



OPEN ACCESS

EDITED BY

Hulin Kuang,
Central South University, China

REVIEWED BY

Amirmasoud Ahmadi,
Max Planck Institute for Biological
Intelligence, Germany
Shichao Kan,
Central South University, China

*CORRESPONDENCE

Lynnet-Samuel J. Teichmann
✉ lynnet-samuel.teichmann@charite.de

RECEIVED 28 October 2024

ACCEPTED 02 January 2025

PUBLISHED 24 January 2025

CITATION

Teichmann L-SJ, Khalil AA, Villringer K,
Fiebach JB, Huwer S, Gibson E and
Galinovic I (2025) Evaluation of Siemens
Healthineers' StrokeSegApp for automated
diffusion and perfusion lesion segmentation
in patients with ischemic stroke.
Front. Neurol. 16:1518477.
doi: 10.3389/fneur.2025.1518477

COPYRIGHT

© 2025 Teichmann, Khalil, Villringer, Fiebach,
Huwer, Gibson and Galinovic. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Evaluation of Siemens Healthineers' StrokeSegApp for automated diffusion and perfusion lesion segmentation in patients with ischemic stroke

Lynnet-Samuel J. Teichmann^{1*}, Ahmed A. Khalil¹,
Kersten Villringer¹, Jochen B. Fiebach¹, Stefan Huwer²,
Eli Gibson² and Ivana Galinovic¹

¹Center for Stroke Research Berlin, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany, ²Siemens Healthineers AG, Erlangen, Germany

Purpose: This study aimed to evaluate the performance of Siemens Healthineers' StrokeSegApp performance in automatically segmenting diffusion and perfusion lesions in patients with acute ischemic stroke and to assess its clinical utility in guiding mechanical thrombectomy decisions.

Methods: This retrospective study used MRI data of acute ischemic stroke patients from the prospective observational single-center 1000Plus study, acquired between September 2008 and June 2013 (clinicaltrials.org; NCT00715533) and manually segmented by radiologists as the ground truth. The performance of the StrokeSegApp was compared against this ground truth using the dice similarity coefficient (DSC) and Bland–Altman plots. The study also evaluated the application's ability to recommend mechanical thrombectomy based on DEFUSE 2 and 3 trial criteria.

Results: The StrokeSegApp demonstrated a mean DSC of 0.60 (95% CI: 0.57–0.63; $n = 241$) for diffusion deficit segmentation and 0.80 (95% CI: 0.76–0.85; $n = 56$) for perfusion deficit segmentation. The mean volume deviation was 0.49 mL for diffusion lesions and –7.69 mL for perfusion lesions. Out of 56 subjects meeting DEFUSE 2/3 criteria in the cohort, it correctly identified mechanical thrombectomy candidates with a sensitivity of 82.1% (95% CI: 63.1–93.9%) and a specificity of 96.4% (95% CI: 81.7–99.9%).

Conclusion: The Siemens Healthineers' StrokeSegApp provides accurate automated segmentation of ischemic stroke lesions, comparable to human experts as well as similar commercial software, and shows potential as a reliable tool in clinical decision-making for stroke treatment.

KEYWORDS

acute ischemic stroke, automated lesion segmentation, MRI analysis software, diffusion-weighted imaging, perfusion-weighted imaging, Siemens Healthineers StrokeSegApp

1 Introduction

The lifetime prevalence of stroke is estimated at 25% (1). Half of those who survive remain chronically disabled, whereby timely and accurate diagnosis and efficient therapy are crucial for the patient's outcome (2, 3). Consequently, magnetic resonance imaging (MRI) has become an integral tool in modern stroke diagnostics (4).

As an indispensable sequence, diffusion-weighted imaging (DWI) provides us with information about the areas of the brain likely to be irreversibly damaged by the infarction. The reduction in the diffusivity of water molecules, present in the situation of cytotoxic edema, unmask the infarct core, which shows up as a drop in the apparent diffusion coefficient (ADC) and an increase in the signal on trace DWI (TraceW) (5).

Perfusion-weighted imaging (PWI), on the other hand, reveals areas of hypoperfused tissue where neuronal dysfunction is potentially reversible. This tissue is potentially at risk of further damage if the insufficient blood flow is not ameliorated. Perfusion maps can be derived from changes in various parameters, such as the relative mean transit time (relMTT), the time to maximum (Tmax), or the cerebral blood flow (CBF) (6).

The juxtaposition of DWI and PWI illustrates the concept of the penumbra, which describes those areas of deficient perfusion that are not part of the irreversibly lost diffusion lesion and thus—despite being currently non-functional—represent vital tissue (7). As these penumbral areas can be saved through timely reperfusion, their quantification has found its way into current recommendations and guidelines for treating stroke patients within the extended or unclear time window (3, 8–10).

Automated image evaluation can support determining penumbra rapidly through image segmentation, whereby deep neural and multilayered convolutional networks in particular have significant potential in neuroradiology (11–14). Although many artificial intelligence (AI) systems already provide reliable results for identifying the ischemic core on MRI (15–19), there are significantly fewer reports on programs that allow for reliable, automatic perfusion segmentation validated on large cohorts (15, 20, 21). In addition, these studies tend to use different statistical approaches, which makes it difficult to compare the applications with each other (22).

The development model of a new generation AI tool from Siemens Healthineers, the research application StrokeSegApp v. 1.3, offers the determination of both lesion types with automatic mismatch calculation. This study had two aims: (a) to provide external and independent validation of the tool's segmentation results and (b) to shed light on whether the assessment of the penumbra by the program is adequate and would provide clinical benefit in decision-making for acute stroke patients qualifying for mechanical thrombectomy (MT).

2 Methods

2.1 Study design

This retrospective study used data from the 1000Plus study (clinicaltrials.org; NCT00715533) (23). The protocol received approval from the local Ethics Committee (EA4/026/08), and all study

participants gave informed consent for participation in the study. Subjects were eligible for the 1000Plus study if they were admitted with the clinical signs of an acute cerebrovascular event within the last 24 h, were able to undergo an MRI scan, and were at least 19 years old. The MRI protocol included diffusion-weighted imaging (DWI), as well as perfusion imaging (PWI), performed using dynamic susceptibility contrast imaging.

This study is reported following the STARD guidelines.

