

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Automated brain computed tomographic densitometry of early ischemic changes in acute stroke

Berend C. Stoel
Henk A. Marquering
Marius Staring
Ludo F. Beenen
Cornelis H. Slump
Yvo B. Roos
Charles B. Majoie

Automated brain computed tomographic densitometry of early ischemic changes in acute stroke

Berend C. Stoel,^{a,*} Henk A. Marquering,^{b,c} Marius Staring,^a Ludo F. Beenen,^b Cornelis H. Slump,^d Yvo B. Roos,^e and Charles B. Majoie^b

^aLeiden University Medical Center, Division of Image Processing, Department of Radiology, Albinusdreef 2 Leiden 2333 AA, The Netherlands

^bAcademic Medical Center, Department of Radiology, Amsterdam, The Netherlands

^cAcademic Medical Center, Department of Biomedical Engineering and Physics, Amsterdam, The Netherlands

^dUniversity of Twente, MIRA Institute for Biomedical Technology and Technical Medicine, Enschede, The Netherlands

^eAcademic Medical Center, Department of Neurology, Amsterdam, The Netherlands

Abstract. The Alberta Stroke Program Early CT score (ASPECTS) scoring method is frequently used for quantifying early ischemic changes (EICs) in patients with acute ischemic stroke in clinical studies. Varying interobserver agreement has been reported, however, with limited agreement. Therefore, our goal was to develop and evaluate an automated brain densitometric method. It divides CT scans of the brain into ASPECTS regions using atlas-based segmentation. EICs are quantified by comparing the brain density between contralateral sides. This method was optimized and validated using CT data from 10 and 63 patients, respectively. The automated method was validated against manual ASPECTS, stroke severity at baseline and clinical outcome after 7 to 10 days (NIH Stroke Scale, NIHSS) and 3 months (modified Rankin Scale). Manual and automated ASPECTS showed similar and statistically significant correlations with baseline NIHSS ($R = -0.399$ and -0.277 , respectively) and with follow-up mRS ($R = -0.256$ and -0.272), except for the follow-up NIHSS. Agreement between automated and consensus ASPECTS reading was similar to the interobserver agreement of manual ASPECTS (differences <1 point in 73% of cases). The automated ASPECTS method could, therefore, be used as a supplementary tool to assist manual scoring. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JMI.2.1.014004]

Keywords: stroke; ASPECTS scoring; computed tomography; image processing; densitometry.

Paper 14092R received Jul. 14, 2014; accepted for publication Mar. 3, 2015; published online Mar. 24, 2015.

1 Introduction

Approximately 70% of acute strokes are caused by hypoxic (ischemic) brain damage as a result of severe cerebral blood circulation failure.¹ The reduced blood flow can lead to infarction of the brain tissue, which can be detected on noncontrast computed tomography (NCCT).² Ischemic tissue on NCCT is characterized by changes in brain parenchyma that reflect either decreased x-ray attenuation (e.g., loss of definition of the lentiform nucleus) or tissue swelling (e.g., hemispheric sulcal effacement and effacement of the lateral ventricle).³ These changes in brain parenchyma in the acute time window within 4.5 h after symptom onset, i.e., early ischemic changes (EICs), correlate with stroke severity and clinical outcome.^{4–6}

To visually identify and quantify subtle EICs in the middle cerebral artery (MCA) territory, the Alberta Stroke Program Early CT score (ASPECTS) scoring method has been introduced,⁶ standardizing the detection and reporting of the extent of ischemic hypodensity. The ASPECTS value is calculated for 10 regions of interest at two standardized levels of the MCA territory, one including the basal ganglia and the cortical MCA areas at the level of the basal ganglia and below (M1 to M3, lentiform and caudate nuclei and internal capsule) and one

above these structures, the supraganglionic level (M4 to M6). An unaffected MCA territory is assigned a total score of 10. For each area with signs of EIC, i.e., parenchymal hypodensity, one point is subtracted from that initial score.

The reliability of visual EICs detection on NCCT is influenced by several factors, including the time window between CT scanning and symptom onset, the vascular territory of the infarct, the experience of the CT reader,⁷ availability of clinical history,⁸ viewing window width, and level settings.^{9,10} Even among experienced clinicians, there is considerable lack of agreement in recognizing and quantifying early radiologic signs of acute stroke on CT.^{11–14} Hence, concern has arisen about the reliability of the detection of EICs on CT, while this is the basis for timely treatment, vital for stroke outcome.¹⁵ Therefore, objective and automated methods are needed to recognize and quantify ischemic brain damage in an early stage, not only for daily clinical practice but also for large-scale clinical trials on the treatment of ischemic stroke.

The quantitative nature of CT allows for a quantitative analysis of density changes as a sign of EICs. Since the density of ischemic brain tissue decreases with time after stroke onset,¹⁶ studies have used brain CT densitometry to identify EICs in acute ischemic stroke. Bendszus et al.¹⁷ described a postprocessing method for the detection of acute MCA territory infarcts. By subtraction of density histograms of the entire hemispheres on

*Address all correspondence to: Berend C. Stoel, E-mail: B.C.Stoel@lumc.nl

CT scans, they showed a significant increase in acute infarct detection rate by human observers. Maldjian et al.¹⁸ identified potential areas of acute ischemia on CT scans by comparing density histograms of the lentiform nucleus and insula with histograms of the structures on the contralateral brain side. ASPECTS areas were not used in these studies, which limits their applicability in clinical practice. Furthermore, no data were reported on the relation with clinical outcome.

Kosior et al.¹⁹ computed an automated atlas-based topographical scoring of acute stroke based on the ASPECTS regions based on MRI. They obtained an auto-ASPECTS from the overlap of ischemic lesions detected on apparent diffusion coefficient maps in MRI data, using atlas-based segmentation of the MCA supply territories. They adopted a segmentation method for artery supply territories in the brain²⁰ using the Fast Talairach Transform (FTT).²¹ The FTT method was optimized and validated for MR images, but an extension to CT images is not available.

