 0.806 and ICC was 0.812 (bootstrap 95% CI 0.748–0.881). The mean difference was 5.0 mL and limits of agreement were –94.6 to 104.7 mL. Proportional bias remained present (*p* < 0.001). In the all-output sensitivity analysis (n = 201), inclusion of the five technically compromised studies widened the limits of agreement to –90.5 to 86.4 mL for core and –191.4 to 219.5 mL for penumbra, with ICCs of 0.665 and 0.610, respectively (Supplementary Table S4).

Threshold-based reclassification in the quality-controlled cohort showed discordance in 17/196 patients (8.7%) at core ≥50 mL, 5/196 (2.6%) at core ≥70 mL and 5/196 (2.6%) at core ≥100 mL. Discordance occurred in 38/196 patients (19.4%) at each mismatch-ratio threshold of >1.2, >1.4 and >1.8, and in 25/196 (12.8%) at absolute mismatch volume >10 mL (Table 3). Independent recalculation confirmed the identical ratio-threshold counts because every finite ratio exceeded 2.00; differences were driven by platform-specific zero-core values. Separate 2 × 2 tables and ratio distributions are provided in Supplementary Table S5A.

A log₁₀(volume + 1) sensitivity analysis used transformed differences defined as log₁₀(Cercare volume + 1) – log₁₀(Vitrea volume + 1). The back-transformed mean difference and 95% limits of agreement, reported as ratios of (Cercare volume + 1) to (Vitrea volume + 1), are provided in Supplementary Table S5B.

3.4. CMRO₂ analysis

In exploratory, hypothesis-generating analyses, CMRO₂ volumes showed strong cross-platform Spearman correlations with Vitrea core (*ρ* = 0.758–0.787) and, as expected for within-platform outputs, stronger correlations with Cercare core (*ρ* = 0.842–0.846; Table 4). CMRO₂ also correlated moderately with discharge mRS (*ρ* = 0.390–0.401; n = 166; Fig. 4). No primary CMRO₂ endpoint was prespecified and no multiplicity correction was applied. Table 5.

In a further exploratory, hypothesis-generating analysis of software-defined large core (≥70 mL), CMRO₂ threshold volumes showed high discrimination, with AUCs of 0.921–0.936 for Vitrea-defined large core and 0.945–0.959 for Cercare-defined large core (Fig. 5). These analyses use software-derived core volume rather than follow-up infarct volume as the reference and should therefore be interpreted as technical association rather than clinical validation.

All correlations *p* < 0.001; reported *p* values are unadjusted. The CMRO₂ analyses were exploratory and hypothesis-generating, with no prespecified primary endpoint or correction for multiple comparisons. AUC analyses used software-defined core ≥70 mL as a technical reference and are exploratory.

3.5. Treatment-classification agreement

The quality-controlled treatment-classification analysis included 65 neuroradiologist-confirmed LVO patients with complete paired perfusion outputs after exclusion of studies 176 and 182 for documented technical compromise. This cohort was drawn from all neuroradiologist-positive cases rather than from the 162-patient CTA agreement cohort.

For EXTEND-IA-style imaging mismatch, agreement was 93.8% (61/65; kappa = 0.818, bootstrap 95% CI 0.618–0.959). For DAWN-style clinical-core mismatch classification, agreement was 87.7% (57/65; kappa = 0.748, bootstrap 95% CI 0.567–0.900). Discordant treatment classifications arose through a combination of modest threshold

