

# **CT reconstruction levels affect automated and reader-based ASPECTS ratings in acute ischemic stroke**

Fatih Seker, Johannes Pfaff, Simon Nagel, Dominik Vollherbst, Stephen Gerry, Markus A. Möhlenbruch, Martin Bendszus, Christian Herweh

From the Department of Neuroradiology at Heidelberg University Hospital, Heidelberg, Germany (F.S., J.P., D.V., M.M., M.B., C.H.), Department of Neurology at Heidelberg University Hospital, Heidelberg, Germany (S.N.), Centre for Statistics in Medicine, University of Oxford, Oxford, UK (S.G.).

Running title: Automated and human ASPECTS with different CT reconstructions

## **Keywords:**

Stroke, ASPECTS, CT, iterative reconstruction, filtered back projection

Corresponding author

PD Dr. Christian Herweh

Im Neuenheimer Feld 400

69120 Heidelberg

GERMANY

Phone: +49 6221 56 7566

Fax: +49 6221 56 4673

E-Mail: [christian.herweh@med.uni-heidelberg.de](mailto:christian.herweh@med.uni-heidelberg.de)

## **Acknowledgements & Disclosures**

Dr Nagel received grants, personal fees, and other from Brainomix, personal fees and other from Bayer, personal fees and other from Böhlinger Ingelheim, personal fees and other from Medtronic, personal fees from Pfizer, and outside submitted work. Dr Pfaff has received payment for lectures from University of Hamburg-Eppendorf and Asklepios Ärzteakademie, travel and meeting expenses from Stryker, and MicroVention Deutschland. Dr Bendszus has received grants and personal fees from Novartis, Guerbet, and Codman, personal fees from Vascular Dynamics, Roche, Teva, Springer, and Bayer Vital, grants from Siemens, Hopp Foundation, and DFG. Dr Möhlenbruch has received personal fees from Codman, MicroVention, Phenox, and Stryker. Dr. Herweh received personal fees from Brainomix. The other authors report no conflicts.

## Abstract

**Background and Purpose:** We investigated whether automated and reader based ASPECTS in acute stroke patients are affected by different CT image reconstruction algorithms.

**Methods:** ASPECTS were assessed by commercial software and four independent blinded readers (2 residents and 2 consultants) from different CT reconstructions (filtered back projection [FBP] and 2 different iterative reconstruction [IR] levels) in 43 acute stroke patients with proximal middle cerebral artery occlusion. Ground truth was provided by an expert with unrestricted data access.

**Results:** The residents showed significant variations between IR levels and had a significantly lower internal consistency across different reconstructions compared to the software, which performed similarly to the consultants. The consultant as well as the software also showed different deviations from ground truth with different IR levels, which were least at IR strength level 2.

**Conclusions:** CT image post-processing affects either automated or human ASPECTS in acute stroke patients. This effect was most pronounced in the less experienced readers, while the software had the most robust performance.

## Introduction

Computer-assisted image analysis based on deep learning has been introduced in clinical radiology recently and is becoming employed increasingly<sup>1</sup>. Besides the possibility to quickly process large amounts of data, another compelling task for such applications is to increase reliability where it is rather poor with human readers. A typical example here is early ischemia signs (EIS) in cerebral stroke which can be challenging, especially to less experienced physician. The Alberta Stroke Program Early CT Score (ASPECTS) was developed to assess EIS more objectively by assigning a numerical and therefore semi-quantitatively score<sup>2</sup>. ASPECTS has been used in recent stroke trials which have proven the benefit of mechanical thrombectomy and is currently recommended as imaging criterion for mechanical thrombectomy (MT) within 6 hours from symptom onset by international guidelines (ESO)<sup>3</sup>. At the same time, interrater reliability in ASPECTS is known to be moderate, only<sup>4</sup>.

e-ASPECTS (Brainomix Ltd, Summertown, Oxford, UK) is a commercial software which automatically determines ASPECTS and has been shown to be non-inferior to stroke experts<sup>5</sup>.

High image quality is a prerequisite for the reliable assessment of EIS and ASPECTS. Various reconstruction algorithms have been established to increase image quality in CT imaging and have already been shown to influence infarct detection<sup>6</sup>. In the present study, we tried to investigate whether different CT reconstructions would influence automated ASPECTS in comparison to those of human readers.