Therefore, we developed an automated brain CT densitometric method to replicate the manual ASPECTS based on the so-called brain density shift (BDS) between contralateral brain areas, thereby systematically quantifying subtle EICs. We hypothesized that this automated ASPECTS quantification correlates with stroke severity and clinical outcome. We designed and optimized the auto-ASPECTS method using a training set of images. Subsequently, the method was validated against manual ASPECTS determined by consensus, and stroke severity at presentation (National Institutes of Health Stroke Scale, NIHSS) and clinical outcome after 7 to 10 days (NIHSS) and 3 months (modified Rankin Scale, mRS) in a separate test set. The agreement between automated and consensus scoring was compared with the interobserver agreement for the global scoring data and per ASPECTS region.

2 Material and Methods

2.1 Patient Data

As part of a prospective randomized multicenter controlled trial²² on acute ischemic stroke, patients were included retrospectively for this study during the period of July 2008 until April 2011. Out of these 73 patients, we randomly selected a training set of 10 patients to optimize the automated method and this optimal method was subsequently evaluated in a separate population of 63 patients (see Table 1). All patients came from densely populated areas and, therefore, had short times from symptom onset to scan. All image data were anonymized to ensure blind assessments. The study protocol was approved

by the medical ethics committee and all patients or their legal representatives provided written informed consent.

All images were acquired with a 64-slice CT scanner (Sensation 64, Siemens Medical Solutions, Forchheim, Germany) with syngo CT 2007s software, using a tube voltage of 120 kVp and a current of 323 mA with a pitch factor of 0.85 and 28.8 mm collimation. Axial slices were reconstructed with a thickness of 1.0 mm and a 0.64 ± 0.12 mm (mean \pm SD) increment, using a moderately soft kernel (Siemens "H31s") and a 512×512 matrix containing the whole brain. The in-plane resolution was 0.46 ± 0.06 mm (mean \pm SD).

2.2 Reference Standard and Clinical Data

Baseline NCCT scans were independently inspected by two experienced radiologists with more than 10 years of experience (C.B.M. and L.F.B.) for signs of EIC occurring in the ASPECTS regions of interest. Additional findings, such as a high degree of atrophy, old ischemic lesions, or scanning artifacts, were also recorded. EICs were defined as parenchymal x-ray hypoattenuation as indicated by a region of abnormally low attenuation relative to other parts of the same structure or the contralateral hemisphere. The observers were blinded for clinical data except for the symptom side. Viewing settings were adjusted based on standard preset viewing window width and center level of 70/35 and 35/30 Hounsfield unit (HU). After individual scoring, discrepancies between both observers were resolved by consensus reading.

Stroke severity was assessed by NIHSS at presentation by the neurologist on duty and clinical outcome was assessed by NIHSS after 7 to 10 days and by mRS at 3 months after symptom onset.^{23,24}

2.3 Automated ASPECTS Method

A schematic overview of the proposed method is shown in Fig. 1 divided into three main steps: ASPECTS labeling, brain densitometry, and EIC detection. During ASPECTS Labeling [Figs. 1(a)–(c)], the ASPECTS regions of interest are defined automatically in the NCCT data. Through brain densitometry [Figs. 1(d) and 1(e)], the density histograms are compared between contralateral sides, yielding BDSs in each ASPECTS region. In the final step, EICs are detected based on these BDSs resulting in an auto-ASPECTS score [Fig. 1(f)]. The overall method was implemented in the image processing and development environment MeVisLab (version 2.2.1).²⁵

In the following subsections, the different steps are presented in more detail along with the optimization.

Table 1 Patient characteristics in the training and test image sets.

	Training set	Test set		
	Total	Follow-up mRS ≤ 2	Follow-up mRS > 2	Total
Number of patients	10	24	39	63
Male/female	6/4	12/12	23/16	35/28
Age (years), mean \pm standard deviation	66.6 \pm 17	68.0 \pm 13	65.8 \pm 13	66.6 \pm 13
Symptom onset-to-scan time (h), first to third quartile	0:46 – 2:36	1:00–2:59	0:51–2:49	0:58–2:52