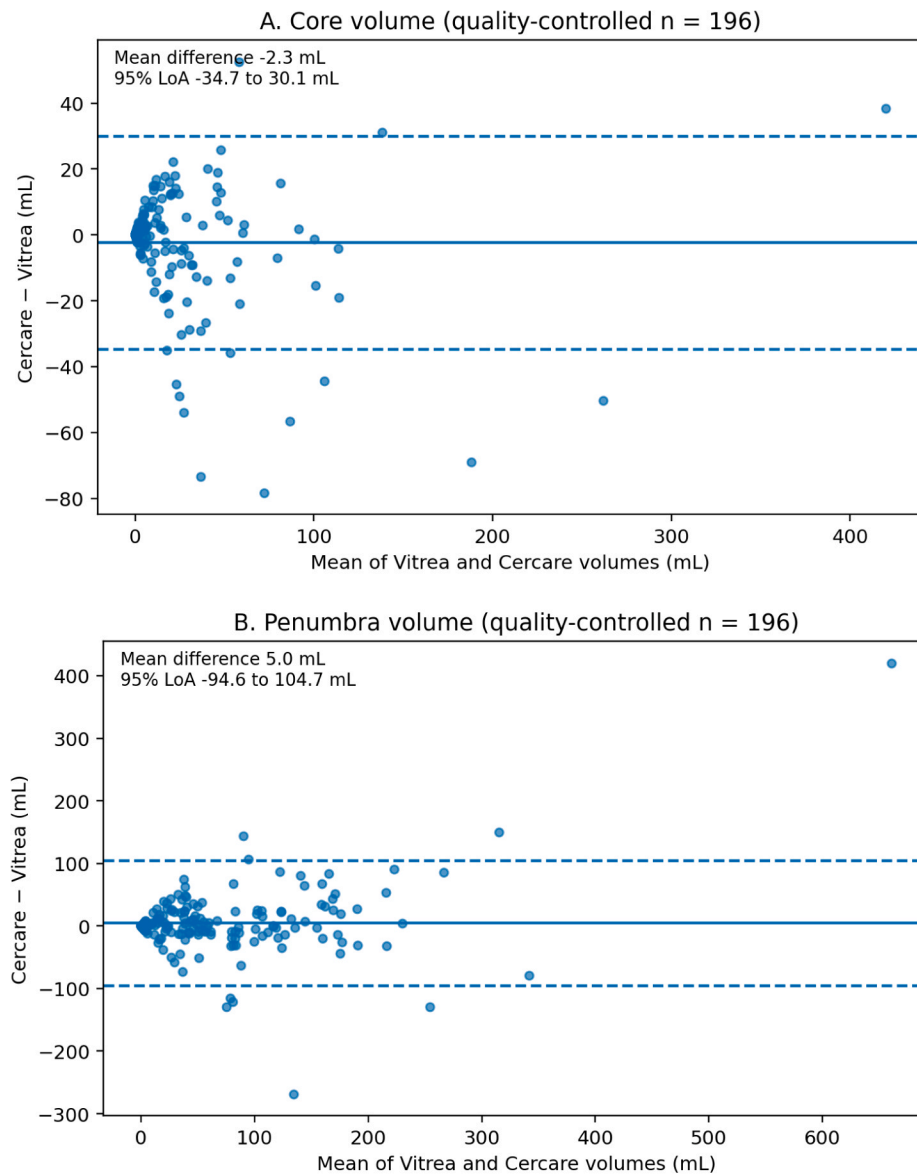


Fig. 2. Quality-controlled Bland-Altman agreement analysis for (A) core volume and (B) penumbra volume (n = 196). Differences were calculated as Cercare minus Vitrea; dashed lines show 95% limits of agreement.

crossings and substantial between-platform differences in estimated core volume. One of the largest disagreements in the all-output analysis occurred in a study with documented technical processing failure. The all-output sensitivity analysis in 67 patients yielded agreement of 92.5% and 88.1%, respectively; study 182 contributed one additional EXTEND-IA-style disagreement and had documented Vitrea VOF construction failure (Supplementary Table S3).

3.6. Centre/scanner-stratified analysis

Centre/scanner-stratified CTA analysis showed no statistically significant difference in LVO agreement between the GE centre (n = 69) and Toshiba centre (n = 93; kappa 0.553 versus 0.707; p = 0.21). In the quality-controlled CTP cohort, core limits of agreement were -25.7 to 27.5 mL at the GE centre (n = 104) and -42.7 to 30.8 mL at the Toshiba centre (n = 92); penumbra limits were -88.1 to 87.5 mL and -99.8 to 121.9 mL, respectively. Core difference variance remained heterogeneous after quality-control exclusions (Brown-Forsythe p = 0.017), whereas penumbra difference variance did not (p = 0.655). Because each scanner was used at only one centre, scanner, acquisition and local

processing effects were completely confounded (Supplementary Table S4).