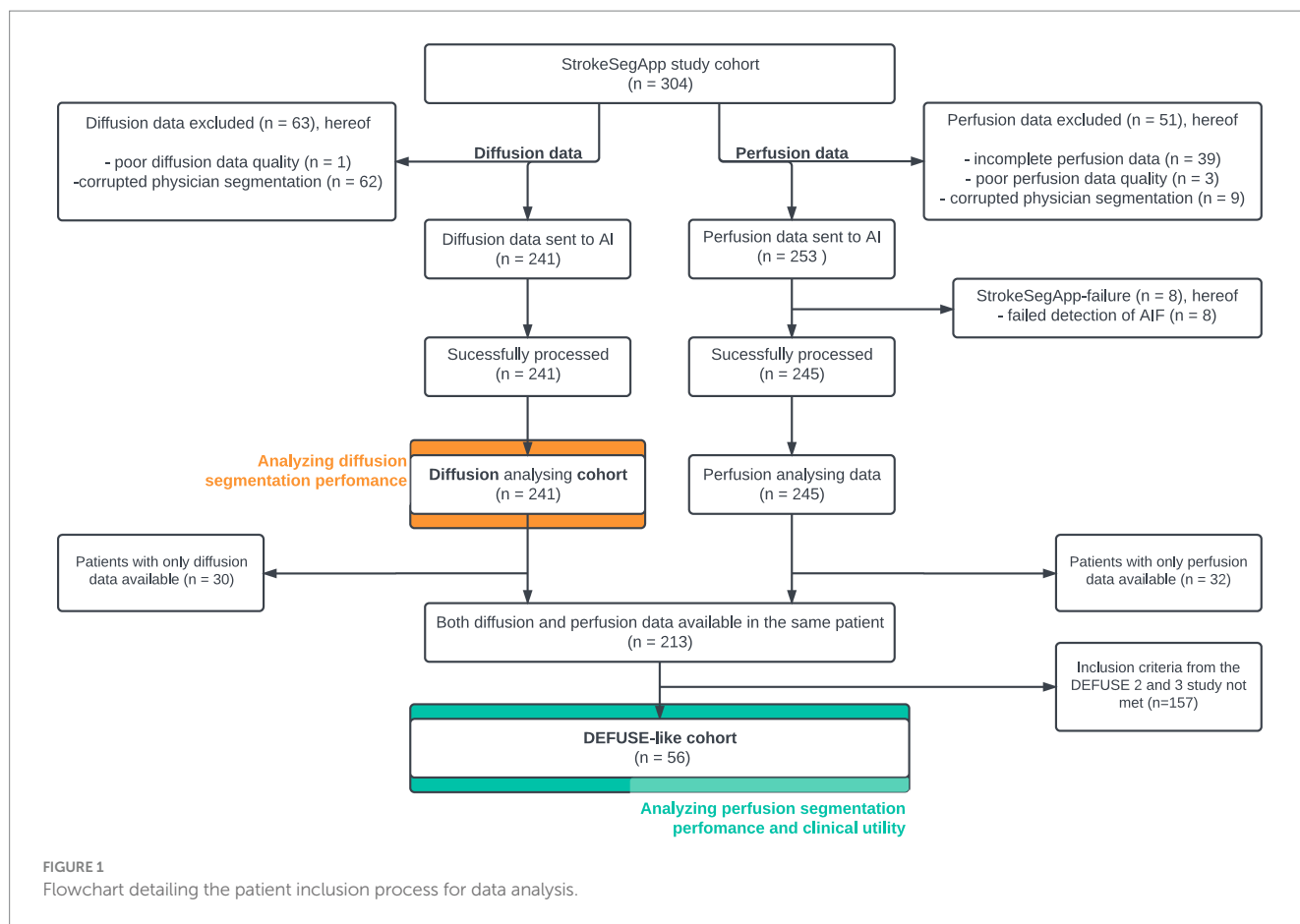
2.2 Participants

Patients were consecutively enrolled between 1 September 2008 and 30 June 2013 at Charité, Campus Benjamin Franklin. In addition to a brain MRI, a detailed medical history was taken. For this study 304 out of 1,437 subjects fulfilled the following inclusion criteria: (a) they had a proven acute stroke, (b) brain perfusion at baseline showed an area of hypoperfusion judged by a radiologist as corresponding to the acute infarct, and (c) this area of hypoperfusion was >10 mL according to the measurements at that time. The last criterion was chosen to only include patients with large enough perfusion deficits to be clinically relevant and credible, i.e., unlikely to be noise or artifacts. As shown in Figure 1, data of diffusion (TraceW and ADC images) and perfusion (dynamic susceptibility contrast perfusion source images) were assessed separately, checked for data quality, and then sent to StrokeSegApp using the software's implemented batch function. The diffusion data that were successfully evaluated by the program formed the **diffusion cohort** for the subsequent analysis of the accuracy of segmentations conducted by the StrokeSegApp.

To assess perfusion segmentation performance and real-life clinical utility of the StrokeSegApp, an additional subgroup was assembled with patients fulfilling the following criteria: (a) time from symptom onset 0–16 h, (b) NIHSS score > 5 at baseline, and (c) age 18–90 years. All subjects had an ICA or MCA (M1, M2) occlusion on magnetic resonance angiography. These criteria correspond to the inclusion in the DEFUSE 2 and 3 studies (8, 24, 25). These subjects are referred to as the **DEFUSE-like cohort**.

2.3 Image acquisition

Our complete imaging protocol has been previously published (22). In short, all examinations were performed on a 3 T MR scanner (MAGNETOM Trio A Tim System, Siemens Healthineers, Forchheim, Germany). DWI was conducted using a spin-echo echo-planar sequence (SE-EPI) [(TE = 93 ms; TR = 7,600 ms; acquisition matrix = 192 × 192 (spatial resolution); slice thickness = 2.5 mm without an interslice gap, b = 0 s/mm² and b = 1,000 s/mm²)]. Dynamic susceptibility contrast was conducted using a single-shot gradient-echo EPI sequence [(TE = 29 ms; TR = 1,390 ms; acquisition matrix = 128 × 128 (spatial resolution); slice thickness = 5 mm; interslice gap = 0.5 mm)] with a fixed dosage of 5 mL Gadovist® (Gadobutrol, 1 M, Bayer Schering Pharma AG, Berlin, Germany) followed by 20 mL of saline, both at an injection rate of 5 mL/s.



3 Test methods

3.1 Ground truth

The ground truth was the manual image delineation, carried out by either of two radiologists (KV or IG), each with more than 5 years of experience in stroke imaging. The radiologists analyzed the images with information on the side and territory of the stroke, but without access to previous MRI images. Diffusion lesions were delineated on TraceW maps with knowledge of the ADC map. Perfusion deficits were delineated in MRICron on relative mean transit time (relMTT) maps, which were created by the NUM/4 Neuro Perfusion (Siemens Healthineers AG, Forchheim, Germany) using standard singular value deconvolution (sSVD). Diffusion and perfusion lesion volumes were calculated. All these were once again checked for accuracy in 2024 (IG) and, if necessary, re-delineated accordingly.

For the DEFUSE-like cohort, a radiologist with 15 years of experience in stroke imaging (IG) manually segmented the perfusion deficits on the Tmax maps generated by the StrokeSegApp using block circulant singular value deconvolution (oSVD).

3.2 Index test

The index test was the AI tool StrokeSegApp version 1.3 (Standalone research package Stroke Segmentation 1.3.0) by Siemens Healthineers AG (Forchheim, Germany), a research application

designed to become an integral part of the neuroradiological routine in the MRI-based diagnosis of ischemic strokes and hemorrhagic lesions.

The StrokeSegApp automatically segments acute infarcts in ADC and TraceW images (b-value 1,000 s/mm²) as well as the perfusion lesions in perfusion-based dynamic susceptibility contrast images. The application provides both diffusion and perfusion lesion volumes and automatically calculates the PWI-DWI-Mismatch ratio.

Although only the default block circulant singular value deconvolution with oscillation index (oSVD) was used in this study, the program offers three more deconvolutional methods, namely singular value decomposition (sSVD, the original method from Ostergaard (26, 27)), block circulant singular value deconvolution with a fixed cutoff (cSVD) (28), and a Tikhonov regularized Fourier deconvolution (29, 30).

The resulting segmentations can be viewed as overlays on top of the underlying contrast maps (e.g., TraceW and PWI). The perfusion maps generated can also be viewed in the application (relative cerebral blood flow, relative cerebral blood volume, relMTT, and Tmax).

Restricted perfusion lesions (shown as Regions of Interest, ROI) are delineated on Tmax maps. Tmax is calculated using a regularized deconvolution method, namely oSVD (28), with an automatically detected global Arterial Input Function (AIF). The AIF is the average of all AIF candidates from the slice with the most suitable AIF candidates. The AIF candidates are selected from prioritized anatomical regions in the brain, following the areas described by Ebinger et al. and Mouridsen et al. (31, 32).

The lesion segmentation and mirroring of the DWI and PWI lesion are based on deep learning models, which were trained on manually segmented and annotated ground truth data (12,989 for DWI, with a mean age of 70 years, 48% female; 770 for PWI with a mean age of 69 years, 47% female), followed by an internal validation with 1,664 (mean age of 72 years, 52% female) and 98 datasets (mean age of 71, 51% female). Internal testing on 1,897 subjects (mean age of 73 years, 53% female) for detecting infarct on DWI revealed a sensitivity of 0.92 and a specificity of 0.95. For PWI (103 subjects, mean age of 72 years, 52% female), Siemens Healthineers did not determine test accuracy.

The data analyzed in this study were neither used for previous model training nor validation.

The StrokeSegApp reviewed MR images for the analysis without access to original referrals, radiological reports, or neuroradiological diagnoses.

All patient's datasets were sent to StrokeSegApp thresholding T_{max} to 6 s. For index testing, the program's default setting was used. For two cases with considerable motion, the "PWI Excessive Motion Threshold" was adjusted (15 → 50 mm) to enable successful processing of the images.

4 Analysis

4.1 Approaches for evaluating metrics of diagnostic precision

4.1.1 Diffusion cohort

To evaluate the diffusion segmentation performance, we first compared the manually segmented lesion volumes with those from the StrokeSegApp. The Bland–Altman plot shows us the volumetric agreement between the ground truth and the index test. We then calculated the Dice–Sørensen coefficient (dice similarity coefficient, DSC) to evaluate the overlap between the two binarized segmentations on the TraceW maps. The DSC is defined as twice the volume of the intersection between two masks divided by the sum of their volumes and serves to quantify their spatial similarity whereby 0 stands for no and 1 for a perfect overlap (22, 33, 34). To assess the impact of lesion size, time to MRI, NIHSS score, and the presence of large vessel occlusion (LVO) on the DSC, a linear model (OLS) was used. Lesion size and time to MRI were included as continuous variables, while LVO was a binary predictor. The NIHSS scores were included as a continuous variable where available, with missing values ($n = 2$) retained in the analysis. The model also included an intercept to account for baseline effects.