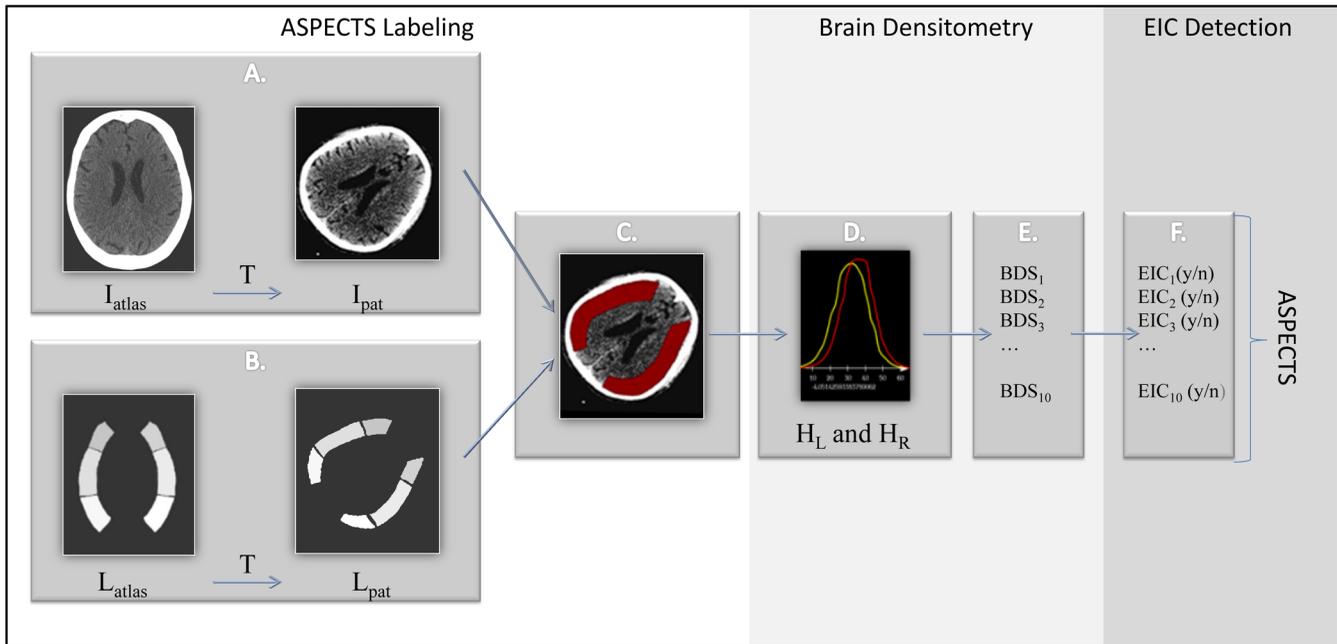


Fig. 1 Flow chart of the automated brain densitometry method based on ASPECTS. (a) The spatial transformation $T(x)$ matching the atlas image with the patient image is computed. (b) The atlas-label image, $L_{atlas}(x)$, is transformed by $T(x)$ to produce a labeled image $L_{pat}(x)$. (c) The atlas is applied as a mask over the patient image. (d) The density distribution is calculated for each ASPECTS region in the left and right hemisphere; H_L and H_R , respectively. (e) For each ASPECTS region, i , the brain density shift (BDS_i) is calculated between contralateral sides. (f) The EIC are detected by thresholding the BDS values, while accounting for irrelevant defects.

2.4 ASPECTS Labeling

The voxels are labeled according to their ASPECTS regions using an atlas-based segmentation method, see Figs. 1(a)–1(c). The atlas is composed of two image volumes: an atlas-intensity image, $I_{atlas}(x)$, containing the original intensity values, and an atlas-label image, $L_{atlas}(x)$, which contains the regions of interest with ASPECTS labels. The atlas-intensity image is registered to the patient's CT image data, $I_{pat}(x)$, yielding a spatial transformation, T . Subsequently, the same transformation is applied to $L_{atlas}(x)$ to obtain the labeled patient CT data, $L_{pat}(x)$.

To account for anatomical differences in the ventricular system²⁶ (including asymmetry of the lateral ventricles), two atlases with different ventricular shapes were generated (each containing only one CT scan). An experienced neuroradiologist (L.F.B.) selected two typical CT scans; one with a “nondilated” and one with a “moderately dilated” ventricular system.

ASPECTS regions of interest as described by Barber et al.⁶ were delineated manually for both image data sets using a free-form editing tool by a trained observer. All ASPECTS regions consist of an eight-slice volume. Delineation was checked for accuracy by an experienced neuroradiologist (L.F.B.).

Image data were registered using the open source software package Elastix,²⁷ version 4.3. Elastix deforms the atlas-intensity image by a coordinate transformation to optimally agree with the patient's image data. First, the atlas images were roughly aligned to the patient image by a rigid transformation. Subsequently, an affine registration with four resolution levels was used to compensate for remaining translation, rotation and scaling differences. Finally, to achieve the most detailed alignment, a nonrigid registration was applied starting with a coarse

B-spline grid, which was refined at four subsequent resolutions. For each of these steps, we used an iterative adaptive stochastic gradient descent method²⁹ with a maximum of 3000 iterations. This maximum was set conservatively without optimizing for speed to ensure convergence in all cases. We used a brain mask for focusing the registration in obtaining similarities within the brain, such as the ventricular system. This brain mask was generated automatically by segmentation of the intracranial structures.³⁰ After registration, the ASPECTS labeling of the patient image data was obtained by transformation of the two atlas-label images. After visual inspection of the registered atlases, the one that agreed best with the patient data was selected.

2.4.1 Optimization of the registration

The registration was optimized by assessing its accuracy and precision while varying registration settings. The alignment of contours of the lateral ventricles and the outer surface of the brain was used as accuracy measure. The brain contour was generated from the brain mask, detected as described above,³⁰ and the contours of the anterior horn and body of the lateral ventricles were determined by region growing. If needed, contours were adjusted manually. It should be noted that the manual adjustment was only needed during optimization of the algorithm. Normal analysis was automated without manual interaction. The registration error at a certain location was defined as the shortest distance between the ventricle and brain contours. The mean error and the variation in the errors were used to determine the accuracy and precision for different transform parameters.

We optimized the nonrigid (B-spline) registration with respect to the use of a brain mask and grid size using six different spacings: 5, 8, 10, 20, 30 and 40 mm. The efficacy of the usage of a brain mask was assessed by comparing the accuracy of registering: (1) without a brain mask; (2) using a brain mask covering brain tissue only; and (3) using a morphologically dilated brain mask ($9 \times 9 \times 1$ kernel size).

2.5 Brain Densitometry

For each ASPECTS region, the contralateral histograms of the density values were computed [Figs. 1(d)–1(e)]. The histograms were normalized (area under the curve is 1) to compensate for differences in volume between contralateral ASPECTS regions of interest.

We examined seven different measures for assessing the BDSs by comparing the density histograms of the left and right hemisphere, HL and HR. These measures were either represented by a difference in distribution features (first, second, and third quartiles, average density, and the peak density value) or by a shift in the entire histograms based on minimizing overlap or maximizing cross-correlation. As the affected side is known in clinical practice by functional symptoms, this information was also used in the automated assessment. The histograms and BDS measures of the ASPECTS regions were calculated in a MeVisLab-based software package (see Fig. 2).

2.6 EIC Detection

Before detecting EICs, we first optimized each measure and evaluated which BDS measure would be the best basis for this detection. This best method was used for the EIC detection and is described in more detail below.