## Methods

### Patient selection

Between 11/2014 and 5/2015, patients with acute occlusion of the M1 segment of the middle cerebral artery and CT and CT angiography as primary diagnostic modality were selected. Our institutional review board approved this study.

### Image acquisition

CT scans were acquired using a 20-row multislice CT (SOMATOM Definition AS, Siemens, Erlangen, Germany). All patients underwent a dedicated multimodal CT protocol (NECT, CTA & CTP) and had CTA-proven proximal MCA occlusion. NECT scans were acquired at 120 kV in X-care technique (Siemens) adjusting the tube current automatically in order to reduce radiation dose. Mean CT Dose Index for NECT was  $53.2 \pm 3.1$  mGy at  $329 \pm 20$  mAs.

NECT scans were reconstructed in axial planes with a slice thickness of 4 mm in FBP (Kernel H41s) and Sinogram-Affirmed Iterative Reconstruction (SAFIRE) in strength levels 2 and 4 out of 5 (Kernel J40s). In total, 3 reconstructions of each NECT per patient were made.

### Image analysis

ASPECTS was assessed automatically by e-ASPECTS software and by four blinded readers independently: a third year and a fourth year radiology resident (DV & FS, respectively) as well as a radiology and a neurology consultant with 5 and 8 years of experience in diagnosis and treatment of stroke (JP & SN, respectively). An unblinded neuroradiology consultant (C.H.) with 15 years of experience in stroke imaging determined the ground truth ASPECTS based on clinical information, baseline multimodal CT and CT perfusion imaging and follow up scans.

## Statistical analysis

Statistical analysis was performed using R version 3.2.2. Agreement of ASPECTS rating with ground truth ASPECTS was assessed using weighted kappa analysis and described as follows<sup>7</sup>: poor agreement ( $\kappa < 0.20$ ), fair agreement ( $0.20 \leq \kappa \leq 0.40$ ), moderate agreement ( $0.40 < \kappa \leq 0.60$ ), good agreement ( $0.60 \leq \kappa \leq 0.80$ ), very good agreement ( $0.80 \leq \kappa \leq 1.00$ ). Bootstrap resampling was used to estimate p-values for the comparisons of Kappa scores. Intraclass correlation coefficients were calculated to assess conformity for particular raters across different reconstruction levels.

## Results

Forty-three patients (mean age  $\pm$  standard deviation,  $73.7 \pm 12.0$  years; 46.5% female) with acute occlusion of the M1 segment of the middle cerebral artery were included in this study. Mean time from stroke onset to CT imaging  $\pm$  standard deviation was  $108 \pm 78$  minutes. In 12 patients, stroke onset time was unknown. In these cases, mean time from last seen well to CT imaging was  $454 \pm 274$  minutes. Median National Institutes of Health Stroke Scale score was 16 (interquartile range [IQR] 11 – 19). Median ground truth ASPECTS was not different between patients with known (median 8; interquartile range; IQR 6.5 – 9) and patients with unknown onset (median 8; IQR 7 - 9.25), respectively ( $p = 0.49$ ; Mann-Whitney). In total, 129 NECT reconstructions were analyzed, please see table 1 for detailed results. Except for consultant 2 all readers performed better with IR compared to FBP, only resident 1 performed better with IR2 than 4, while resident 2 and consultant 1 performed similarly with both IR levels. eASPECTS performed similarly to the consultant with the better results with each respective level. The both consultants were more consistent across different levels compared to the residents in terms of intra-rater agreement, while eASPECTS showed the highest consistency which also was significantly higher than that of the residents (0.92 vs 0.79 & 0.73, respectively;  $p=0.03$  & 0.01, respectively).

## Discussion

In this study, CT reconstruction algorithms (FBP, IR 2 and IR 4) had an influence on all raters including e-ASPECTS software. Overall, IR 2 allowed best recognition of early ischemic changes. Especially the residents showed a significant improvement from FBP ( $\kappa = 0.52$ ) to IR 2 ( $\kappa = 0.69$ ), while the consultant and e-ASPECTS showed little improvement. All the raters showed similar agreement with ground truth when using IR 2. This is of note, since interrater reliability in ASPECTS has been reported to be moderate<sup>4</sup>. IR may, therefore, help reducing disagreements in trials and in clinical routine. Although the software's reliability with regard to the ground truth also varied with different CT reconstructions, it had the highest internal consistency which also differed significantly from that of both residents. Thus, its performance was least affected by different image qualities. This is important in the clinical context especially in referral situations where different CT scanning qualities can be challenging in making diagnosis and treatment decisions.

