

# Performance of Automated ASPECTS Software and Value as a Computer-Aided Detection Tool

J. Lambert, J. Demeestere, B. Dewachter, L. Cockmartin, A. Wouters, R. Symons, L. Boomgaert, L. Vandewalle, L. Scheldeman, P. Demaerel, and R. Lemmens



## ABSTRACT

**BACKGROUND AND PURPOSE:** ASPECTS quantifies early ischemic changes in anterior circulation stroke on NCCT but has interrater variability. We examined the agreement of conventional and automated ASPECTS and studied the value of computer-aided detection.

**MATERIALS AND METHODS:** We retrospectively collected imaging data from consecutive patients with acute ischemic stroke with large-vessel occlusion undergoing thrombectomy. Five raters scored conventional ASPECTS on baseline NCCTs, which were also processed by RAPID software. Conventional and automated ASPECTS were compared with a consensus criterion standard. We determined the agreement over the full ASPECTS range as well as dichotomized, reflecting thrombectomy eligibility according to the guidelines (ASPECTS 0–5 versus 6–10). Raters subsequently scored ASPECTS on the same NCCTs with assistance of the automated ASPECTS outputs, and agreement was obtained.

**RESULTS:** For the total of 175 cases, agreement among raters individually and the criterion standard varied from fair to good (weighted  $\kappa$  = between 0.38 and 0.76) and was moderate (weighted  $\kappa$  = 0.59) for the automated ASPECTS. The agreement of all raters individually versus the criterion standard improved with software assistance, as did the interrater agreement (overall Fleiss  $\kappa$  = 0.15–0.23;  $P < .001$  and .39 to .55;  $P = .01$  for the dichotomized ASPECTS).

**CONCLUSIONS:** Automated ASPECTS had agreement with the criterion standard similar to that of conventional ASPECTS. However, including automated ASPECTS during the evaluation of NCCT in acute stroke improved the agreement with the criterion standard and improved interrater agreement, which could, therefore, result in more uniform scoring in clinical practice.

**ABBREVIATION:** CAD = computer-aided detection and diagnosis


ASPECTS was developed as a method to quantify early ischemic changes in the anterior circulation on NCCT. Low ASPECTS was associated with poor functional outcome and increased rates of symptomatic intracranial hemorrhage in patients with acute ischemic stroke who underwent thrombolysis.<sup>1</sup> ASPECTS was subsequently used to select patients with a higher pretreatment chance of achieving good functional outcome in randomized controlled trials on endovascular stroke

treatment and has been incorporated in the American Heart Association guidelines on the management of acute stroke for the selection of thrombectomy candidates in the early time window.<sup>2</sup> However, reported interrater agreement for ASPECTS varied considerably in previous publications.<sup>3–6</sup> A systematic review on this topic concluded that there may be insufficient precision to use ASPECTS as a treatment decision guide.<sup>7</sup> Automated software or artificial intelligence tools have been suggested in the literature as a possible solution for this problem.<sup>8</sup> Several automated or semiautomated software packages based on artificial intelligence have been developed and validated in acute stroke diagnostics, and studies have shown that automated ASPECTS correlates with outcome in patients with large-vessel occlusions treated with mechanical thrombectomy.<sup>9,10</sup> However, recently, a large study documented only moderate agreement of an automated ASPECTS tool compared with the expert raters. This finding argues against an artificial intelligence–only approach without case-by-case validation of the results by physicians.<sup>11</sup> In this study, we compared the agreement of automated ASPECTS and human raters with different

Received March 6, 2023; accepted after revision June 14.

From the Departments of Radiology (J.L., B.D., L.C., R.S., L. B., P.D.) and Neurology (J.D., L.V., L.S., R.S.), University Hospitals Leuven, Leuven, Belgium; Departments of Imaging and Pathology (J.L., B.D., P.D.), Neuroscience (J.D., A.W., L.V., L.S., R.L.), and Experimental Neurology (J.D., A.W., L.V., L.S., R.L.), Laboratory of Neurobiology, Katholieke Universiteit Leuven, University of Leuven, Leuven, Belgium; and Imelda Hospital (R.S.), Bonheiden, Belgium.

Please address correspondence to Julie Lambert, MD, Department of Radiology, University Hospitals Leuven, Herestraat 49, 3000 Leuven, Belgium; e-mail: julie.lambert@uzleuven.be

 Indicates article with online supplemental data.

<http://dx.doi.org/10.3174/ajnr.A7956>

levels of experience levels with the criterion standard in scoring patients who underwent thrombectomy for anterior circulation acute ischemic stroke. In addition, we assessed the impact of the time interval between symptom onset and imaging on the performance of automated ASPECTS. Finally, we evaluated the benefit of automated software in assisting the raters while scoring ASPECTS, by providing software output in addition to the baseline NCCT. We examined the effect on the performance and agreement regarding the evaluation of these early ischemic changes.