4.1.2 DEFUSE-like cohort

For the DEFUSE-like cohort, only patients with both diffusion and perfusion data available were included (see Figure 1). We compared the diffusion data for the new subgroup in the same way as in the diffusion cohort. The evaluation of segmentation performance for perfusion data was carried out by comparing the physician's T_{max} ROI and the StrokeSegApp's T_{max} ROI with thresholding T_{max} each to 6 s. Here, too, we created a Bland–Altman plot and calculated the DSC. We also used a linear model in the same way. As all subjects had an LVO as an inclusion criterion, this was not analyzed further here.

Finally, to determine the clinical utility of the software, we assessed the decision for or against mechanical thrombectomy (MT) in line with the criteria of the DEFUSE 3 study, based on the manual and automatically segmented DWI and PWI data. In the DEFUSE 3 study, patients with the following mismatch criteria were eligible for MT: infarct volume < 70 mL, penumbra volume ≥ 15 mL, and mismatch ratio (reduced perfusion/core infarction) ≥ 1.8 (8). We then calculated sensitivity and specificity for the StrokeSegApp versus the ground truth and furthermore determined the likelihood ratio (likelihood-ratio test, LRT).

The statistical analyses were performed with MATLAB Version 9.1.0.441655 (R2016b),¹ and statistical graphics were created in IBM SPSS Statistics Version 29.0.0.0 (241).² The MRI images in the figures were visualized using FSLeves Version 1.4.5.³

5 Results

5.1 Participants

We included 241 subjects for the diffusion segmentation analysis (diffusion cohort) and 56 patients for analyzing perfusion segmentation and indication for MT (DEFUSE-like cohort). Cases of missing sequences or data compromised by excessive patient movement (defined as head motion of more than 50 mm) were excluded. Patients were also rejected if the physician's original segmentation showed technical processing errors resulting from incorrectly saved physician's lesion segmentations (ROI) that were not congruent with the original brain map and could not be coregistered. The StrokeSegApp failed to analyze 8 perfusion datasets as it was unable to identify an AIF in these scans (see Figure 1).

Figure 2 shows an example of a successful segmentation of both lesion types in one patient, and Table 1 summarizes the baseline demographic characteristics of participants for both cohorts.

5.2 Test results for the diffusion cohort

The StrokeSegApp was able to correctly identify a DWI lesion in 212 out of 241 (88%). Of the 29 missed detections of diffusion lesions, all infarcts were smaller than 10 mL, and 20 (69%) were smaller than 0.5 mL.

The Bland–Altman plot (Figure 3) shows the volumetric agreement between both methods. The mean difference between the two methods was 0.49 mL, suggesting only a very small systematic difference between the measurements of the two methods. Especially at higher mean values, we see that the datapoints increasingly approach the limits of agreement and extreme outliers occur. Even though a few of these show poor spatial overlap with the ground truth (Figure 4), the majority of cases demonstrate acceptable or good

1 MathWorks, Natick, MA, USA, <https://www.mathworks.com>.

2 IBM Corp., Armonk, NY, USA, <https://www.ibm.com/products/spss-statistics>.

3 FMRIB Analysis Group, University of Oxford, Oxford, UK, <https://fsl.fmrib.ox.ac.uk/fsl/fsleaves>.

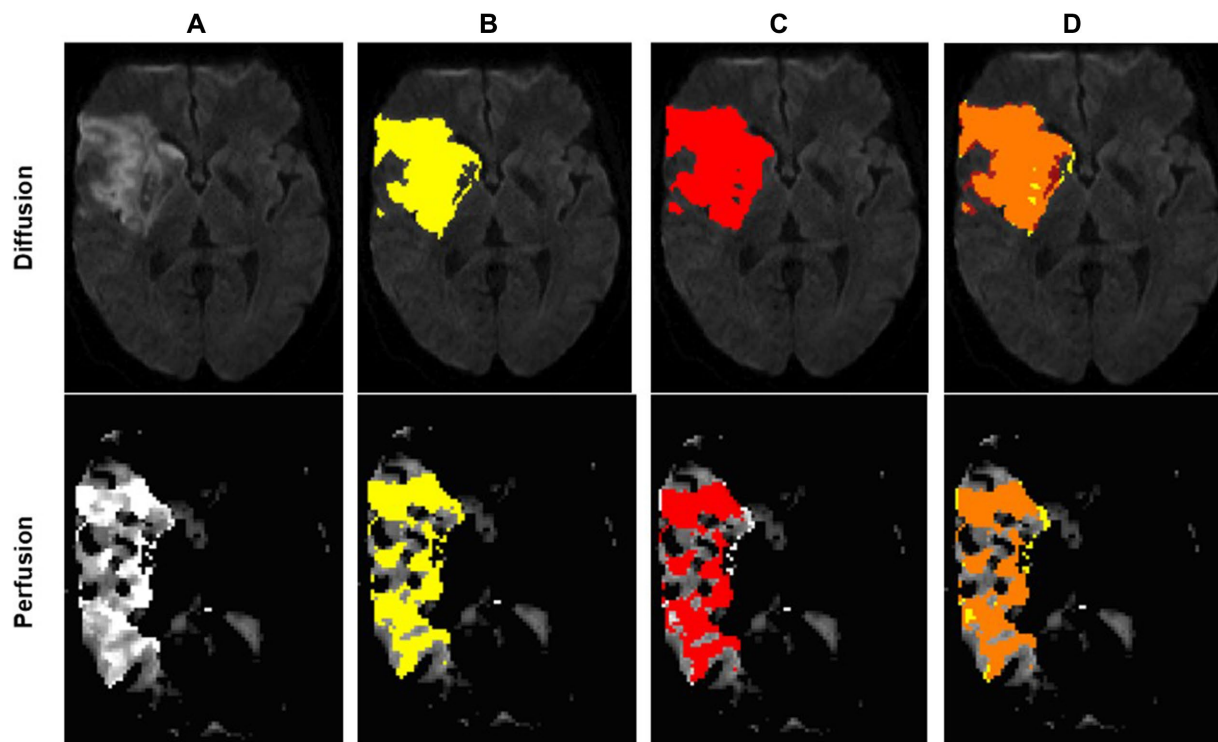


FIGURE 2

Example of an optimal segmentation of the program compared to the ground truth. The patient had a proximal M1 occlusion on the right side and suffered an infarction of the basal ganglia and the frontal operculum. The diffusion segmentation is shown above on TraceW maps, perfusion can be viewed below on Tmax maps thresholded to 6 s. The diffusion and perfusion segmentation are not displayed in the same slice. DSC for diffusion was 0.87 with an overestimation of the lesion size by the program of 9.63 mL. For perfusion, the dice score was even higher at 0.94, whereby the StrokeSegApp underestimated the deficit slightly by 4.37 mL. Following the DEFUSE 3 criteria, the approaches came to the same conclusion that a mechanical thrombectomy is not indicated because the penumbra is too small. (A) Blank scan used for segmentation, (B) ground truth ROI, (C) StrokeSegApp ROI, and (D) both ROI superimposed, whereby orange indicates areas of overlapping mapping, yellow shows areas missed by the StrokeSegApp and red displays diffusion/perfusion deficits overestimated by the application.

segmentation performance, despite a large volume difference (Figure 5). For lesions <10 mL, the median percentage volume deviation was 22.6% (IQR 7.6–68.9%), while for lesions >10 mL, it was 27.9% (IQR 14.7–53.3%). These findings suggest that, although absolute differences can occasionally be substantial—particularly for larger lesions, which considerably affect the limits of agreement—the relative differences remain consistent and fall within an acceptable range.

The mean of the DSC for the 212 subjects in whom the software recognized a diffusion lesion was 0.60 (95% CI: 0.57–0.63). The spatial overlap varied greatly depending on the size of the lesion. For infarcts ≥ 10 mL ($n = 88$), the DSC was 0.72 (95% CI: 0.67–0.76). For small diffusion lesions (<10 mL, $n = 124$), it was markedly lower at 0.53 (95% CI: 0.49–0.57). Figure 6 shows the distribution of the DSC.

To better understand the factors influencing segmentation performance, a linear model was applied. The analysis confirmed that lesion size and time to MRI are significant predictors of the DSC. For every additional 1 mL of lesion size, the DSC increased by 0.003 (95% CI: 0.002–0.005, $p < 0.001$), and for every hour from stroke onset to MRI, the DSC increased by 0.014 (95% CI: 0.009–0.019, $p < 0.001$). However, neither the presence of a large vessel occlusion ($p = 0.816$) nor the NIHSS score ($p = 0.704$) was found to significantly influence DSC values in this dataset.