The ability of the different measures of BDS to detect affected ASPECTS regions of interest was optimized regarding their threshold parameters [i.e., D_A , D_O , D in Eq. (3)] by receiver operating characteristic (ROC) curve analysis using the consensus ASPECTS reading as the reference standard. After optimization, we selected the best method for detecting EICs in an ASPECTS region (i), with the largest area under the ROC curve (see results section). This method is based on the overlap $O_i(\Delta d)$ between the left and right histogram ($H_{L,i}$ and $H_{R,i}$, respectively), after applying a density shift, Δd :

$$O_i(\Delta d) = 1 - \frac{1}{2} \sum_d |H_{L,i}(d) - H_{R,i}(d + \Delta d)|. \quad (1)$$

The value Δd between -40 and 40 HU for which $O_i(\Delta d)$ was largest was defined as the BSD_i

$$BDS_i = \operatorname{argmax}_{-40 \leq \Delta d \leq 40} O_i(\Delta d). \quad (2)$$

In some cases, brain densitometry could be affected by cerebrospinal fluid between the gyri in extensive atrophic brains or old ischemic scars which may be included in the ASPECTS regions, or by imaging artifacts. Therefore, additional findings (including imaging artifacts) were detected separately based on the dissimilarity of the contralateral histograms after applying the density shift by a large difference in the first quartile ($BDS_{q1,i}$). Subsequently, the results were used to exclude these areas automatically. The ultimate detection of an EIC_i in the i 'th ASPECTS region was thus determined by

$$EIC_i = \begin{cases} 0 & \left(\max_{-40 \leq \Delta d \leq 40} O_i(\Delta d) < D_A \wedge |BDS_{q1,i}| > D_O \right) \vee |BDS_i| \leq D \\ 1 & |BDS_i| > D \end{cases}, \quad (3)$$

where $BDS_{q1,i}$ is the density shift in the first quartile and optimal thresholds for defect detection $D_A = 84.5\%$ and $D_O = 3$ HU; and for EIC detection, $D = 1.5$ HU.

Finally, the automated ASPECTS score was calculated by following the ASPECTS definition:

$$\text{autoASPECTS} = 10 - \sum_{i=1}^{10} EIC_i. \quad (4)$$

2.7 Validation and Statistical Analysis

The best method, optimized in the training set, was subsequently validated using a separate test set of 63 patients (see Table 1). In a confounder analysis, correlations were evaluated between variables that should not be associated (for example, between pixel size and BDS). The Spearman's correlation coefficient was used to assess the association between ASPECTS (automated and manual) and the clinical outcome parameters. The variability between automated and consensus ASPECTS and the variability between observers was studied by Bland–Altman analysis. The agreement between ASPECTS measurements was evaluated by first dichotomizing all scores using a threshold of 7 (one of the common cut-off points for treatment^{6,31}), followed by constructing confusion matrices and computing the percent agreement between observers, and between automated and consensus ASPECTS.

Similarly, the agreement between observers and between automated and consensus ASPECTS was also evaluated in detecting EICs in different regions by calculating the percent agreement for each region separately. A McNemar test was conducted in these paired data to determine the significance of the differences in agreement.

A p -value below 0.05 was considered to indicate statistical significance.

3 Results

3.1 Method Optimization

The accuracy and precision of the different registration strategies are represented in Fig. 3. The use of a brain mask improved the ventricle registration accuracy considerably (from 6 to 2 mm), at the expense of a decreased accuracy of brain contour alignment. By dilating the brain mask, the registration accuracy of the brain contours improved with only a minor decrease in accuracy in registering the ventricles. Using B-spline grid sizes of 5, 8, and 10 mm gave comparable accuracies. Therefore, we choose a spacing of 10 mm (to allow for smooth deformations) in combination with a dilated mask in the following experiments.

Except for the measurements based on a shift in the histogram peaks, all measures had similar ROC areas under the curve, where BDS based on maximizing the histogram overlap with a threshold of 1.5 HU performed best with an area under the curve of 71.6%. This method detected the affected ASPECTS regions in the training set with a sensitivity of 75% and specificity of 74%.



Fig. 2 Capture of the graphical interface of the automated brain densitometry method based on the ASPECTS regions. The top row represents the different BDS measures per ASPECTS regions. The bottom rows are visualizations of the density distributions of the ASPECTS regions. From left to right, top to bottom: caudate nucleus; anterior and posterior part of internal capsule (for the ultimate quantification, these two regions were merged into one region, conform manual ASPECTS); lentiform nucleus; insular ribbon; M1; M2; M3; M4; M5, and M6. Red (gray) curves represent histograms of the left hemisphere and yellow (white) curves of the right hemisphere. A leftward shift of the yellow curve towards the lower density values indicates a hypodensity in the right hemisphere. The internal capsula area was divided in an anterior and a posterior part, to ensure detection of EICs when only a single part of the capsula was affected.

The detection of additional findings (including artifacts) in the ASPECTS regions of various measures was also optimized by ROC analysis, where the visual classification of defects by the radiologists was considered the reference standard. Additional findings were detected best with a sensitivity of 78% and specificity of 81%.

One analysis, including image registration and quantification, takes approximately 15 min on a standard PC. The method

has not been optimized for speed, however, and parallelization can significantly reduce analysis time.³²

3.2 Confounder Analysis

In the auto-ASPECTS analyses, the atlas with the “moderately dilated” ventricular system was used in 10 patients and the atlas with the “nondilated” ventricular system in the remaining 53