## MATERIALS AND METHODS

### **Patient Selection and Imaging Collection**

We retrospectively collected imaging data of consecutive patients who underwent thrombectomy for acute ischemic stroke at the University Hospitals Leuven (Belgium) between 2015 and 2018, irrespective of the time from symptom onset. For transferred patients, we collected pretreatment imaging performed at the referral hospital. The scan protocol of the main referral centers for NCCT was like the scan protocol at the Leuven University Hospital. More detailed information about acquisition parameters can be found in the Online Supplemental Data. While 3 mm was the standard MPR soft-tissue section thickness for the thrombectomy center and also for both main referral centers, for some cases only 1- or 5-mm slices were available. For each case, the best available axial soft-tissue series was imported from the PACS system to the scoring platform. We did not apply angulation correction, filter, or fixed window settings.

### **ASPECTS Rating**

Raters used the ViewDEX scoring platform (<https://academic.oup.com/rpd/article-abstract/139/1-3/42/1599429?redirectedFrom=fulltext>), which randomized cases, for ASPECT scoring.<sup>12</sup> Raters scored ASPECTS in both hemispheres, blinded to clinical information and follow-up imaging. They were able to view the images in the window levels of their preference. Each of 10 predefined regions of 1 side was scored as normal or abnormal (based on visible blurring of contours and swelling and/or hypodensity of the brain parenchyma) to obtain ASPECT scores.<sup>1</sup> If the rater selected the unaffected hemisphere to quantify ASPECTS, this was documented and another rating was requested for the affected side. The software of the automated ASPECTS was RAPID ASPECTS (iSchemaView). The images were remotely processed by the RAPID server. The cases for which the software rated the unaffected side were reprocessed on the server to obtain the score for the affected hemisphere.

The criterion standard was defined as the consensus rating of 2 experienced neuroradiologists (J.L. and P.D.), who also had access to the automated software. Each neuroradiologist scored the cases separately, and discordant cases were read together to reach consensus. This criterion standard rating was performed on high-definition external monitors, DICOM-calibrated, with appropriate lighting. Five raters individually rated the baseline NCCTs: rater 1 (a radiology resident [B.D.]), rater 2 (a stroke neurologist [J.D.]), rater 3 (a neurology resident [L.V.]), and raters 4 and 5 ([R.S.] and [L.B.], both radiologists with a special interest in neuroradiology).

For the evaluation of the software as a computer-aided detection and diagnosis (CAD) tool, we provided the raters with the same randomized images on the scoring platform several weeks after the initial reading. A file with both of their previous scores and the automated ASPECTS output (with images) was available for each case. Raters were then instructed to re-evaluate the scores of the NCCT with the knowledge of this additional software information and to make changes deemed appropriate.

**Statistical Analysis.** Weighted  $\kappa$  was calculated to assess the agreement between the overall ASPECT scores of individual raters as well as the agreement between the criterion standard ratings versus conventional and automated ASPECTS, respectively.

The interpretation of weighted  $\kappa$  values is identical to that of standard  $\kappa$ ,<sup>13</sup> and the classification as proposed by Landis and Koch was applied.<sup>14</sup> The Hotelling T-squared test was applied to compare weighted  $\kappa$  values, for example, to compare the interrater agreement for the readings with and without automated ASPECTS.<sup>15</sup>

The ASPECT scores were dichotomized according to guideline recommendations into 2 groups with either poor (0–5) or good (6–10) ASPECTS. The Cohen  $\kappa$  was used to calculate the agreement with the criterion standard for dichotomized ASPECTS.<sup>2,6</sup> For the comparison of >2 raters, the Fleiss  $\kappa$  was applied. Additionally, sensitivity and specificity of the dichotomized scores of the raters compared with the criterion standard were calculated. Moreover, we calculated the area under the curve values as a measure of the accuracy and compared the area under the curve values without and with CAD by using the DeLong test.

To investigate the impact of time between symptom onset and imaging acquisition on interrater agreement, we binned patients into 3 time intervals (0–1 hour, 1–2 hours, 2–6 hours). All analyses were performed using SPSS v28.0 (IBM), except for the Hotelling T-squared test and the DeLong test, for which R statistical and computing software (<http://www.r-project.org/>) was used.

## RESULTS

### **Demographics of the Study Cohort**

We collected imaging data for 226 patients, of whom 52 were excluded (15 without an occlusion in the anterior circulation, 21 without NCCT, 11 with insufficient imaging quality, 2 with software-processing errors, and 3 who were double-registered in the database). Baseline characteristics for the 174 remaining patients are presented in Table 1. For 1 patient, we included 2 NCCTs because this person underwent 2 thrombectomies for a large-vessel occlusion, first in 1 hemisphere, followed by a second in the other hemisphere within a 3-day time interval. Time from symptom onset to imaging was known for 173 patients: Forty-six patients (27%) were imaged 0–1 hour after the onset of symptoms, 71 (41%) between 1 and 2 hours, 40 (23%) between 2 and 6 hours, and only 16 (9%) >6 hours after symptom onset.