5.3 Test results for DEFUSE-like cohort

We first looked at diffusion and perfusion results separately before analyzing them together as part of the decision for or against MT. Diffusion segmentation performance was similar to that in the diffusion cohort with a DSC of 0.61 (95% CI: 0.53–0.68).

In terms of perfusion, the program was able to correctly determine the presence of hypoperfusion corresponding to the area of the acute stroke in all 56 cases. Volumetric agreement on perfusion lesion volumes is shown in Figure 7A. With a mean deviation of -7.69 mL, the StrokeSegApp consistently and slightly underestimates the size of the perfusion lesions. As the size of the deficit grows, we can see that the scatter of the datapoints increases volume-wise. Nevertheless, large volume deviations—and even outliers—are usually accompanied by good segmentation performance in terms of DSC (see Figure 8). Large lesions are associated with better automated segmentation performance than small ones as seen in Figure 7B.

The mean DSC for perfusion lesions was 0.80 (95% CI 0.76–0.85). Focused on the lesions that have a volume of ≥ 10 mL ($n = 46$), we see a DSC of 0.85 (95% CI: 0.83–0.87). See Figure 6 for a summary of the data's distribution. The OLS explained 20.2% of the variability in DSC ($R^2 = 0.202$, $p = 0.0079$). Lesion size ($p = 0.093$), and time to MRI ($p = 0.082$) showed trends toward significance, with a DSC increase of 0.0009 per 1 mL lesion size and 0.0077 per hour from stroke onset to

TABLE 1 Baseline patient characteristics.

Variable	Diffusion cohort ¹ (<i>n</i> = 241)	DEFUSE-like cohort ¹ (<i>n</i> = 56)
Age in years	71 (12)	72 (12)
Sex		
Female	110 (45.6%)	33 (58.9%)
Male	131 (54.4%)	23 (47.1%)
Time window		
Time from stroke to MRI in hours	8.11 (6.9)	5.3 (4.9)
0–6 h	124 (51.5%)	36 (64.3%)
6–16 h	75 (31.1%)	20 (35.7%)
16–24 h	42 (17.4%)	–
LVO distribution		
Cases of LVO	134 (55.6%), hereof	56 (100%), hereof
Only ICA (right, left or both)	23 (17.2%)	8 (14.3%)
Only M1 (right, left or both)	36 (26.9%)	16 (28.6%)
Only M2 (right, left or both)	48 (35.8%)	17 (30.4%)
ICA and MCA (right, left or both)	27 (20.1%)	15 (26.8%)
NIHSS²		
0	23 (9.5%)	0 (0%)
1–4	97 (40.2%)	0 (0%)
5–15	97 (40.2%)	36 (64.3%)
16–20	18 (7.5%)	16 (28.6%)
21–42	6 (2.5%)	4 (7.1%)
Diffusion lesion size (mL):		
<0.5	41 (17.0%)	2 (3.6%)
≥0.5 to <10	112 (46.5%)	16 (28.6%)
≥10	88 (36.5%)	38 (67.8%)
Perfusion lesion size (mL):		
<10	–	10 (17.9%)
≥10	–	46 (82.1%)

¹Mean (SD); *n* (%).

²NIHSS, National Institutes of Health Stroke Scale.

MRI, respectively. The NIHSS ($p = 0.222$) did not significantly influence DSC.

To evaluate the clinical utility of the StrokeSegApp, we compared the decision for and against MT between the physician and the software. Out of 28 cases where MT would have been indicated based on the ground truth, StrokeSegApp recommended MT for 23 subjects (sensitivity of 82.1%; 95% CI: 63.1–93.9%). In cases where MT would not be recommended by the ground truth ($n = 28$), the AI segmentation reached the same conclusion in 27 patients (specificity of 96.4%; 95% CI: 81.7–99.9%); see Table 2 for cross-tabulation. In all five cases where the values determined by the program did not indicate MT, this was because the automatically calculated penumbra

was too small. In one case, the diffusion lesion was overestimated. Visual inspection confirmed that the areas identified by the StrokeSegApp aligned well with the infarct core and the corresponding perfusion deficit. However, the tool's delineation boundaries differed enough from those of the radiologist to result in a significant volumetric discrepancy. The likelihood ratios LR+ and LR– were 22.81 and 0.19. The large positive LR (>10) implies that a positive test result greatly increases the probability of recommending MT. Conversely, an LR– of 0.19 suggests that a negative test result is moderately effective in ruling out the need for MT.

6 Discussion

We carried out an independent evaluation of the StrokeSegApp using a large retrospective real-life cohort of patients (>200 datasets) who underwent MRI screening for acute ischemic stroke. The StrokeSegApp showed very good reliability and accuracy in recognizing acute ischemic changes on DWI and was also able to segment the ischemic core with a mean difference of 0.49 mL compared to human radiology experts. In a fully automated fashion, the AI was able to produce usable perfusion maps for 97% of patients in the sample. Its segmentations of hypoperfusion showed a high degree of spatial overlap with a human rater and a mean volumetric difference of only 7.68 mL. The linear model results confirm that lesion size and time to MRI are the primary determinants of segmentation performance, with larger lesions and later imaging times associated with higher dice similarity coefficient values. For the diffusion cohort, these effects were pronounced, with lesion size and timing significantly influencing DSC. In the DEFUSE-like cohort, lesion size and time to MRI showed trends toward significance, but their impact was less marked, likely reflecting the homogeneity of patients with large vessel occlusion. The NIHSS did not significantly influence DSC in either cohort, suggesting that lesion size and time to MRI already account for stroke severity and occlusion-related variability. These findings align with expectations and emphasize the importance of lesion size and timing in determining segmentation accuracy. The lower DSC observed in smaller lesions should be understood in the context that segmentation errors in objects with a small volume are penalized more harshly. Given their clinical importance in early-stage stroke assessment, proven strategies to improve the segmentation of small lesions could include, for example, data enrichment with additional real or synthetic cases (35) and lesion-volume-based data sampling and loss functions (36).

When the StrokeSegApp is used as part of a decision matrix for recanalization in patients with a large vessel occlusion in the anterior circulation (based on the criteria of the DEFUSE 2 and DEFUSE 3 clinical trials), it would provide 82% sensitivity and 96% specificity in selecting candidates for reperfusion therapy, compared to a human gold standard. While the sensitivity result is promising, the automated analysis would have ruled out thrombectomy in an additional five cases. Even though visual inspection confirmed that the areas identified by the software corresponded well to the infarct core and matching perfusion deficit, the delineation boundaries differed sufficiently from those selected by the radiologist in all cases, leading to rejection due to a PWI/DWI mismatch <1.8. Such strict decision criteria are necessary in large clinical trials such as DEFUSE, but they hold less significance in routine clinical practice, where information

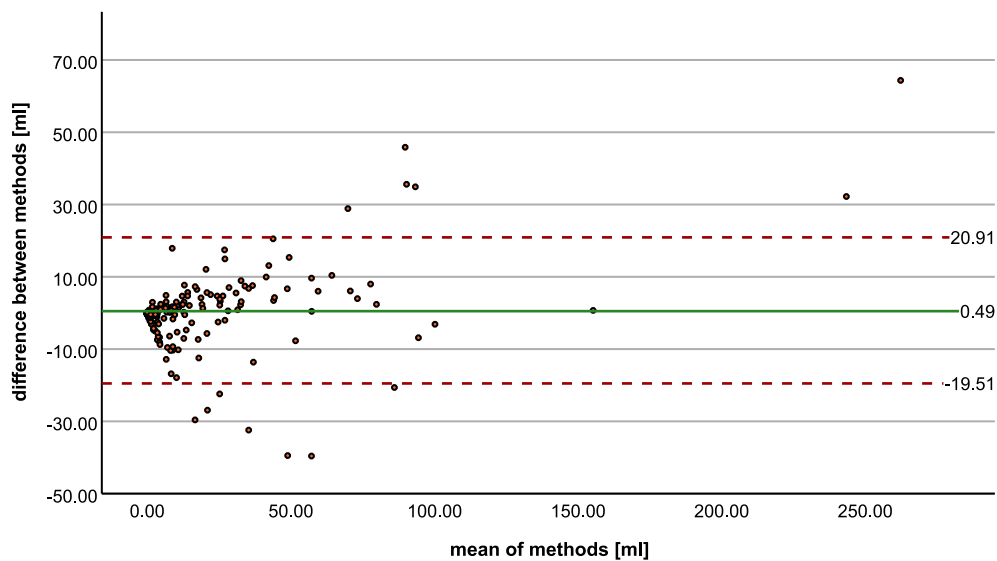


FIGURE 3

Bland–Altman plot shows the agreement between radiologist and StrokeSegApp for volumes of diffusion lesions.