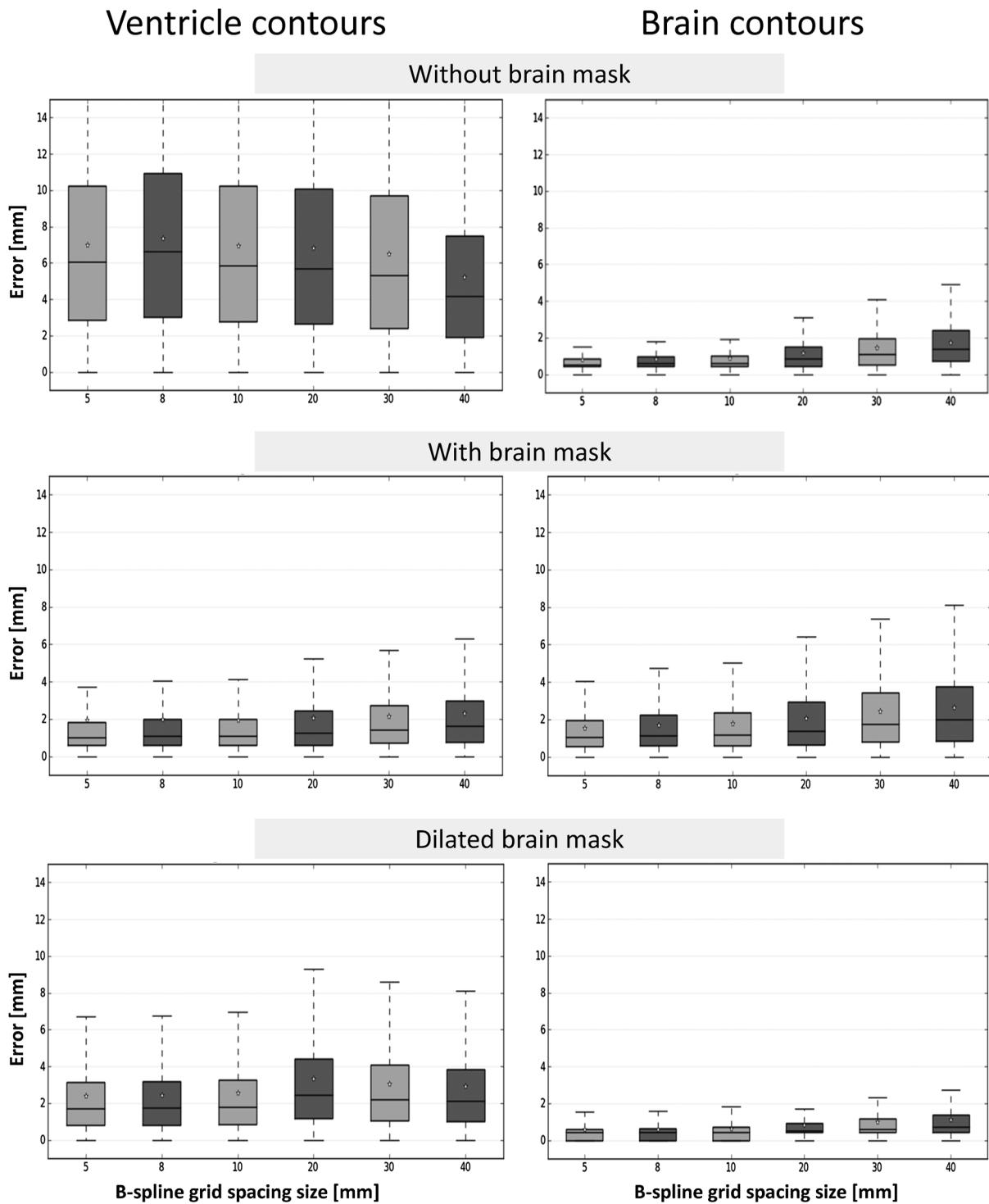


Fig. 3 Box plots of the residual registration errors in the training set of 10 patients, showing the effect of the use of brain masks. The left column shows the residual distances to the ventricle contours and the right column the distances to the brain contours.

patients. It was shown that mRS scores in the dilated group were 1.3 points higher than in the group with nondilated ventricular systems (95% confidence interval (95%CI): [0.1 to 2.5], $p = 0.03$). However, this can be explained by the additional effect of age, since age is associated with mRS (Spearman's $R = 0.64$, $p = 0.036$) and with the enlargement of cerebral ventricles. Since the choice of atlas is based on ventricle dilation, the two groups differ in age by 11 years, 95%CI: [2.8 to 20.0],

$p = 0.01$, with an accompanying difference in mRS scores. Therefore, the choice of atlas was not a confounding factor but a correct adaptation to the patient's anatomy.

No other potentially confounding factors were found.

3.3 Associations with Clinical Parameters

The correlation between the auto-ASPECTS score and clinical parameters is presented in Table 2, together with the manual

Table 2 Nonparametric (Spearman's rho) correlations among ASPECTS scores and clinical baseline and outcome parameters, in the test image set.

ASPECTS (n = 63):	Baseline NIHSS	NIHSS after 7 to 10 days	mRS after 3 months
Auto-ASPECTS	-0.277*	-0.192 ^{ns}	-0.272*
Observer 1	-0.358**	-0.285*	-0.107 ^{ns}
Observer 2	-0.315*	-0.273*	-0.091 ^{ns}
Consensus	-0.399**	-0.367**	-0.256*

Note: * Significant at the 0.05 level.; ** Significant at the 0.01 level.; ^{ns}not significant.

ASPECTS, scored separately and in consensus. All patients were included in a multicenter randomized open-label trial, treated with standard alteplase or alteplase with the early addition of intravenous aspirin, where ASPECTS scoring was not used as inclusion criterion. Since early administration of intravenous aspirin had no effect on the outcome, treatment was not considered a confounder in this study.²² Due to the standardization of the acquisition protocol, we used all trial data from one center with the same CT scanner. For both consensus and auto-ASPECTS, there was a weak but statistically significant correlation with baseline NIHSS. The consensus scoring correlated slightly better than the automated methods and individual observers. After 7 to 10 days, the correlation between NIHSS was maintained for the consensus and individual scores, but the automated methods did not correlate significantly. The correlations with mRS score at 3 months were lower than with (baseline and follow-up) NIHSS, with a slightly higher correlation for the automated method. The scores from the individual observers, however, did not correlate significantly with mRS.