### **Performance of Conventional ASPECTS Raters and Automated Software versus the Criterion Standard**

The 175 NCCTs were scored by human raters (conventional ASPECTS) and processed by the automated software. An overview of all score ranges is provided in Table 2. The frequency of readers

**Table 1: Patient characteristics**

Characteristics	
No. of patients	174
Age (yr) <sup>a</sup>	72 (33–96)
Female sex (%)	94 (54%)
NIHSS <sup>b</sup>	18 (13–22)
Time from symptom onset to imaging (min) <sup>a</sup>	161 (5–1503)
No. of patients with known time interval from symptom onset to imaging	173
0–1 hour	46 (27%)
1–2 hours	71 (41%)
2–6 hours	40 (23%)
>6 hours	16 (9%)
NCCT acquired at referral hospital (%)	50 (29%)

<sup>a</sup>Data are mean (range).

<sup>b</sup>Data are median (interquartile range).

**Table 2: ASPECTS ratings overview with medians, interquartile ranges, and dichotomized score proportions**

Raters	Median (IQR)	0–5	6–10
Criterion standard	8 (7–10)	14%	86%
RAPID ASPECTS	7 (6–9)	17%	83%
Rater 1	9 (8–9)	7%	93%
Rater 1 with CAD	8 (7–9)	9%	91%
Rater 2	8 (6–10)	17%	83%
Rater 2 with CAD	8 (6–10)	19%	81%
Rater 3	9 (8–10)	8%	92%
Rater 3 with CAD	8 (7–10)	12%	88%
Rater 4	9 (7–10)	11%	89%
Rater 4 with CAD	8 (7–10)	14%	86%
Rater 5	8 (7–10)	11%	89%
Rater 5 with CAD	8 (6–9)	18%	82%

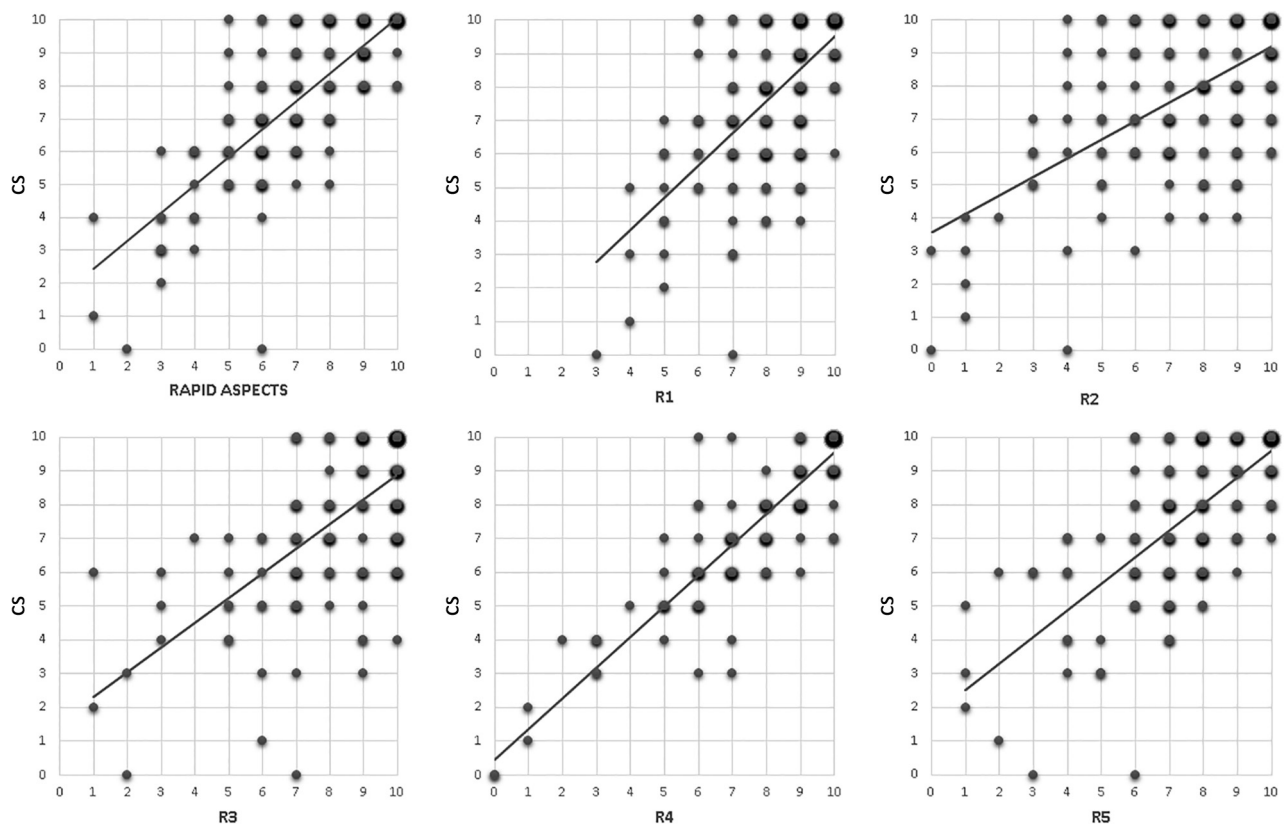
**Note:**—IQR indicates interquartile range.