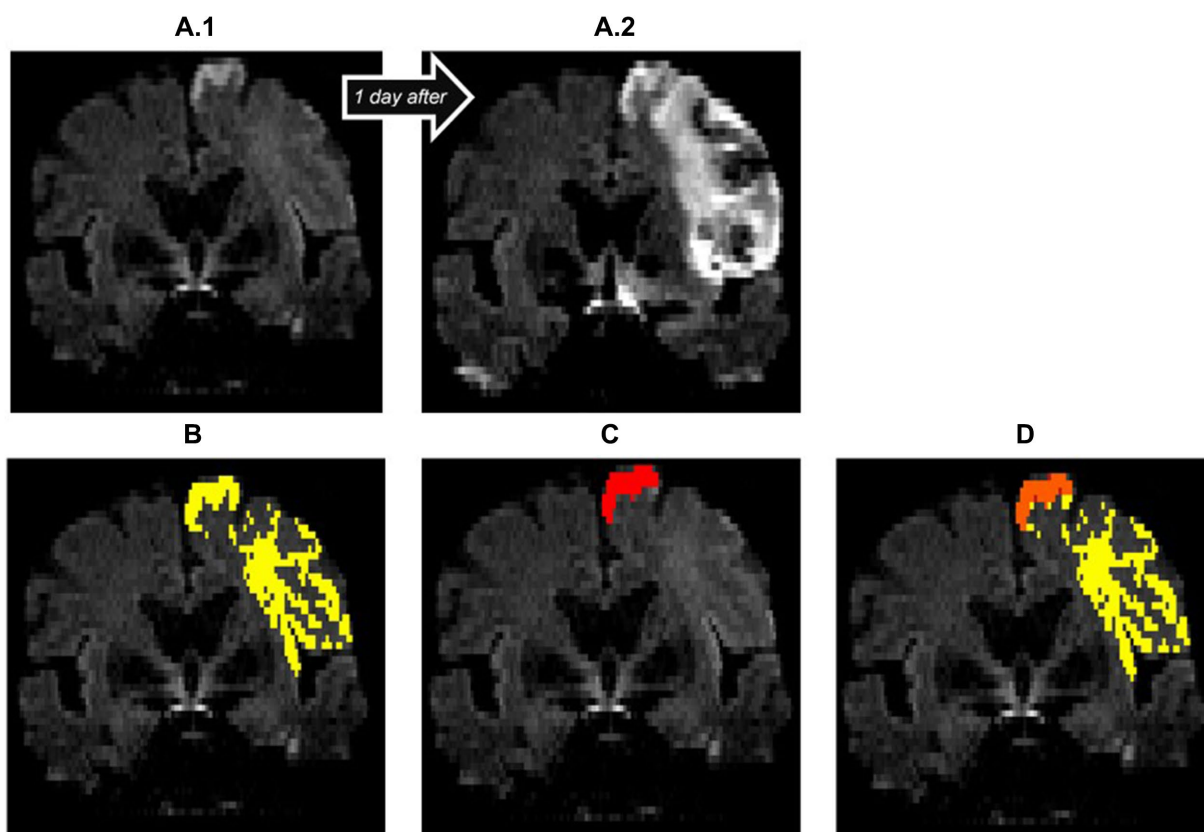


FIGURE 4

Insufficient diffusion segmentation by the AI on a TraceW map. The clearly hyperintense cortical infarction of the superior frontal gyrus near the cerebral falx was correctly recognized. The remaining faint cortical and subcortical lesions of the frontal lobe, which are hyperintense compared to the contralateral hemisphere on day 1 (A.1) and clearly demarcated as part of the acute stroke on follow-up imaging on day 2 (A.2), were not recognized by the StrokeSegApp. The dice score is poor (0.11), the program only delineates 2.13 mL of the 31.73 mL lesion. (A.1) blank scan used for segmentation, (A.2) follow-up imaging on day 2, (B) ground truth ROI, (C) StrokeSegApp ROI, (D) both ROI superimposed, whereby orange indicates areas of overlapping mapping, yellow shows areas missed by the StrokeSegApp and red displays diffusion deficits overestimated by the application.

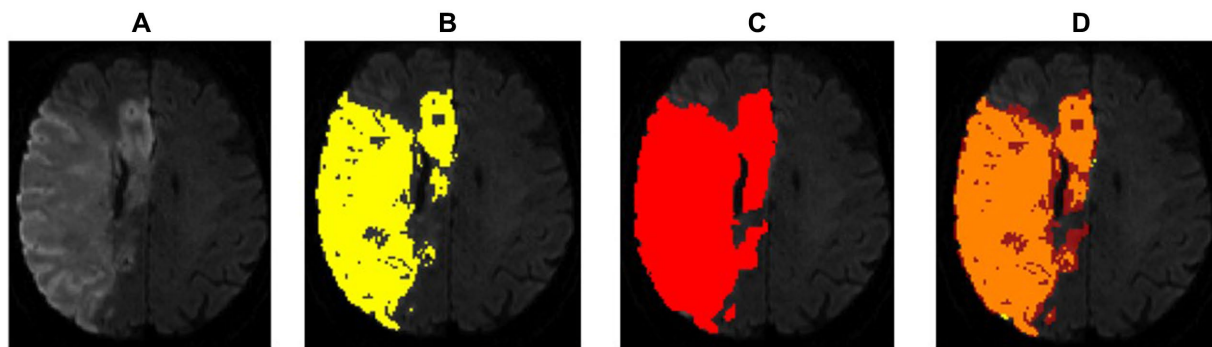


FIGURE 5

Example of a regionally correct segmentation in DWI with a high dice score (0.87), but a large deviation in lesion volumes. For such large lesions, even slight discrepancies in the thresholds used for delineation by the radiologist and segmentation software, respectively, can lead to large absolute differences in volume. In this case, the StrokeSegApp overestimated the DWI volume by 64.35 mL. However, at that end of the range, such thresholding differences would not typically have clinical relevance; therefore, pure volumetric comparisons are not necessarily the most suitable metric for quantifying a program's performance. (A) Blank scan used for segmentation, (B) ground truth ROI, (C) StrokeSegApp ROI, and (D) both ROI superimposed, whereby orange indicates areas of overlapping mapping, yellow shows areas missed by the StrokeSegApp and red displays diffusion deficits overestimated by the application.