3.4 Interobserver Agreement

Figure 4 shows the results of the Bland-Altman analysis. The interobserver differences and mean values did not follow a

normal distribution, and an association was observed between the interobserver difference and the mean score (Spearman's $R = 0.67$, $p < 0.01$), despite the outlier with a seven-point difference [Fig. 4(b)]. Therefore, there was a significant bias between observers that was dependent on the magnitude of the score, and as a consequence, a reliable limit of agreement could not be calculated. However, the over-all agreement between automated and consensus was comparable to the agreement between observers, with a slightly higher variability between the observers. The average difference (SD) between consensus and auto-ASPECTS was -0.54 (1.20) points, and the average difference (SD) between observers was -0.81 (1.62) points. In 73% of the cases, the auto-ASPECTS deviated from the consensus score by one point or less; the same exact rate occurred for the interobserver agreement: in 73% of cases, the difference was one point or less.

The agreement among the different ASPECT measurements after dichotomizing with a threshold 7 is presented in Table 3. The percent agreement between the individual observers was 83%, whereas the percent agreement between automated and consensus scoring was 89%. These percent agreements did not differ significantly, however, between the two comparisons (McNemar test).

The agreement in detecting EIC's in the individual ASPECTS regions is shown in Fig. 5, between observers and between the automated method and consensus ASPECTS. Overall, the automated method found more EICs than the consensus scoring (87 versus 52, out of 629 regions, that were free of artifacts) and observer 2 found more EICs than observer 1 (82 versus 31, out of 629 regions). The percent agreement between observers ranged between 67% and 97%, whereas the percent agreement between automated method and consensus ASPECTS ranged between 73% and 89%. In the regions M2, M3, and M6, the interobserver agreement was significantly higher than for the automated method and consensus scoring (McNemar test). Of the 52 manually detected ASPECTS regions, five EICs were missed by the automated method due to the exclusion of these regions due to additional findings such as high degree of atrophy, old ischemic lesions, or scanning artifacts.

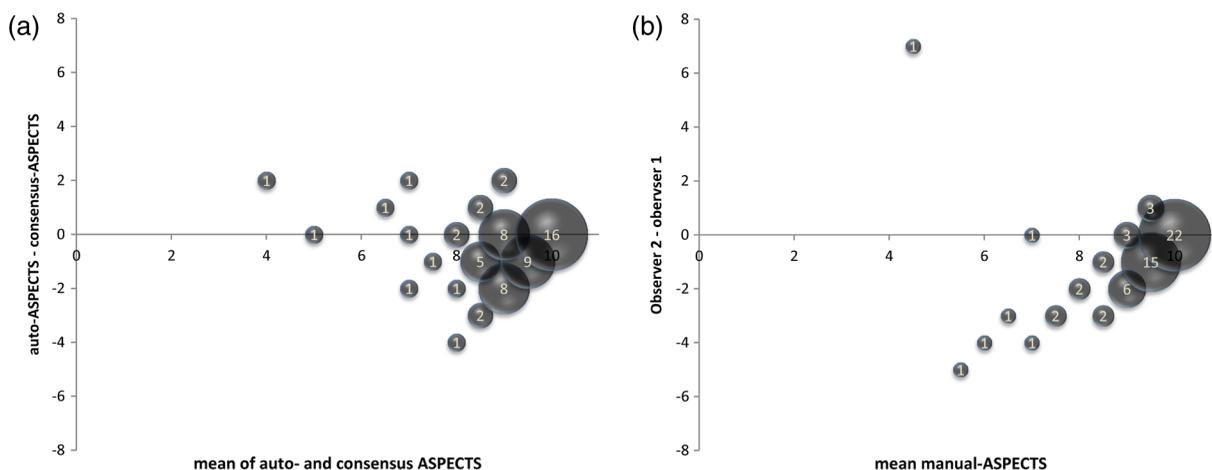


Fig. 4 Bland-Altman plots of (a) auto-ASPECTS versus consensus-ASPECTS and (b) differences between observers. The size of the circles and the value inside indicate the number of cases with the same result.

Table 3 Confusion matrices between observers and between automated and consensus ASPECTS (equal to or less than 7 versus higher than 7), in the test image set.

	Agreement:		Observer 1		Agreement:		Consensus	
	83%	≤ 7	> 7	≤ 7	89%	≤ 7	> 7	
Observer 2	≤ 7	1	10	Auto-ASPECTS	≤ 7	4	6	
	>7	1	51		>7	1	52	

4 Discussion

We have designed, optimized, and validated an automated method for brain densitometry in ASPECTS regions of interest to detect and quantify EICs on NCCT images, using atlas-based segmentation to automatically delineate the ASPECTS regions. In the validation against clinical parameters, we found variable associations. Both manual and automated ASPECTS correlated with baseline stroke severity (NIHSS). The individual manual scores still correlated with NIHSS after 7 to 10 days, where the automated method did not. On the other hand, the manual ASPECTS scoring by the individual observers did not correlate with clinical dependency scores (mRS) after 3 months, where the automated method did correlate. The poor and absent correlations could be due to the early or reversible stage of the

ischemic lesion, however, determining the cause of this difference was beyond the scope of this study.

Except for selecting the optimal atlas (normal versus enlarged ventricles), the method is fully automated. For this reason, no test-retest was performed as the remainder of the method is completely observer independent. The atlas-based segmentation method for classifying voxels into specific ASPECTS regions was optimized by minimizing the residual registration errors of the brain and ventricle contours. From these experiments, it became clear that brain masks are needed to focus the image registration to brain tissue only. However, although this improved the alignment of the ventricles, the accuracy of registering the outer contours of the brain deteriorated due to over-focusing. Therefore, a dilated brain mask was proposed in