scoring the unaffected side was low (rater 1: 3%; rater 2: 5%; rater 3: 0%; rater 4: 0%; rater 5: 1%) and similar to that of the software (2%). The intrarater agreement for a subset of 20 cases per reader is presented in the Online Supplemental Data.

We compared the individual rating with the criterion standard consensus scores.

The agreement with the criterion standard of conventional ASPECTS varied substantially among the raters, ranging between fair and good (rater 1: weighted  $\kappa = 0.41$ ; rater 2: weighted

$\kappa = 0.38$ ; rater 3: weighted  $\kappa = 0.42$ ; rater 4: weighted  $\kappa = 0.76$ ; rater 5: weighted  $\kappa = 0.49$ ). Automated analysis resulted in moderate agreement with the criterion standard (weighted  $\kappa = 0.59$ ). Agreement of the dichotomized ASPECTS of the software and raters compared with the criterion standard produced comparable results (automated ASPECTS: Cohen  $\kappa = 0.51$ ; rater 1: Cohen  $\kappa = 0.43$ ; rater 2: Cohen  $\kappa = 0.42$ ; rater 3: Cohen  $\kappa = 0.40$ ; rater 4: Cohen  $\kappa = 0.74$ ; rater 5: Cohen  $\kappa = 0.43$ ). **Figure 1** shows the scatterplots of the 10-point-scale scores of both the automated software versus the criterion standard and the human raters versus the criterion standard. We conclude that for the separate software analysis, the automated scoring did not provide an advantage in scoring precision over the human raters.



**FIG 1.** Scatterplots with trendlines of individual raters and automated software (*x*-axis) versus the criterion standard (*y*-axis). The size of each dot is proportionate to the number of cases with that combination of scores. R indicates rater.

### Impact of Time from Symptom Onset to Imaging on ASPECT Scores

One-hundred fifty-seven patients (90% of the 174) presented within the 0- to 6-hour time window after stroke onset and were selected for this analysis. The median time between onset to imaging in this subgroup was 80 minutes (interquartile range = 55–66 minutes). We divided the patients into three groups according to time intervals: 0–1 hour, 1–2 hours, and 2–6 hours between onset and imaging. The agreement with the criterion

standard improved with time from symptom onset for the automated ASPECTS and 2 of 5 raters. These results are listed in Table 3.

**Table 3: Impact of time from onset to imaging on ASPECTS**

Paired Agreement	0–1 Hour	1–2 Hours	2–6 Hours	P Value
Rater 1 versus criterion standard	0.26	0.41	0.54	.10
Rater 2 versus criterion standard	0.38	0.42	0.32	.77
Rater 3 versus criterion standard	0.27	0.34	0.58	.03
Rater 4 versus criterion standard	0.72	0.68	0.84	.04
Rater 5 versus criterion standard	0.55	0.46	0.50	.71
Automated ASPECTS versus criterion standard	0.37	0.66	0.67	.002

Note:—Data are weighted  $\kappa$  values.

**Table 4: Overall scores agreement with the expert criterion standard for readings without and with CAD**

Overall Scores	Weighted $\kappa$ without CAD	Weighted $\kappa$ with CAD	P Value
Rater 1 versus criterion standard	0.41 (0.04)	0.58 (0.03)	<.001
Rater 2 versus criterion standard	0.38 (0.05)	0.52 (0.05)	.001
Rater 3 versus criterion standard	0.42 (0.04)	0.68 (0.03)	<.001
Rater 4 versus criterion standard	0.76 (0.03)	0.84 (0.02)	<.001
Rater 5 versus criterion standard	0.49 (0.04)	0.57 (0.04)	.04