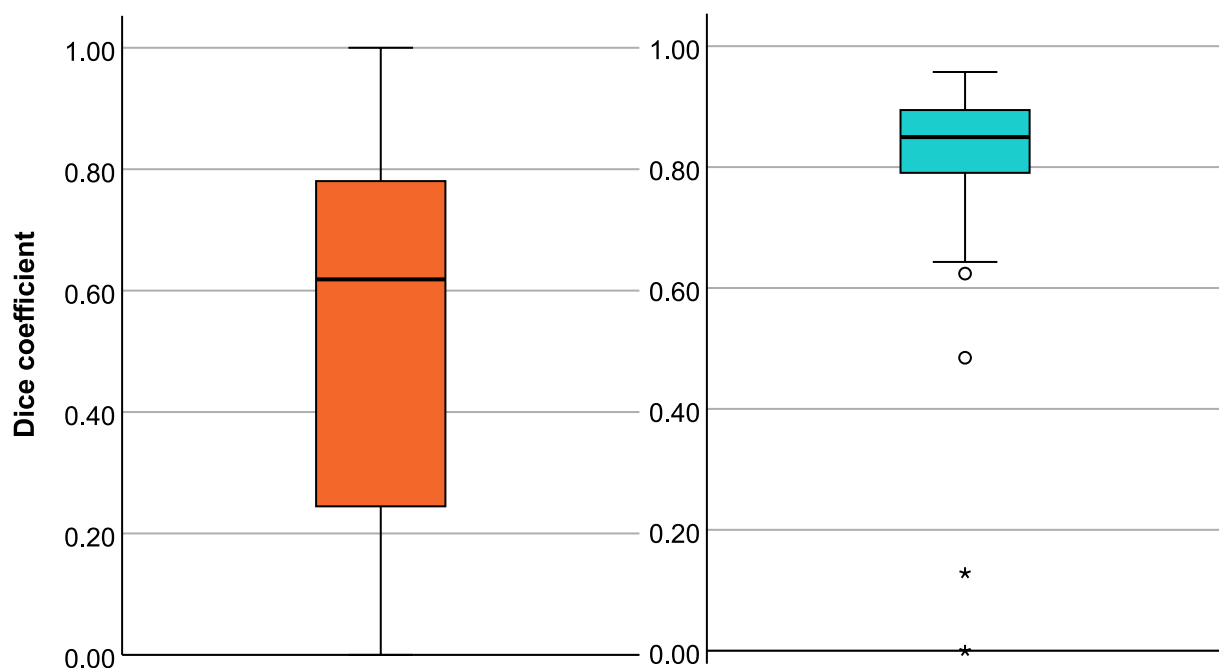


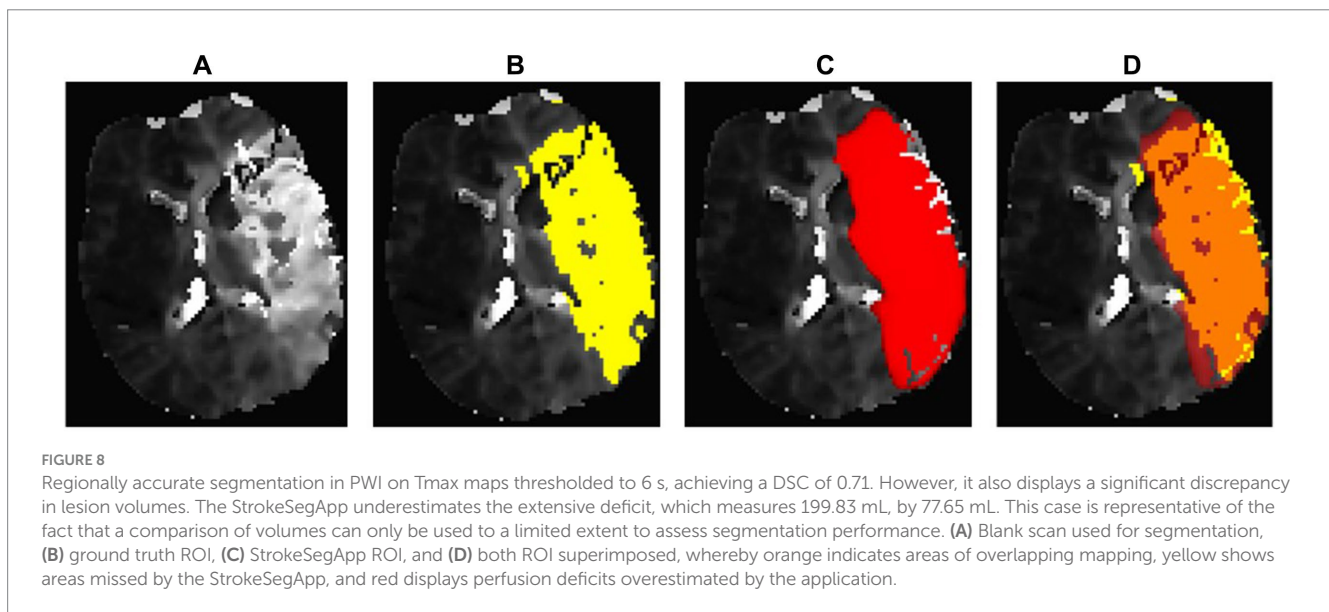
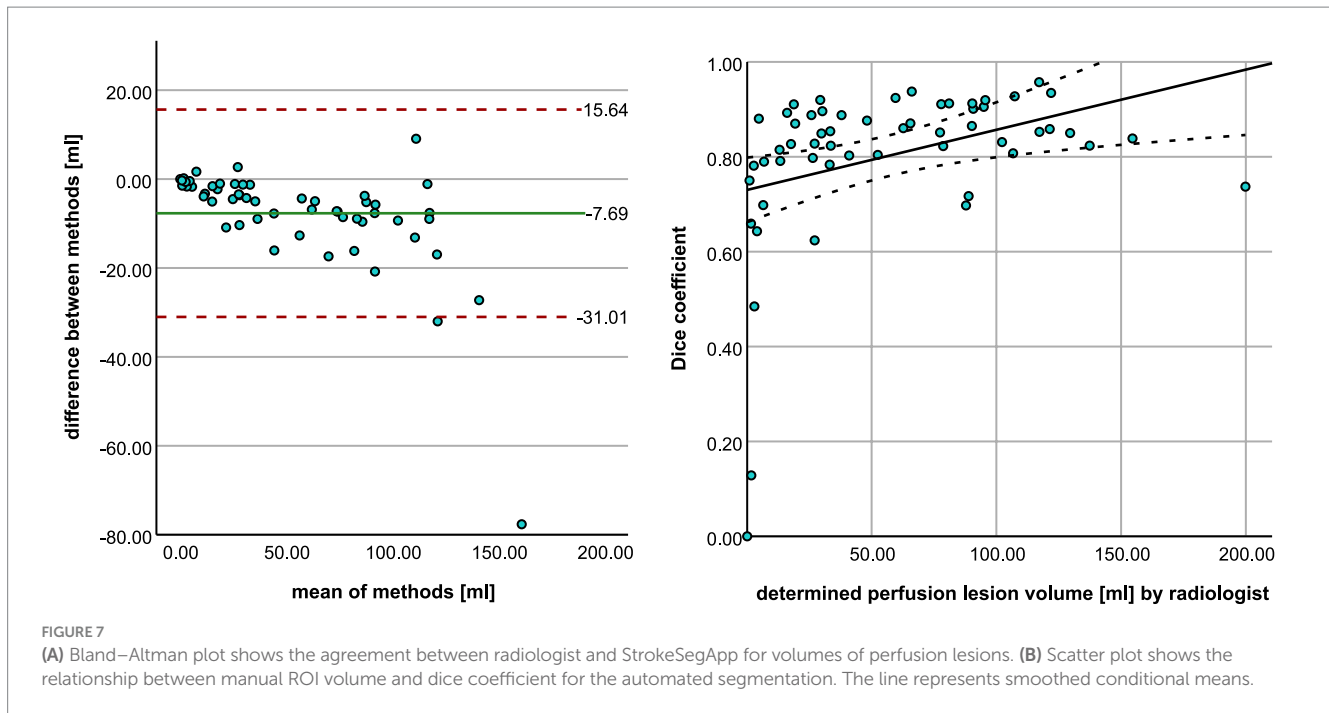
FIGURE 6

Box plots of dice coefficient (DSC) for the diffusion segmentation (A) and the perfusion segmentation (B).

is interpreted more holistically without adhering rigidly to specific thresholds. Additionally, some degree of interrater variability is expected, both in human and software analyses, which underscores the challenges of achieving perfect alignment.

A direct comparison of Siemens Healthineers' software with other programs that are commercially available cannot be made, as publications on segmentation performance in ischemic strokes rarely include a comparison of spatial overlap and volumetric congruence between the software providing automated segmentation and a ground truth. In contrast, the number of publications on the

performance of non-commercial applications for segmenting diffusion lesions is quite high—mostly with the provision of dice scores. In a study also using data from the 1000Plus study (but not using the same cohort as here), the algorithm developed by the research group achieved a DSC of 0.43 compared to manual segmentations (37). The ISLES challenges—an open competition to foster the development and innovation of advanced tools for ischemic stroke analysis—reflect the development in ischemic stroke lesion segmentation of non-profit participants as a steady improvement. While the top 2018 model reached a dice score of 0.51, recent advancements have pushed this to



0.69, and the 2022 challenge has seen scores as high as 0.78, demonstrating progress in segmentation accuracy across different imaging modalities (38).

The StrokeSegApp can compete with those results and performs comparably to commercial software solutions on which data are available. In one publication ($n = 131$), RapidAI showed a mean deviation of diffusion lesion volumes of -1.95 mL (StrokeSegApp: 0.49 mL) with limits of agreement of -12.89 mL to 9 mL (StrokeSegApp: -19.51 mL to 20.91 mL) (39). The e-Stroke software distributed by Brainomix showed a mean volume deviation of 4 mL in the segmentation of CT perfusion for determining the infarct core

($n = 551$) (40). OleaSphere from Olea Medical achieved a dice score of 0.52 (StrokeSegApp: 0.60) compared to a ground truth segmented by physicians using DWI ($n = 75$) (41).