4. Discussion and conclusion

4.1. Principal findings

In this dual-centre retrospective study, Cercare showed substantial agreement with neuroradiologist interpretation for CTA LVO detection. Patient-level audit identified that technical quality materially affected CTP agreement estimates: exclusion of five documented technically compromised studies substantially narrowed limits of agreement and increased ICCs. Nevertheless, residual limits remained clinically relevant, particularly for penumbra and threshold-based classification. CMRO₂ volumes showed strong cross-platform rank-based associations with Vitrea-derived core and moderate association with discharge mRS, but these findings remain exploratory and hypothesis-generating.

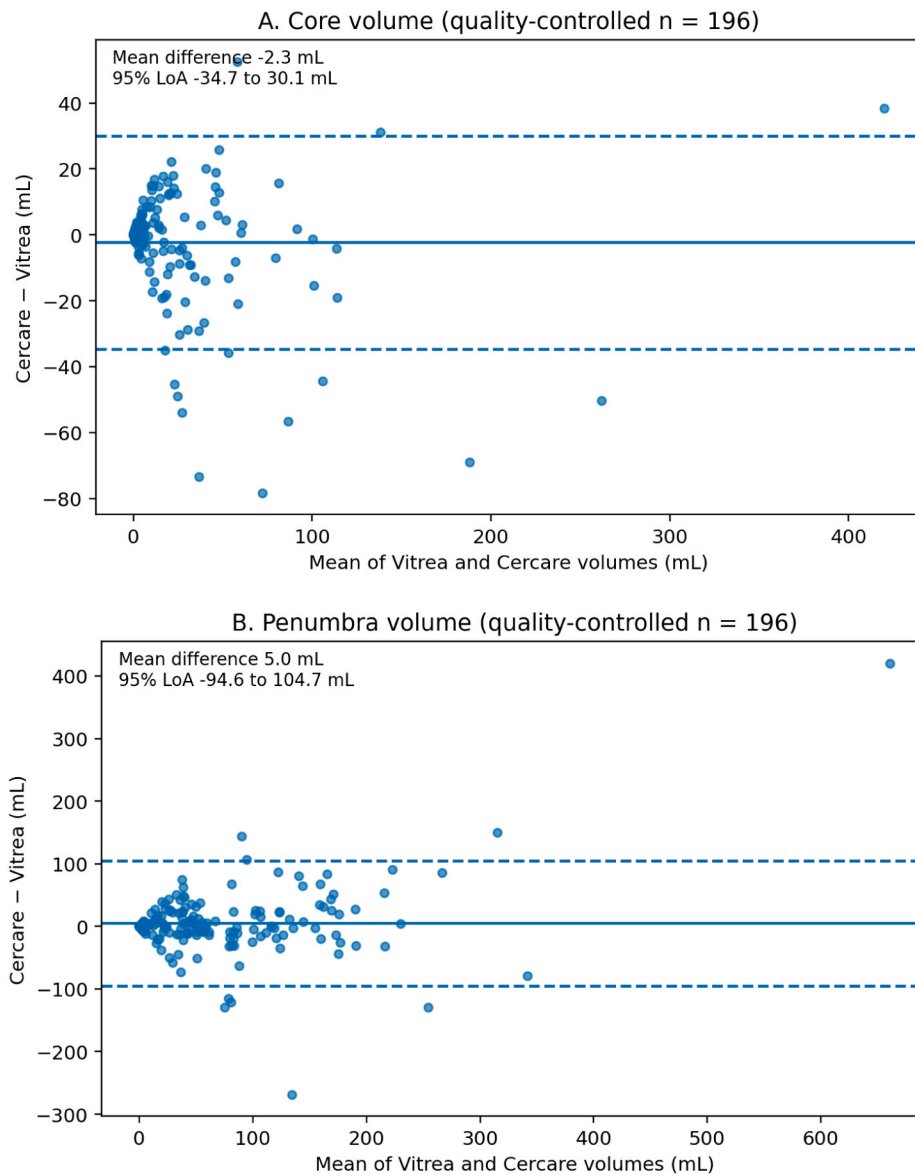


Fig. 3. Quality-controlled correlation analysis for (A) core volume and (B) penumbra volume (n = 196). The dashed line is the line of identity and the solid line is the fitted regression line.

4.2. LVO detection

Cercare showed positive percent agreement of 82.3% and negative percent agreement of 83.0% with expert neuroradiologist interpretation. These values describe agreement with a single expert reader rather than diagnostic accuracy against a definitive standard. Any potential role in prioritising image review requires prospective workflow evaluation, because this study did not assess alert time, reporting time, transfer time, treatment delay or patient-level workflow benefit.