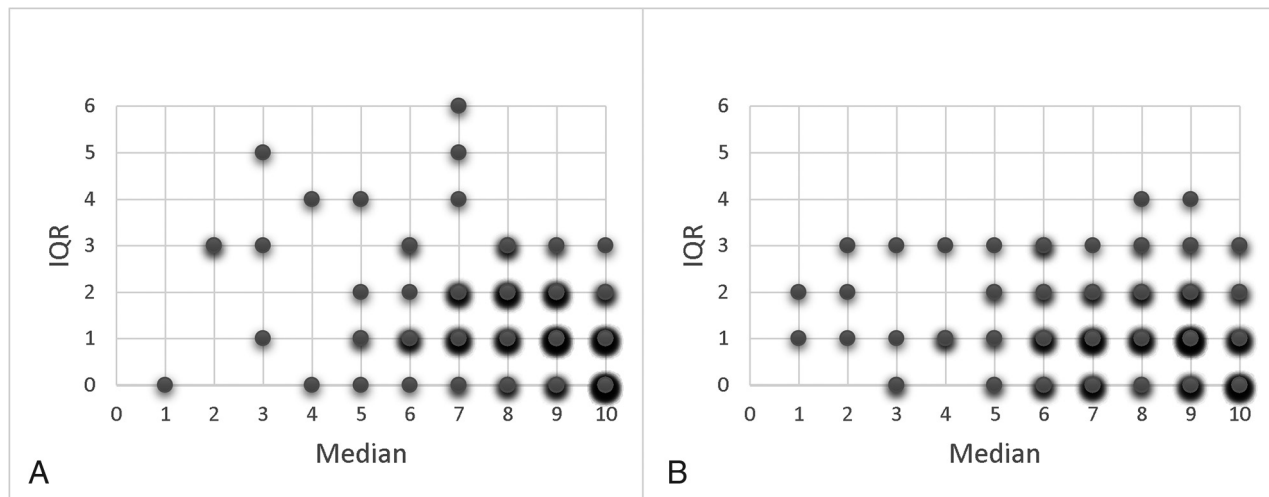
Note:—Standard errors of the  $\kappa$  coefficients are in parentheses.

**Table 5: Dichotomized scores agreement with the criterion standard for readings without and with CAD**

Dichotomized Scores	Cohen $\kappa$ without CAD	Cohen $\kappa$ with CAD	P Value
Rater 1 versus criterion standard	0.43 (0.11)	0.61 (0.09)	.12
Rater 2 versus criterion standard	0.42 (0.09)	0.63 (0.08)	.02
Rater 3 versus criterion standard	0.40 (0.10)	0.55 (0.09)	.19
Rater 4 versus criterion standard	0.74 (0.08)	0.86 (0.06)	.07
Rater 5 versus criterion standard	0.43 (0.10)	0.58 (0.09)	.14

Note:—Standard errors of the  $\kappa$  coefficients are in parentheses.

Automated ASPECTS for Computer-Aided Detection. We identified an improvement in agreement with the criterion standard for the conventional ASPECTS ratings with the use of automated ASPECTS output as CAD. The Online Supplemental Data show the scatterplots of all raters without and with CAD, compared with the criterion standard. For rater 1, the weighted  $\kappa$  increased from 0.41 to 0.58; for rater 2, from 0.38 to 0.52; for rater 3, from 0.42 to 0.68; for rater 4, from 0.76 to 0.84; and for rater 5, from 0.49 to 0.57 (Tables 4 and 5). For dichotomized ASPECTS, we could not identify this significantly improved agreement of the CAD ratings for most raters, though an increasing trend of the Cohen  $\kappa$  values was observed for all raters (0.43–0.61,  $P = .12$  for rater 1; 0.42–0.63,  $P = .02$  for rater 2; 0.40–0.55,  $P = .19$  for rater 3; 0.74–0.86,  $P = .07$  for rater 4; and 0.43–0.58,  $P = .14$  for rater 5; Tables 4 and 5). The diagnostic performance of the dichotomized readings for each reader is provided in the Online Supplemental Data. Overall, the accuracy (area under the curve) improved with the assistance of the automated software from 0.72 to 0.82 ( $P = .004$ ). Also, the overall Fleiss  $\kappa$  for the interrater agreement of all 5 raters improved with CAD from 0.15 to 0.23 ( $P < .001$ ). We identified the same increase for the overall Fleiss  $\kappa$  of the binary scores with CAD (from 0.39 to



**FIG 2.** Score distribution plots illustrating the improvement of the interrater agreement of the 5 raters without (A) and with (B) CAD. The intra-class correlation coefficient of the raters was 0.55 (95% CI, 0.48–0.62) without CAD compared with 0.73 (95% CI, 0.68–0.78) with CAD. IQR indicates interquartile range.

0.55;  $P = .01$ ). The reduction in variability for the ASPECTS rating with CAD is illustrated in Fig 2.