Regarding perfusion, a study using RapidAI reported a mean deviation of 7.60 mL (StrokeSegApp: -7.69 mL) from a reference determined by radiologists ($n = 18$) (25). In using more than just a mere volumetric comparison, our voxel-based approach to the validation of perfusion segmentation against a ground truth adds an additional level of rigor in terms of validation of automated segmentation software as compared to the current literature. This underlines the need for a standardized guideline for evaluation

TABLE 2 Cross-tabulation of decisions for or against MT based on data from the radiologist (ground truth) and the StrokeSegApp.

	Ground truth: MT indicated	Ground truth: MT not indicated	Total
StrokeSegApp: MT indicated	23	1	24
StrokeSegApp: MT not indicated	5	27	32
Total	28	28	56

Sensitivity	Specificity	Likelihood Ratio (LR)	
		LR+	LR–
82.1%	96.4%	22.81	0.19
95% CI: 63.1–93.9%	95% CI: 81.7–99.9%	95% CI: 5.13–631	95% CI: 0.06–0.45

metrics, as proposed by Müller et al. (22), which helps manufacturers align their evaluations and enables a robust comparison of the relative strengths and weaknesses of different AI applications. With a dice score of 0.8 (or 0.85 for lesions ≥ 10 mL), the StrokeSegApp shows very good overlap with human raters and sets standards for current and future applications.

In a real-time clinical context, the StrokeSegApp could be integrated either directly at the scanner or at a connected processing node, where the calculation is automatically triggered upon the arrival of new data. With both approaches the resulting DICOM image and report data could be delivered to the PACS reading station. Current calculation times for the entire stroke segmentation pipeline range from approximately 1–2 min on standard computer hardware making the use of the software feasible in an actual clinical workflow. The processing time can increase considerably in cases with substantial artifacts, such as excessive patient movement. Beyond delays, poor contrast and strong noise can also adversely affect the analysis. This occurs in two main ways. First, poor contrast in the original perfusion-weighted images, particularly when the contrast bolus dip is reduced, can degrade the quality of pre-fusion map generation. Second, segmentation performance may decline, as it is closely tied to the accuracy of the perfusion maps. To address these challenges, Siemens has implemented a method in the current software version that automatically rejects processing if minimal quality requirements are not met. The rejection criteria include: (a) insufficient contrast bolus drop in the PWI data relative to baseline noise, (b) excessive motion artifacts between timepoints, and (c) an arterial input function that is too weak to produce accurate perfusion maps. The output includes, in addition to the endpoint results (mismatch volumes and ratio) visualizations of segmented lesions in the context of the original data, details about the arterial input function used in the calculation, and brain extraction masks. This additional information aids radiologists in decision-making, builds trust in the derived values, and helps identify errors in cases where the automatic calculation fails. Incorporating mechanisms for users to provide feedback about the application to the manufacturer would further enhance its usability and reliability.

The integration of AI into neuroimaging, particularly in the context of segmenting stroke lesions, demonstrates the transformative potential for diagnostic and therapeutic advancements but also raises significant ethical concerns. A core challenge lies in the dependency of AI models on the quality and consistency of input data, spanning training, validation, and clinical application phases. Errors stemming from algorithmic biases in segmentation (42) can directly affect clinical decisions and patient outcomes. Moreover, ensuring a diverse and representative dataset for training and validation is critical to addressing disparities in healthcare (43), especially for marginalized groups, who risk underrepresentation in neuroimaging repositories (44). The necessity for transparency in training strategies, algorithmic decisions, and the management of sensitive patient data is paramount (44). This transparency ensures ethical AI deployment, builds trust, and enhances clinical applicability. Collaborative interdisciplinary approaches that merge clinical, technical, and ethical expertise are essential to mitigate biases and promote equitable AI applications across diverse populations.

Our results should be viewed in the context of the study's limitations. This is a single-center retrospective analysis of data acquired between 2009 and 2013; therefore, we cannot rule out that the StrokeSegApp may perform differently on data acquired using contemporary scanning protocols. The ground truth comes from manual segmentation, which is prone to subjective variability.

7 Conclusion

The Siemens Healthineers StrokeSegApp provides automated segmentation of diffusion and dynamic susceptibility contrast perfusion on MR images in acute ischemic stroke patients, achieving a high degree of accuracy compared to a human expert. Therefore, it is comparable to other existing and already commercially available solutions.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

Ethics statement

The studies involving humans were approved by Ethikkommission der Charité – Universitätsmedizin Berlin EA4/026/08. This retrospective study used data from the 1000Plus study (clinicaltrials.org; NCT00715533). All study participants gave informed consent for participation in the study.

Author contributions

L-ST: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing – review & editing, Validation. AK: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review &

editing. KV: Data curation, Resources, Writing – review & editing. JF: Supervision, Writing – review & editing. SH: Software, Writing – review & editing. EG: Software, Writing – review & editing. IG: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding for part of the publication costs of this article is provided by the Charité Open Access Publication Fund.

Acknowledgments

The authors gratefully acknowledge Uchralt Temuulen and Ralf Mecke for their consulting and technical support.

Conflict of interest

Outside the submitted work JF reports consulting and advisory board fees from AbbVie, AC Immune, Alnylam Pharmaceuticals, Alzheon, Artemida, BioClinica/Clario, Biogen, Bristol Myers Squibb, Brainomix, Cerevast, C2N Diagnostics, Daiichi-Sankyo, Eisai, Eli

Lilly, F. Hoffmann-LaRoche AG, GlaxoSmithKline, Guerbet, Ionis Pharmaceuticals, IQVIA, Janssen, Julius Clinical, jung diagnostics, Lantheus Medical Imaging, Lumosa Therapeutics, Merck, Novo Nordisk, Octapharma AG, Premier Research, ProPharma Group, Prothena Biosciences, Regeneron Pharmaceuticals, Roche, Syneos, Tau Rx, Vertex Pharmaceuticals, and Worldwide Clinical Trials.

SH and EG were employed by company Siemens Healthineers AG and developed and provided the StrokeSegApp.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Campbell BCV, Khatri P. Stroke. *Lancet*. (2020) 396:129–42. doi: 10.1016/S0140-6736(20)31179-X
- Donkor ES. Stroke in the 21(st) century: a snapshot of the burden, epidemiology, and quality of life. *Stroke Res Treat*. (2018) 2018:3238165. doi: 10.1155/2018/3238165
- Lin MP, Tsiygoulis G, Alexandrov AV, Chang JJ. Factors affecting clinical outcome in large-vessel occlusive ischemic strokes. *Int J Stroke*. (2015) 10:479–84. doi: 10.1111/ijls.12406
- Czap AL, Sheth SA. Overview of imaging modalities in stroke. *Neurology*. (2021) 97:S42–s51. doi: 10.1212/WNL.00000000000012794
- Go KG. The normal and pathological physiology of brain water. *Adv Tech Stand Neurosurg*. (1997) 23:47–142. doi: 10.1007/978-3-7091-6549-2_2
- Tong E, Sugrue L, Wintermark M. Understanding the neurophysiology and quantification of brain perfusion. *Top Magn Reson Imaging*. (2017) 26:57–65. doi: 10.1097/RMR.0000000000000128
- Yi Y, Liu Z, Wang M, Sun M, Jiang X, Ma C, et al. Penumbra in acute ischemic stroke. *Curr Neurovasc Res*. (2021) 18:572–85. doi: 10.2174/1567202619666211231094046
- Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med*. (2018) 378:708–18. doi: 10.1056/NEJMoa1713973
- Deutsche Gesellschaft für Neurologie (DGN). *Acute therapy of ischemic stroke*. 5th edition. Berlin: DGN. (2023). Available from: <https://dgn.org/leitlinie/akuttherapie-des-ischemischen-schlaganfalls> (Accessed April 14, 2024).
- Berge E, Whiteley W, Audebert H, De Marchis GM, Fonseca AC, Padiglioni C, et al. European stroke organisation (ESO) guidelines on intravenous thrombolysis for acute ischaemic stroke. *Eur Stroke J*. (2021) 6:I–lxii. doi: 10.1177/2396987321989865
- Krizhevsky A, Sutskever I, Hinton GE. Image net classification with deep convolutional neural networks. *Commun ACM*. (2017) 60:84–90. doi: 10.1145/3065386
- Lui YW, Chang PD, Zaharchuk G, Barboriak DP, Flanders AE, Wintermark M, et al. Artificial intelligence in neuroradiology: current status and future directions. *AJNR Am J Neuroradiol*. (2020) 41:E52–9. doi: 10.3174/ajnr.A6681
- Olthof AW, van Ooijen PMA, Rezazade Mehrizi MH. Promises of artificial intelligence in neuroradiology: a systematic technographic review. *Neuroradiology*. (2020) 62:1265–78. doi: 10.1007/s00234-020-02424-w
- Yedavalli VS, Tong E, Martin D, Yeom KW, Forkert ND. Artificial intelligence in stroke imaging: current and future perspectives. *Clin Imaging*. (2021) 69:246–54. doi: 10.1016/j.clinimag.2020.09.005
- Ben Alaya I, Limam H, Kraiem T. Applications of artificial intelligence for DWI and PWI data processing in acute ischemic stroke: current practices and future directions. *Clin Imaging*. (2022) 81:79–86. doi: 10.1016/j.clinimag.2021.09.015
- Bridge CP, Bizzo BC, Hillis JM, Chin JK, Comeau DS, Gauriau R, et al. Development and clinical application of a deep learning model to identify acute infarct on magnetic resonance imaging. *Sci Rep*. (2022) 12:2154. doi: 10.1038/s41598-022-06021-0
- Krag CH, Müller FC, Gandrup KL, Raaschou H, Andersen MB, Břejnebo MW, et al. Diagnostic test accuracy study of a commercially available deep learning algorithm for ischemic lesion detection on brain MRIs in suspected stroke patients from a non-comprehensive stroke center. *Eur J Radiol*. (2023) 168:111126. doi: 10.1016/j.ejrad.2023.111126
- Nael K, Gibson E, Yang C, Ceccaldi P, Yoo Y, Das J, et al. Automated detection of critical findings in multi-parametric brain MRI using a system of 3D neural networks. *Sci Rep*. (2021) 11:6876. doi: 10.1038/s41598-021-86022-7
- Subudhi A, Dash P, Mohapatra M, Tan RS, Acharya UR, Sabut S. Application of machine learning techniques for characterization of ischemic stroke with MRI images: a review. *Diagnostics (Basel)*. (2022) 12:2535. doi: 10.3390/diagnostics12102535
- Deutschmann H, Hinteregger N, Wiefßpeiner U, Kneihsl M, Fandler-Höfler S, Michenthaler M, et al. Automated MRI perfusion-diffusion mismatch estimation may be significantly different in individual patients when using different software packages. *Eur Radiol*. (2021) 31:658–65. doi: 10.1007/s00330-020-07150-8