However, the 11 false-negative and 17 false-positive cases emphasise that AI should support, not replace, expert image review. False negatives are clinically important because missed LVO may delay transfer or thrombectomy assessment, whereas false positives may trigger unnecessary escalation. A descriptive analysis of discordant cases by vessel segment and centre/scanner is provided in [Supplementary Table S1](#); formal blinded adjudication of artefact pattern, stenosis grade and collateral status was not performed.

Because the comparator was a single neuroradiologist rather than consensus interpretation or angiographic confirmation, these results

should be interpreted as agreement with one expert reader rather than definitive diagnostic accuracy. CTA interpretation also has recognised observer variability in suspected acute ischaemic stroke [19]. Future validation should include consensus or angiographic reference standards and structured adjudication of discordant cases by vessel segment, image quality and artefact pattern.

4.3. Perfusion analysis

For core and penumbra volumes, correlation coefficients suggested that both platforms broadly tracked between-patient variation. The quality-controlled analysis showed high core ICC and moderate-to-high penumbra ICC, but Bland-Altman limits remained wide enough to affect individual decisions. This distinction is important: good correlation does not establish interchangeability. The all-output analysis further demonstrated that a small number of technically compromised studies can markedly widen limits of agreement and reduce ICCs.

Because no follow-up infarct-volume reference standard was available, this study cannot determine whether either software package more

Table 3
Threshold-based reclassification analysis in the quality-controlled paired CTP cohort (n = 196).

Threshold	Discordant cases, n (%)	Direction of discordance
Core ≥50 mL	17 (8.7%)	Vitrea ≥50, Cercare <50: 8; Vitrea <50, Cercare ≥50: 9
Core ≥70 mL	5 (2.6%)	Vitrea ≥70, Cercare <70: 4; Vitrea <70, Cercare ≥70: 1
Core ≥100 mL	5 (2.6%)	Vitrea ≥100, Cercare <100: 5; Vitrea <100, Cercare ≥100: 0
Mismatch ratio >1.2	38 (19.4%)	Vitrea >1.2, Cercare ≤1.2: 17; Vitrea ≤1.2, Cercare >1.2: 21
Mismatch ratio >1.4	38 (19.4%)	Vitrea >1.4, Cercare ≤1.4: 17; Vitrea ≤1.4, Cercare >1.4: 21
Mismatch ratio >1.8	38 (19.4%)	Vitrea >1.8, Cercare ≤1.8: 17; Vitrea ≤1.8, Cercare >1.8: 21
Absolute mismatch volume >10 mL	25 (12.8%)	Vitrea >10, Cercare ≤10: 10; Vitrea ≤10, Cercare >10: 15
Absolute mismatch volume >15 mL	31 (15.8%)	Vitrea >15, Cercare ≤15: 11; Vitrea ≤15, Cercare >15: 20
Absolute mismatch volume >50 mL	29 (14.8%)	Vitrea >50, Cercare ≤50: 11; Vitrea ≤50, Cercare >50: 18

accurately estimates true infarct core. The analysis evaluates agreement between software outputs, not biological accuracy.
The observed variability is likely to reflect differences in perfusion

modelling and, importantly, different core definitions: Vitrea used an rCBV <40% threshold whereas Cercare used rCBF <30%. This makes the comparison pragmatic rather than algorithmically harmonised: it evaluates whether the outputs available to clinicians in routine workflow can be substituted for each other, not whether the platforms are measuring the same biological quantity. Penumbra estimates may be even more sensitive to such differences because they depend on both Tmax-defined hypoperfusion and subtraction of estimated core. This may explain why penumbra limits of agreement were wider than those for core, and why modest proportional bias was observed for penumbra volumes.

Treatment-classification discordance arose through a mixture of modest threshold crossings and substantial between-platform core-volume differences. Examples included 31.72 versus 84.29 mL and 111.15 versus 32.77 mL in quality-controlled cases. The technically compromised all-output case with core estimates of 226.00 versus 29.63 mL contributed an additional EXTEND-IA-style disagreement and was excluded from the quality-controlled primary analysis. Threshold reclassification remained evident after quality control, including 2.6% discordance at the core ≥70 mL threshold and 19.4% at mismatch-ratio thresholds. These findings support reviewing automated maps with source imaging and clinical context, particularly when platform outputs differ materially. The five documented compromised cases and the basis for their exclusion are detailed in [Supplementary Table S6](#).