## DISCUSSION

We compared the performance of automated ASPECTS software with human readers and identified moderate agreement for most visual assessments and automated ASPECTS compared with the criterion standard. Providing readers with the automated software such as CAD improved the agreement with the criterion standard and accuracy. Most important, the interrater agreement increased with the assistance of the output from the software, suggesting enhanced uniformity in reading ASPECTS by CAD.

The performance of the automated software in comparison with the consensus score is in line with previous studies revealing similar, moderate agreement in conventional assessment of baseline NCCT by raters and automated software analysis compared with experts.<sup>11,16-18</sup> Some studies analyzing results for the individual regions did identify differences in agreement depending on the ASPECTS region.<sup>19,20</sup> Two previous validation studies on RAPID ASPECTS software showed an outperformance of the software versus the human raters.<sup>21,22</sup> A first study found RAPID ASPECTS to be more accurate than experienced raters in patients with large hemispheric infarcts compared with a diffusion-weighted imaging ASPECTS, which, of course, differs from our methodology.<sup>22</sup>

A second study showed near-perfect agreement with a criterion standard.<sup>21</sup> Agreement with the criterion standard in our study was lower, possibly due to methodologic differences related to the definition of the criterion standard. First, the criterion standard expert raters in our study were exposed to only baseline NCCT data, while in the other study, follow-up MR imaging, which revealed the final infarct location, was included. Also, the NCCT acquisition and reconstructions sometimes slightly differed among cases in our data (as a reflection of imaging included from multiple centers), while this was standardized in the other study. Section thickness possibly affects human raters and automated ASPECTS differently. Thinner slices can improve the performance of automated analyses resulting from the higher resolution, but these could also negatively affect human performance because the resolution will differ from imaging data evaluated in daily clinical practice.<sup>23</sup> Surprisingly, while most hospitals are evolving to thinner slices, a recent study on the validation of 2 automated ASPECTS software programs found 5 mm to be the optimal section thickness.<sup>18,24</sup>

Several publications reported on the interrater variability of ASPECTS.<sup>3-6</sup> A study with 100 raters found high variability among even experienced radiologists and neuroradiologists.<sup>25</sup> In many hospitals, stroke care is provided by less experienced raters (particularly in out-of-office hours). Therefore, we studied the potential role of automated software to assist readers in improving agreement with the criterion standard but also accuracy and interrater agreement because these may be of great benefit.

Our results revealing increased agreement with a consensus criterion standard in readers assisted by the automated ASPECTS have been reported previously.<sup>26</sup> This information may not be extrapolated to all available packages because suboptimal software

tools may even negatively influence readers, thereby decreasing their performance.<sup>27</sup>

To our knowledge, previous studies did not report on the effect of automated ASPECTS on overall interrater agreement. We hypothesize that part of the reason for the amelioration of the interrater agreement reported here is due to the visualization of the automated ASPECTS on 2 standardized slices with uniform angulation, filter, and windowing. Presenting this uniform set of selected images by the software might counteract part of the image variability that ASPECTS raters encounter when evaluating NCCTs.

In clinical management of acute ischemic stroke, agreement on a dichotomized ASPECTS (good versus poor) may be more relevant because it is used to assess eligibility for thrombectomy.<sup>28</sup> In this study, we defined ASPECTS as good versus poor on the basis of the current guidelines. Although one might assume the interrater variability to be less in the dichotomized approach, the individual  $\kappa$  values were rather similar for the weighted and Cohen  $\kappa$ . We assume the imbalance of our data set; the more normal ASPECTS areas than ischemic areas as well as the low proportion of ASPECTS scores of  $<6$  can partially explain this. Considering the recently published trials in large-core patients, a similar analysis would be of interest after dichotomizing ASPECTS in  $<3$  compared with  $\geq 3$ .<sup>29,30</sup> Because our population was skewed toward higher ASPECTS, we could, unfortunately, not perform this dichotomized low ASPECTS analysis in this cohort.

Our study has several limitations. First, we included only patients who underwent thrombectomy. Although in our clinical center, we typically do not exclude patients from thrombectomy presenting in the 0- to 6-hour time window solely on the basis of poor ASPECTS, it is possible that patients with low ASPECTS are underrepresented in this cohort, possibly affecting interrater agreement. Second, we wanted the experienced raters to have access to the automated ASPECTS output for the consensus criterion standard rating because they could have all imaging information. Potentially, this choice could have introduced bias toward the automated software results, but we assumed experienced readers to be less prone to this bias. Third, the sample size may have been too small and homogeneous in relation to the time between onset to imaging (patients imaged beyond 6 hours after symptom onset were underrepresented) to robustly study the role of time from onset to imaging on the performance of human and automated ratings. We believe that future studies should focus on including patients in the later time window to study the effect of the time between onset and imaging on the agreement of ASPECTS reading.