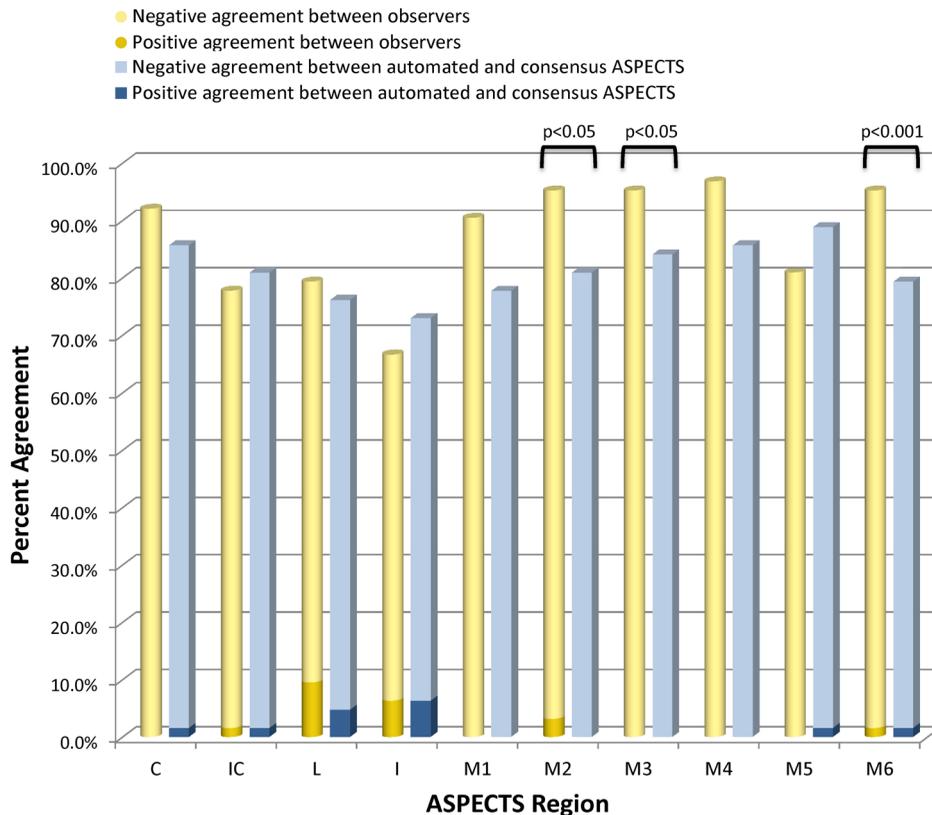


Fig. 5 Percent agreement between observers (cylinders) and between automated and consensus scoring (blocks), for each ASPECTS region. The lower part of the columns represents the percentage agreement on positive EIC findings (i.e., the percentage of cases where both methods indicate a positive EIC finding), the upper part represents the negative agreement (% agreement of a negative EIC finding). From left to right: caudate nucleus (C); internal capsule (IC); lentiform nucleus (L); insular ribbon (I); M1; M2; M3;M4, M5 and M6.

which the gradient between the skull and brain tissue is included in the registration, thereby improving the registration quality of the outer contours without significantly degrading the registration of ventricle contours.

B-spline grid spacing, which determines the smoothness of the deformation, did not influence the accuracy as much as the usage of masks, since sizes between 5 and 10 mm gave comparable results. Therefore, we choose a spacing of 10 mm to allow for the smoothest possible deformations.

This optimized segmentation method obtained an accuracy between 0.45 mm (at the brain surface) and 1.8 mm (at the ventricle). For comparison, Pexman et al.³³ reported at least a 10 mm difference in the size of each “M” area between the minimum and maximum chosen by human observers. Despite this variability, they still found an excellent interobserver agreement of ASPECTS scoring. Therefore, the accuracy of our segmentation method for labeling the ASPECTS regions may be sufficient for clinical use.

Signs of EICs were indicated by a BDS between the histograms of contralateral ASPECTS regions. To optimize this BDS method, a comparison with manual consensus scoring was made and we obtained an optimal method with a threshold of 1.5 HU. A sensitivity of 75% and specificity of 74% was obtained.

In the early stages of cerebral ischemia, there is a maximum change of only 1.6 to 4.4 HU within the first 2 to 2.5 h of ischemia.^{15,34–36} As our patient group was scanned between 0:58 and 2:52 h after symptom onset (first and third quartile), the optimal threshold of 1.5 HU is consistent with this reported range of density changes.

In a study on quantifying EICs by comparing density histograms from contralateral sides,¹⁸ the nucleus caudatus, lentiform nucleus, and insula were labeled by registration of the NCCT scan to the Talairach³⁷ template. Unfortunately, no objective measure for the quality of the alignment was reported. Furthermore, this method is not based on ASPECTS regions, which makes it less applicable in clinical practice.

In the MRI atlas-based scoring method from Kosior et al.,¹⁹ the average errors in localization of the Talairach landmarks using FTT ranged between 0.08 and 1.49 mm, depending on the location of the landmark in the brain.²¹ The registration accuracy reported in our study fits well within this range.

In the study by Kosior et al.,¹⁹ the mean symptom onset-to-scan time was 11.2 ± 11.7 h (average \pm standard deviation) for MRI scans, compared to a range of only [0:58 to 2:52] h for our NCCT scans (Table 1). As a longer onset-to-scan time increases the visibility of EICs, results from this study are not comparable to our data.

This study has a number of limitations. The method compares density distributions between contralateral sides. As such, it does not account for lesions, such as remote infarcts or encephalomalacia, on the contralateral side. The detection and correction of such lesions could improve accuracy.

Because the patients included in this study live in a densely populated area, onset-to-scan time is relative short. Especially for these patients, it is difficult to detect the (very) EICs and an automated method may have added value. As a result of this short onset-to-scan time, the ASPECTS scores were higher, resulting in a different ASPECTS distribution than in less-densely populated areas with a less extensive public infrastructure or less stroke awareness. Furthermore, since all patients in the population were part of a randomized trial, it was required that they were eligible for alteplase. Guidelines for management

of patients with acute stroke suggest that hypodensity is the only meaningful imaging finding in patients eligible for standard recombinant tissue plasminogen activator treatment.³⁸ Therefore, all patients were eligible for standard treatment because they all arrived within 4.5 h after symptom onset. However, because ASPECTS is mainly used in clinical research and not yet in clinical practice, we believe that the added value of this score is highest for the population arriving within 6 h after stroke onset for treatment effectiveness studies. The validation in patients with a longer onset-to-scan time is prudent, but was beyond the scope of this study. Such a validation should be performed with a more diverse patient population with a larger window of onset-to-scan times. Moreover, it is expected that the automated method has a higher accuracy for more severe strokes since the hypodensity increases with time, making the histogram shifts more apparent.