Table 4
CMRO₂ correlations with perfusion-derived core volumes and discharge mRS.

CMRO ₂ threshold	Vitrea core ρ (95% CI)	Cercare core ρ (95% CI)	Discharge mRS ρ (95% CI)	AUC: Vitrea core ≥70 mL	AUC: Cercare core ≥70 mL
CMRO ₂ <0.20	0.758 (0.692–0.811)	0.845 (0.801–0.881)	0.390 (0.253–0.512)	0.921	0.945
CMRO ₂ <0.26	0.771 (0.708–0.821)	0.846 (0.801–0.881)	0.396 (0.259–0.517)	0.930	0.957
CMRO ₂ <0.32	0.782 (0.722–0.831)	0.843 (0.798–0.879)	0.401 (0.265–0.521)	0.933	0.955
CMRO ₂ <0.38	0.787 (0.728–0.835)	0.842 (0.796–0.878)	0.397 (0.261–0.518)	0.936	0.959

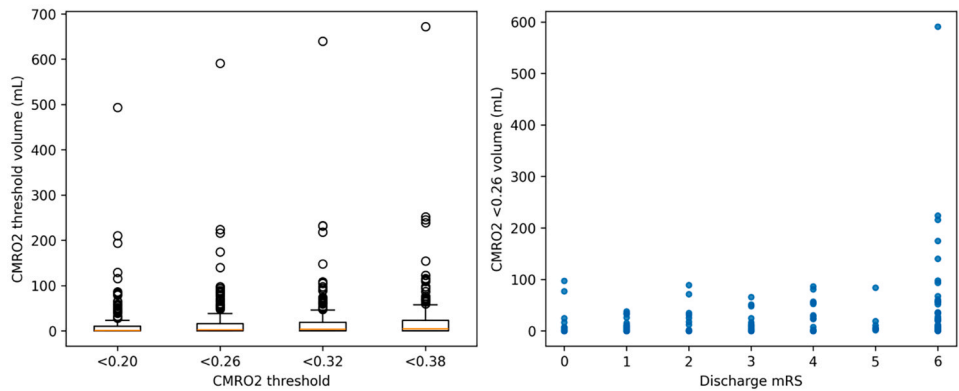


Fig. 4. Exploratory, hypothesis-generating CMRO₂ analysis. (A) Distribution of CMRO₂ threshold volumes. (B) Association between CMRO₂ <0.26 vol and discharge mRS.

Table 5
Agreement for treatment classification in the quality-controlled cohort (n = 65).

Criterion	Software	Meets classification, n (%)	Agreement, n (%)	Kappa (95% CI)	Cross-tabulation*
EXTEND-IA-style imaging mismatch	Vitrea	50/65 (76.9%)	61/65 (93.8%)	0.818 (0.618–0.959)	No/No = 12; No/Yes = 3; Yes/No = 1; Yes/Yes = 49
	Cercare	52/65 (80.0%)			
DAWN-style clinical-core mismatch	Vitrea	29/65 (44.6%)	57/65 (87.7%)	0.748 (0.567–0.900)	No/No = 34; No/Yes = 2; Yes/No = 6; Yes/Yes = 23
	Cercare	25/65 (38.5%)			

* Cross-tabulation cells are Vitrea/Cercare classification counts. Analysis includes neuroradiologist-confirmed LVO patients with complete paired perfusion outputs after exclusion of two documented technically compromised studies. DAWN-style classification applied age-, NIHSS- and core-volume thresholds only and does not represent full DAWN trial eligibility. CI, confidence interval.