Another limitation of our study is that we evaluated only 1 vendor software type. A different automated ASPECTS software program (Brainomix; <https://www.brainomix.com/>) was shown to be noninferior to neuroradiologists in several studies.<sup>16,31,32</sup> A comparison between Brainomix and Frontier software found higher agreement of the raters with Brainomix than with Frontier.<sup>17</sup> A study comparing 3 different automated ASPECTS tools reported on a convincing grade of agreement among them, underscoring the potential of all 3 for decision support.<sup>18</sup>

The current discussion on introducing artificial intelligence in clinical practice is often directed at evaluating software as a replacement for human raters. Alternatively, we could focus on the software as a decision-aiding tool in clinical practice. In clinical scenarios, frequently a patient with presumed stroke is not immediately assessed by a criterion standard expert. Therefore, our study with raters of different backgrounds, all working in the acute stroke flow, is clinically relevant. The introduction of artificial intelligence for ASPECTS could benefit many patients because it can potentially support raters with lower levels of expertise. Because even intrarater agreement of ASPECTS is known to be low,<sup>7</sup> the advantage of the software (if, of course, accurate) is providing reproducible assistance. ASPECTS with assistance of artificial intelligence has the potential to improve the accuracy of various raters over all hospital settings, especially in the absence of the expert raters. The results of our study are suggestive of such a benefit, though its potential needs to be confirmed on a larger scale by future studies. These studies could focus on reading ASPECTS in patients presenting in later time windows beyond 6 hours to assess the performance of the automated software and the value of the output in assisting readers in scoring ASPECTS. In addition, these studies could include physicians from various disciplines involved in acute stroke care with differences in expertise to validate the improvement in performance and agreement.

## CONCLUSIONS

We determined automated ASPECTS and human raters to have similar agreement compared with the criterion standard. Using the automated ASPECTS output as a CAD tool improved the agreement with the criterion standard, accuracy, and interrater agreement. Our findings suggest that the application of this automated analysis as an assistance tool for reading NCCTs in patients with acute ischemic stroke will result in more uniform and accurate scoring of ASPECTS.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