In recent years, several studies have demonstrated that IR improves image quality of spiral CT of the head<sup>8</sup>. Bodelle et al. have shown that evident infarct demarcation in NECT is more conspicuous when using IR, especially in strength level 4<sup>6</sup>. This study, in contrast to the present one, used cases with fully demarcated infarcts, only. In the acute phase, i.e. within the first hours from symptom onset, detection of early signs of infarction is much more challenging but can be truly beneficial. Therefore, we have included cases in which the ASPECTS was clinically relevant with regard to treatment decisions.

Our study has several limitations. This study as well as previous ones included only a moderate number of patients. We tried to compensate for that by establishing an expert-based ground truth for comparison of reader-based and automated results.

Whereas simultaneous diffusion weighted MRI could provide the most exact ground truth in case of acute stroke<sup>9</sup>, this was not feasible in the present study for the sake of time since all patients were considered eligible for emergency endovascular therapy. Also NECT scans were acquired using a single CT scanner by one particular vendor and its IR technique at one institution. Therefore, the transferability to other manufacturers might be impaired.



## References

1. Zaharchuk G, Gong E, Wintermark M, Rubin D, Langlotz CP. Deep Learning in Neuroradiology. *AJNR Am J Neuroradiol* 2018 [Epub ahead of print].
2. Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. Alberta Stroke Programme Early CT Score. *Lancet* 2000;355:1670-4.
3. Wahlgren N, Moreira T, Michel P, et al. Mechanical thrombectomy in acute ischemic stroke: Consensus statement by ESO-Karolinska Stroke Update 2014/2015, supported by ESO, ESMINT, ESNR and EAN. *Int J Stroke* 2016;11:134-47.
4. McTaggart RA, Jovin TG, Lansberg MG, et al. 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.
5. 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.
6. Bodelle B, Wichmann JL, Scholtz JE, et al. Iterative Reconstruction Leads to Increased Subjective and Objective Image Quality in Cranial CT in Patients With Stroke. *AJR Am J Roentgenol* 2015;205:618-22.
7. Fleiss JL, Cohen J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educ Psychol Meas* 1973;33:613-19.

8. Wenz H, Maros ME, Meyer M, et al. Image Quality of 3rd Generation Spiral Cranial Dual-Source CT in Combination with an Advanced Model Iterative Reconstruction Technique: A Prospective Intra-Individual Comparison Study to Standard Sequential Cranial CT Using Identical Radiation Dose. PLoS One 2015;10:e0136054.
9. 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.

Table 1: Correlation of ASPECTS ratings with ground truth ASPECTS

	Resident1	Resident2	Consultant1	Consultant2	eASPECTS
FBP	0.58 (0.30- 0.82)	0.36 (0.17- 0.65)	0.68 (0.37- 0.86)	0.66 (0.38- 0.83)	0.70 (0.46- 0.86)
IR 2	0.68 (0.43- 0.86)	0.50 (0.29- 0.70)	0.72 (0.46- 0.86)	0.55 (0.28- 0.77)	0.71 (0.49- 0.85)
IR 4	0.50 (0.25- 0.68)	0.54 (0.31- 0.71)	0.72 (0.47- 0.87)	0.56 (0.28- 0.74)	0.62 (0.36- 0.82)
ICC	0.79* (0.66-0.88)	0.73# (0.60-0.84)	0.84 (0.67-0.93)	0.81 (0.71-0.9)	0.92 (0.81-0.97)

Weighted Kappa correlation coefficients as well as intraclass correlation coefficients (ICC; bottom line) and 95% confidence intervals, respectively. Note that ICC values of Resident 1 & 2 differ significantly from that of eASPECTS ( $p=0.03$  &  $0.01$ , respectively). FBP: filtered back projection. IR 2 and IR 4: iterative Reconstructions at strength levels 2 and 4 (out of 5).