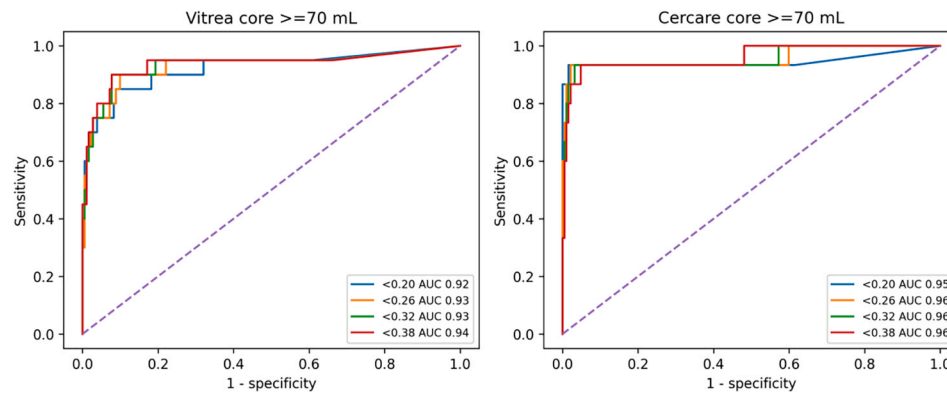


Fig. 5. Exploratory, hypothesis-generating ROC curves for CMRO₂ threshold volumes predicting software-defined large core volume (≥ 70 mL) using Vitrea-derived and Cercare-derived core volumes as technical references.

4.4. CMRO₂

CMRO₂ threshold volumes correlated with both Cercare and Vitrea core volumes and with discharge mRS, providing hypothesis-generating evidence of biological plausibility as a marker of metabolically compromised tissue. This is plausible because oxygen metabolism is closely related to irreversible tissue injury, and recent CT perfusion work has evaluated CMRO₂-derived thresholds against DWI infarct core [20]. The within-platform correlations are expected, as both measures share the Cercare processing pipeline, and they support internal consistency rather than independent validation. The more informative finding is the cross-platform correlation with Vitrea core, which provides preliminary evidence that CMRO₂ tracks a signal detectable across different perfusion platforms.

The association with discharge mRS is encouraging but should be interpreted cautiously. Discharge outcome is influenced by baseline stroke severity, reperfusion treatment, comorbidity, complications and rehabilitation destination, and is not equivalent to the conventional 90-day functional endpoint. The CMRO₂ analysis should therefore be considered exploratory and hypothesis-generating rather than confirmatory. Future studies should compare CMRO₂ with follow-up infarct volume, reperfusion status and longer-term functional outcome before clinical thresholds are adopted.

4.5. Clinical Implications and limitations

The present study assessed agreement between imaging outputs and did not measure alert time, reporting time, transfer time, treatment delay or patient outcome attributable to software use. The findings should therefore not be interpreted as evidence of workflow benefit or clinical utility. Quality control should precede interpretation of automated maps, and software-specific outputs should be reviewed together with source imaging, clinical severity, time from onset and treatment feasibility [21,22]. Local validation remains important when acquisition and processing workflows differ.

This study evaluates agreement rather than diagnostic accuracy. Limitations include the retrospective design, use of a single neuroradiologist comparator without consensus adjudication or angiographic confirmation, incomplete data for some variables, differing sample sizes across analyses and absence of validation against final infarct volume. [Supplementary Table S1](#) provides only a descriptive breakdown of discordant LVO cases by recorded vessel segment and centre/scanner. Image quality, artefact pattern, CTA phase or timing, collateral status, stenosis grade, follow-up imaging and treatment confirmation were not systematically reassessed. The technical-quality analysis relied on existing case-review notes rather than a new blinded source-image adjudication. Although the initial Vitrea AIF and/or VOF curve was manually adjusted in 41/214 cases (19.2%), a centre-specific

breakdown and detailed adjustment logs were not available. Off-site Cercare processing was performed by a Cercare Medical ApS employee whose role was limited to automated image processing and who had no access to clinical data or the neuroradiologist's interpretation. For study 182, the stored Cercare penumbra value of 1414.68 mL could not be independently verified against source maps, segmentation masks, acquisition coverage or units; this technically compromised case was excluded from the quality-controlled analysis and retained only in the all-output sensitivity analysis.

The study compared two real-world software workflows rather than harmonised research pipelines. Centre and scanner were completely confounded, and the observed heterogeneity cannot be attributed independently to scanner manufacturer. Both CTP protocols used two-slab coverage of the basal ganglia and centrum semiovale; limited z-axis coverage may have underestimated infarcts extending beyond these regions.