- Barber PA, Demchuk AM, Zhang J, et al. **Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy.** *Lancet* 2000;355:1670–74 [CrossRef Medline](#)
- Powers WJ, Rabinstein AA, Ackerson T, et al. **Guidelines for the Early Management of Patients with Acute Ischemic Stroke: 2019 Update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke—A Guideline For Healthcare Professionals from the American Heart Association/American Stroke Association.** *Stroke* 2019;50:E344–418 [CrossRef Medline](#)
- Coutts SB, Demchuk AM, Barber PA, et al; VISION Study Group. **Interobserver variation of ASPECTS in real time.** *Stroke* 2004;35:e103–05 [CrossRef Medline](#)
- Gupta AC, Schaefer PW, Chaudhry ZA, et al. **Interobserver reliability of baseline noncontrast CT Alberta Stroke Program Early CT score for intra-arterial stroke treatment selection.** *AJNR Am J Neuroradiol* 2012;33:1046–49 [CrossRef Medline](#)
- McTaggart RA, Jovin TG, Lansberg MG, et al; DEFUSE 2 Investigators. **Alberta stroke program early computed tomographic scoring performance in a series of patients undergoing computed tomography and MRI: reader agreement, modality agreement, and outcome prediction.** *Stroke* 2015;46:407–12 [CrossRef Medline](#)
- Nicholson P, Hilditch CA, Neuhaus A, et al. **Per-region interobserver agreement of Alberta Stroke Program Early CT Scores (ASPECTS).** *J Neurointerv Surg* 2020;12:1069–71 [CrossRef Medline](#)
- Farzin B, Fahed R, Guilbert F, et al. **Early CT changes in patients admitted for thrombectomy: intrarater and interrater agreement.** *Neurology* 2016;87:249–56 [CrossRef Medline](#)
- Wilson AT, Dey S, Evans JW, et al. **Minds treating brains: understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke.** *Expert Rev Cardiovasc Ther* 2018;16:143–45 [PMC\] CrossRef Medline](#)
- Pfaff J, Herweh C, Schieber S, et al. **E-ASPECTS correlates with and is predictive of outcome after mechanical thrombectomy.** *AJNR Am J Neuroradiol* 2017;38:1594–99 [CrossRef Medline](#)
- Olive-Gadea M, Martins N, Boned S, et al. **Baseline ASPECTS and e-ASPECTS correlation with infarct volume and functional outcome in patients undergoing mechanical thrombectomy.** *J Neuroimaging* 2019;29:198–202 [CrossRef Medline](#)
- Mair G, White P, Bath PM, et al. **External validation of e-ASPECTS software for interpreting brain CT in stroke.** *Ann Neurol* 2022;92:943–57 [CrossRef Medline](#)
- Håkansson M, Svensson S, Zachrisson S, et al. **ViewDEX: an efficient and easy-to-use software for observer performance studies.** *Radiat Prot Dosimetry* 2010;139:42–51 [CrossRef Medline](#)
- McHugh ML. **Interrater reliability: the  $\kappa$  statistic.** *Biochem Med (Zagreb)* 2012;22:276–82 [Medline](#)
- Landis JR, Koch GG. **The measurement of observer agreement for categorical data.** *Biometrics* 1977;33:159–74 [Medline](#)
- Vanbelle S. **Comparing dependent  $\kappa$  coefficients obtained on multilevel data.** *Biom J* 2017;59:1016–34 [CrossRef Medline](#)
- Sundaram VK, Goldstein J, Wheelwright D, et al. **Automated ASPECTS in acute ischemic stroke: a comparative analysis with CT perfusion.** *AJNR Am J Neuroradiol* 2019;40:2033–38 [CrossRef Medline](#)
- Goebel J, Stenzel E, Guberina N, et al. **Automated ASPECT rating: comparison between the Frontier ASPECT Score software and the Brainomix software.** *Neuroradiology* 2018;60:1267–72 [CrossRef Medline](#)
- Hoelter P, Muehlen I, Goelitz P, et al. **Automated ASPECT scoring in acute ischemic stroke: comparison of three software tools.** *Neuroradiology* 2020;62:1231–33 [CrossRef Medline](#)
- Neuhaus A, Seyedsaadat SM, Mihal D, et al. **Region-specific agreement in ASPECTS estimation between neuroradiologists and e-ASPECTS software.** *J Neurointerv Surg* 2020;12:720–23 [CrossRef Medline](#)
- Austein F, Wodarg F, Jürgensen N, et al. **Automated versus manual imaging assessment of early ischemic changes in acute stroke: comparison of two software packages and expert consensus.** *Eur Radiol* 2019;29:6285–92 [CrossRef Medline](#)
- Maegerlein C, Lehm M, Friedrich B, et al. **Automated calculation of the Alberta Stroke Program Early CT Score: feasibility and reliability.** *Radiology* 2019;291:141–48 [CrossRef Medline](#)
- Albers GW, Wald MJ, Mlynash M, et al. **Automated calculation of Alberta Stroke Program Early CT Score: validation in patients with large hemispheric infarct.** *Stroke* 2019;50:3277–79 [CrossRef Medline](#)
- Löffler MT, Sollmann N, Mönch S, et al. **Improved reliability of automated ASPECTS evaluation using iterative model reconstruction from head CT scans.** *J Neuroimaging* 2021;31:341–44 [CrossRef Medline](#)
- Chen Z, Shi Z, Lu F, et al. **Validation of two automated ASPECTS software on non-contrast computed tomography scans of patients with acute ischemic stroke.** *Front Neurol* 2023;14 [CrossRef Medline](#)
- van Horn N, Kniep H, Broocks G, et al. **ASPECTS interobserver agreement of 100 investigators from the TENSION study.** *Clin Neuroradiol* 2021;31:1093–100 [CrossRef Medline](#)

26. Delio PR, Wong ML, Tsai JP, et al. **Assistance from automated ASPECTS software improves reader performance.** *J Stroke Cerebrovasc Dis* 2021;30:105829 [CrossRef Medline](#)
27. Ernst M, Bernhardt M, Bechstein M, et al. **Effect of CAD on performance in ASPECTS reading.** *Inform Med Unlocked* 2020;18:100295 [CrossRef](#)
28. Cagnazzo F, Derraz I, Dargazanli C, et al. **Mechanical thrombectomy in patients with acute ischemic stroke and ASPECTS  $\leq$ 6: a meta-analysis.** *J Neurointerv Surg* 2020;12:350–55 [CrossRef Medline](#)
29. Huo X, Ma G, Tong X, et al; ANGEL-ASPECT Investigators. **Trial of Endovascular Therapy for Acute Ischemic Stroke with Large Infarct.** *N Engl J Med* 2023;388:1272–83 [CrossRef Medline](#)
30. Sarraj A, Hassan AE, Abraham MG, et al; SELECT2 Investigators. **Trial of Endovascular Thrombectomy for Large Ischemic Strokes.** *N Engl J Med* 2023;388:1259–71 [CrossRef Medline](#)
31. Herweh C, Ringleb PA, Rauch G, et al. **Performance of e-ASPECTS software in comparison to that of stroke physicians on assessing CT scans of acute ischemic stroke patients.** *Int J Stroke* 2016;11:438–45 [CrossRef Medline](#)
32. Nagel S, Sinha D, Day D, et al. **e-ASPECTS software is non-inferior to neuroradiologists in applying the ASPECT score to computed tomography scans of acute ischemic stroke patients.** *Int J Stroke* 2017;12:615–22 [CrossRef Medline](#)