21. Wardlaw JM, Mair G, von Kummer R, Williams MC, Li W, Storkey AJ, et al. Accuracy of automated computer-aided diagnosis for stroke imaging: a critical evaluation of current evidence. *Stroke*. (2022) 53:2393–403. doi: 10.1161/STROKEAHA.121.036204
22. Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes*. (2022) 15:210. doi: 10.1186/s13104-022-06096-y
23. Hotter B, Pittl S, Ebinger M, Oepen G, Jegzentis K, Kudo K, et al. Prospective study on the mismatch concept in acute stroke patients within the first 24 h after symptom onset - 1000Plus study. *BMC Neurol*. (2009) 9:60. doi: 10.1186/1471-2377-9-60
24. Ducroux C, Khoury N, Lecler A, Blanc R, Chetrit A, Redjem H, et al. Application of the DAWN clinical imaging mismatch and DEFUSE 3 selection criteria: benefit seems similar but restrictive volume cut-offs might omit potential responders. *Eur J Neurol*. (2018) 25:1093–9. doi: 10.1111/ene.13660
25. Lansberg MG, Straka M, Kemp S, Mlynash M, Wechsler LR, Jovin TG, et al. MRI profile and response to endovascular reperfusion after stroke (DEFUSE 2): a prospective cohort study. *Lancet Neurol*. (2012) 11:860–7. doi: 10.1016/S1474-4422(12)70203-X
26. Ostergaard L, Sorensen AG, Kwong KK, Weisskoff RM, Gyldensted C, Rosen BR. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part II: experimental comparison and preliminary results. *Magn Reson Med*. (1996) 36:726–36. doi: 10.1002/mrm.1910360511
27. Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: mathematical approach and statistical analysis. *Magn Reson Med*. (1996) 36:715–25. doi: 10.1002/mrm.1910360510
28. Wu O, Østergaard L, Weisskoff RM, Benner T, Rosen BR, Sorensen AG. Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix. *Magn Reson Med*. (2003) 50:164–74. doi: 10.1002/mrm.10522
29. Calamante F, Gadian DG, Connelly A. Quantification of bolus-tracking MRI: improved characterization of the tissue residue function using Tikhonov regularization. *Magn Reson Med*. (2003) 50:1237–47. doi: 10.1002/mrm.10643
30. Gall P, Emerich P, Kjølbj BF, Kellner E, Mader I, Kiselev VG. On the design of filters for Fourier and oSVD-based deconvolution in bolus tracking perfusion MRI. *MAGMA*. (2010) 23:187–95. doi: 10.1007/s10334-010-0217-8
31. Ebinger M, Brunecker P, Jungehülsing GJ, Malzahn U, Kunze C, Endres M, et al. Reliable perfusion maps in stroke MRI using arterial input functions derived from distal middle cerebral artery branches. *Stroke*. (2010) 41:95–101. doi: 10.1161/STROKEAHA.109.559807
32. Mouridsen K, Christensen S, Gyldensted L, Ostergaard L. Automatic selection of arterial input function using cluster analysis. *Magn Reson Med*. (2006) 55:524–31. doi: 10.1002/mrm.20759
33. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, et al. Optimizing the dice score and Jaccard index for medical image segmentation. *Theory Pract*. (2019). doi: 10.1007/978-3-030-32245-8_11
34. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. (2015) 15:29. doi: 10.1186/s12880-015-0068-x
35. Zhao G, Yoo Y, Re TJ, Das J, Hesheng W, Kim MM, et al. 3D-2D GAN based brain metastasis synthesis with configurable parameters for fully 3D data augmentation. In: Colliot O, Isgum I, editors. *Medical Imaging 2023: Image Processing*. SPIE. (2023) p. 124640J. doi: 10.1117/12.2654281
36. Yoo Y, Gibson E, Zhao G, Re TJ, Parmar H, Das J, et al. Extended nn U-net for brain metastasis detection and segmentation in contrast-enhanced magnetic resonance imaging with a large multi-institutional data set. *Int J Radiat Oncol Biol Phys*. (2025) 121:241–9. doi: 10.1016/j.ijrobp.2024.07.2318
37. Gosch V, Villringer K, Galinovic I, Ganeshan R, Piper SK, Fiebach JB, et al. Automated acute ischemic stroke lesion delineation based on apparent diffusion coefficient thresholds. *Front Neurol*. (2023) 14:1203241. doi: 10.3389/fneur.2023.1203241
38. Wagner DT, Tilmans L, Peng K, Niedermeier M, Rohl M, Ryan S, et al. Artificial intelligence in neuroradiology: a review of current topics and competition challenges. *Diagnostics (Basel)*. (2023) 13:2670. doi: 10.3390/diagnostics13162670
39. Kim YC, Lee JE, Yu I, Song HN, Baek IY, Seong JK, et al. Evaluation of diffusion lesion volume measurements in acute ischemic stroke using encoder-decoder convolutional network. *Stroke*. (2019) 50:1444–51. doi: 10.1161/STROKEAHA.118.024261
40. Mallon D, Fallon M, Blana E, McNamara C, Menon A, Ip CL, et al. Real-world evaluation of Brainomix e-stroke software. *Stroke Vasc Neurol*. (2023) 9:497–504. doi: 10.1136/svn-2023-002859
41. Woo I, Lee A, Jung SC, Lee H, Kim N, Cho SJ, et al. Fully automatic segmentation of acute ischemic lesions on diffusion-weighted imaging using convolutional neural networks: comparison with conventional algorithms. *Korean J Radiol*. (2019) 20:1275–84. doi: 10.3348/kjr.2018.0615
42. Lock C, Tan NSM, Long IJ, Keong NC. Neuroimaging data repositories and AI-driven healthcare-global aspirations vs. ethical considerations in machine learning models of neurological disease. *Front Artif Intell*. (2023) 6:1286266. doi: 10.3389/frai.2023.1286266
43. Borchert RJ, Azevedo T, Badhwar A, Bernal J, Betts M, Bruffaerts R, et al. Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: a systematic review. *Alzheimers Dement*. (2023) 19:5885–904. doi: 10.1002/alz.13412
44. Boyle AJ, Gaudet VC, Black SE, Vasdev N, Rosa-Neto P, Zukotynski KA. Artificial intelligence for molecular neuroimaging. *Ann Transl Med*. (2021) 9:822. doi: 10.21037/atm-20-6220