Confidence intervals for ICCs are wide, and multiple exploratory CMRO₂ analyses were performed without correction. No primary endpoint was prespecified for the CMRO₂ analysis. Discharge mRS was used rather than 90-day mRS, and the CMRO₂ findings are exploratory and hypothesis-generating rather than confirmatory. No formal fairness or demographic subgroup analysis was undertaken. Prospective studies should include consensus or angiographic LVO standards, follow-up imaging, longer-term outcomes and assessment of workflow impact.

4.6. Conclusion

Cercare AI showed substantial agreement with expert neuroradiologist interpretation for LVO detection, noting that this reflects agreement with a single reader rather than diagnostic accuracy against angiographic confirmation. Quality control materially improved CTP agreement estimates, but residual individual-patient variability and threshold reclassification indicate that Cercare and Vitrea outputs should not be treated as interchangeable. The DAWN-derived analysis represents age-, NIHSS- and core-volume threshold classification rather than full DAWN trial eligibility. CMRO₂ showed exploratory, hypothesis-generating cross-platform associations with Vitrea-derived core volume and discharge mRS, but prospective validation against follow-up infarct imaging, reperfusion status and 90-day outcome is required.

CRedit authorship contribution statement

Kausik Chatterjee: . **Allam Harfoush:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Saranya Kalyanasundaram Lakshmi:** Writing – review & editing, Data curation. **Kumar Pulapula:** Writing – review & editing, Formal analysis. **Anastasia**

Liaretidou: Writing – review & editing, Data curation. **Parthasarathy Karunakaran:** Writing – review & editing, Data curation. **Ciprian Rusu:** Writing – review & editing, Data curation. **Patricia O'keefe:** Writing – review & editing, Data curation. **Fathalla Elnagi:** Writing – review & editing, Investigation, Conceptualization.

Ethics declaration

The authors have NOT obtained informed consent from participants or their legal representatives. The study was conducted in accordance with local institutional guidelines and ethical approval was obtained (AMaT Ref: Stro/QI/2026-27/02). Institutional approval was obtained and the requirement for informed consent was waived because of the retrospective design.

This study was performed in compliance with relevant laws, regulatory frameworks and guidelines where the research took place. This study was approved by the NHS. (Approval No. AMaT Ref: Stro/QI/2026-27/02).

Declaration of Generative AI and AI-Assisted Technologies in the manuscript preparation Process

During the preparation of this work the author(s) used Gemini in order to improve language, formatting, and structural organisation. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Ethics approval

Institutional approval was obtained and the requirement for informed consent was waived because of the retrospective design.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request, subject to institutional review board approval and data-sharing agreements. Analysis code is available upon reasonable request. Commercial AI software code is proprietary and not publicly available.

Funding

No funding was received for this study.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Cercare Medical ApS provided access to the AI platform for study purposes, which constitutes in-kind support. A Cercare Medical ApS employee performed the off-site automated image processing but had no access to clinical data or the neuroradiologist's interpretation. The employee did not participate in study design, statistical analysis, interpretation or manuscript preparation. The authors declare no other competing interests.

Acknowledgements

The authors thank the radiology and neurology staff at both participating centres for their contributions to patient care and data collection. The authors thank the radiology and stroke teams at [Countess of Chester] and [Whiston Hospital] for their support of this study. We acknowledge Cercare Medical ApS for providing access to the

AI platform for study purposes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jocn.2026.112299>.

References

- [1] Feigin VL, Norrving B, Mensah GA. Global burden of stroke. *Circ Res* 2017;120(3): 439–48. <https://doi.org/10.1161/CIRCRESAHA.116.308413>.
- [2] Saver JL, Goyal M, Bonafe A, et al. Stent-retriever thrombectomy after intravenous t-PA versus t-PA alone in stroke. *N Engl J Med* 2015;372:2285–95. <https://doi.org/10.1056/NEJMoa1415061>.
- [3] Goyal M, Demchuk AM, Menon BK, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med* 2015;372:1019–30. <https://doi.org/10.1056/NEJMoa1414905>.
- [4] Kakkar P, Kakkar T, Patankar T, Saha S. Current approaches and advances in the imaging of stroke. *Dis Model Mech* 2021;14:dmm048785. <https://doi.org/10.1242/dmm.048785>.
- [5] Stib MT, Vasquez J, Dong MP, et al. Detecting large vessel occlusion at multiphase CT angiography by using a deep convolutional neural network. *Radiology* 2020; 297:640–9. <https://doi.org/10.1148/radiol.2020200334>.
- [6] Luijten SPR, Wolff L, Duvekot MHC, et al. Diagnostic performance of an algorithm for automated large vessel occlusion detection on CT angiography. *J Neurointerv Surg* 2022;14:794–8. <https://doi.org/10.1136/neurintsurg-2021-017842>.
- [7] Matsoukas S, Morey J, Lock G, et al. AI software detection of large vessel occlusion stroke on CT angiography: a real-world prospective diagnostic test accuracy study. *J Neurointerv Surg* 2023;15:52–6. <https://doi.org/10.1136/neurintsurg-2021-018391>.
- [8] van Leeuwen KG, Meijer FJA, Schalekamp S, et al. Cost-effectiveness of artificial intelligence aided vessel occlusion detection in acute stroke: an early health technology assessment. *Insights Imaging* 2021;12:133. <https://doi.org/10.1186/s13244-021-01077-4>.
- [9] Kamalian S, Kamalian S, Maas MB, et al. CT cerebral blood flow maps optimally correlate with admission diffusion-weighted imaging in acute stroke but thresholds vary by postprocessing platform. *Stroke* 2011;42:1923–8. <https://doi.org/10.1161/STROKEAHA.110.610618>.
- [10] Bivard A, McElduff P, Spratt N, Levi C, Parsons M. Defining the extent of irreversible brain ischemia using perfusion computed tomography. *Cerebrovasc Dis* 2011;31:238–45. <https://doi.org/10.1159/000321897>.
- [11] Koopman MS, Berkhemer OA, Geuskens RREG, et al. Comparison of three commonly used CT perfusion software packages in patients with acute ischemic stroke. *J Neurointerv Surg* 2019;11:1249–56. <https://doi.org/10.1136/neurintsurg-2019-014822>.
- [12] Bivard A, Churilov L, Ma H, et al. Does variability in automated perfusion software outputs for acute ischemic stroke matter? Reanalysis of EXTEND perfusion imaging. *CNS Neurosci Ther* 2022;28:139–44. <https://doi.org/10.1111/cns.13756>.
- [13] Alwood BT, Meyer DM, Ionita C, et al. Multicenter comparison using two AI stroke CT perfusion software packages for determining thrombectomy eligibility. *J Stroke Cerebrovasc Dis* 2024;33(7):107750. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2024.107750>.
- [14] Campbell BCV, Mitchell PJ, Kleinig TJ, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med* 2015;372:1009–18. <https://doi.org/10.1056/NEJMoa1414792>.
- [15] Nogueira RG, Jadhav AP, Haussen DC, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med* 2018;378:11–21. <https://doi.org/10.1056/NEJMoa1706442>.
- [16] Albers GW, Marks MP, Kemp S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med* 2018;378:708–18. <https://doi.org/10.1056/NEJMoa1713973>.
- [17] Sounderajah V, Guni A, Liu X, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med* 2025;31:3283–9. <https://doi.org/10.1038/s41591-025-03953-8>.
- [18] Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. *Radiol Artif Intell* 2024;6:e240300. <https://doi.org/10.1148/ryai.240300>.
- [19] Mair G, von Kummer R, Adami A, et al. Observer reliability of CT angiography in the assessment of acute ischaemic stroke: data from the Third International Stroke Trial. *Neuroradiology* 2015;57:1–9. <https://doi.org/10.1007/s00234-014-1441-0>.
- [20] Otgonbaatar C, Song H, Jung KH, et al. Quantification of infarct core volume in patients with acute ischemic stroke using cerebral metabolic rate of oxygen in CT perfusion. *AJNR Am J Neuroradiol* 2024;45:1432–40. <https://doi.org/10.3174/ajnr.A8360>.
- [21] Allmendinger AM, Tang ER, Lui YW, Spektor V. Imaging of stroke: part 1, perfusion CT—overview of imaging technique, interpretation pearls, and common pitfalls. *AJR Am J Roentgenol* 2012;198:52–62. <https://doi.org/10.2214/AJR.10.7255>.
- [22] Chung CY, Hu R, Peterson RB, Allen JW. Automated processing of head CT perfusion imaging for ischemic stroke triage: a practical guide to quality assurance and interpretation. *AJR Am J Roentgenol* 2021;217:1401–16. <https://doi.org/10.2214/AJR.21.26139>.