Although techniques have been described that increase visibility of EICs on NCCT scan,^{9,39–41} comparative data are limited since most methods are based on MR imaging, accuracy is not always reported and none have been correlated to clinical outcome. Although diffusion-weighted MRI is considered the gold standard for assessing EICs, MRI is rarely available in the emergency setting of patients with acute stroke and consequently it is not the standard care.⁴² Our method was, therefore, developed for NCCT, which is the standard initial imaging examination for acute stroke, since it is fast, inexpensive, widely and easily available, and provides whole brain coverage.⁴³ A limitation of this study is, therefore, that no gold-standard (MRI) of the EICs was available. We could only compare the automated scoring with manual scoring, which had considerable observer variability. The variation of the manual ASPECTS limits the value of the reference measurement, however, accurately it resembles the expected variation in clinical practice. The variation in manual ASPECTS was a main motivation to initiate this study on an objective automated approach knowing that a direct comparison with observers cannot prove any improvements.

The automated ASPECTS method excludes regions where a small old ischemic scar is found (extreme hypoattenuation). Therefore, this rules out situations where new EICs occurs in proximity of an old infarct, which can be picked up by visual scoring. Out of 52 EICs detected by the consensus scoring, 5 were missed by the automated method. As this is a substantial limitation, new image analysis methods need to be developed focused on this particular situation.

Another possible limitation of our method is the fact that histogram analysis does not include local image information. As a result, small focal lesions may not contribute to a detectable shift in the histogram. Therefore, in order to improve the performance of the method in detecting focal lesions, spatial information needs to be included, requiring a different approach. These methods may, however, be less sensitive to subtle BDSs in diffuse hypoattenuation regions. Thus, hybrid approaches, where histogram-based methods are combined with methods that quantify spatial/texture information, may be developed.

Because of the retrospective standardization of the image acquisition protocol, we used CT data from the same CT scanner from only one center. Therefore, additional validation is needed to prove the effectiveness of the method in a general setting.

Some user-interaction was still needed in the study since the appropriate atlas needed to be selected manually. A detection

method of the ventricles could be used to determine the ventricle volume and select the dilated or nondilated atlas automatically. Alternatively, the residual normalized cross-correlation after registration could be used as a basis for this atlas selection. Adding more atlases is not needed since the two atlases cover the anatomical variances sufficiently, as the residual registration errors were independent of the atlas used and small errors were obtained over the entire training set.

To evaluate the impact of the different scores on the decision to initiate treatment, when using a threshold value of 7, the data were dichotomized using this threshold. There was agreement between individual observers in 83%, whereas agreement occurred in 89% between automated and consensus scoring. Please note that kappa values were not presented in this study since the kappa statistic is affected by prevalence, and the distribution of low and high scores was indeed skewed in our study population. Using a threshold value of 7, the decision for initiating treatment would be different between the individual observers in 11 out of 63 cases (17%), whereas between automated and consensus scoring, this decision would be different in 7 out of 63 cases (11%). This difference was, however, not statistically significant.

The agreement analysis for each ASPECTS region (see Fig. 5) showed that the observers agreed less in the internal capsule (78%), lentiform nucleus (79%) and insular ribbon (67%), compared to the other regions, whereas the agreement between the automated and consensus scoring was less variable, but generally at a lower level than the observer agreement. In the regions M2, M3, and M6, the interobserver agreement was significantly higher. This difference in agreement was mainly determined by the difference in agreement in negative EIC detections; for the positive detections, the interobserver agreement and agreement between automated and consensus ASPECTS was more evenly divided over the different regions. This may explain the difference in results between local and regional agreement.

In this paper, an automated ASPECTS scoring method has been presented as an alternative to manual ASPECTS score. The method can also be used, however, as a supplementary tool to assist manual scoring. The display of the shifts in density may help physicians in deciding whether a region of interest is affected. The added value of using our tool in such a manner needs to be evaluated in a future study in a population with a wide range of ASPECTS scores. It should be noted that ASPECTS is currently not a part of standard diagnosis workflow but is mainly used in clinical research.^{4,5} However, with the advent of alternative stroke therapies such as intraarterial treatment, ASPECTS has the potential to be included in the acute setting in the near future.

In conclusion, we have demonstrated the feasibility of a stroke severity measurement by automated brain densitometry of contralateral ASPECTS regions. This study provides an automated method to transform brain atlases to a patient's NCCT scan and to automatically detect and quantify EICs. This automated method was in agreement with manual consensus scoring in 73% without bias or outliers, in contrast with individual observers. With the same limits of agreement between observers and between automated and consensus scoring, we can conclude that the automated method performed comparable to the manual ASPECTS scoring, keeping in mind that in the acute setting, ASPECTS is not scored in consensus but by a single observer. Since our method obtained similar correlations with clinical

outcome measures, it shows the potential to assist or eventually replace manual ASPECTS assessments as indicators of stroke severity.

Acknowledgments

A large portion of this work has been performed as a graduation project by F. E. van Rijn, at the Institute of Technical Medicine, University of Twente, Enschede, The Netherlands.

References

1. M. A. Foulkes et al., "The Stroke Data Bank: design, methods, and baseline characteristics," *Stroke* **19**(5), 547–554 (1988).
2. J. Astrup, B. K. Siesjo, and L. Symon, "Thresholds in cerebral ischemia—the ischemic penumbra," *Stroke* **12**(6), 723–725 (1981).
3. K. W. Muir et al., "Can the ischemic penumbra be identified on non-contrast CT of acute stroke?," *Stroke* **38**(9), 2485–2490 (2007).
4. A. J. Yoo et al., "Impact of pretreatment noncontrast CT Alberta Stroke Program Early CT Score on clinical outcome after intra-arterial stroke therapy," *Stroke* **45**(3), 746–751 (2014).
5. O. A. Berkhemer et al., "A randomized trial of intraarterial treatment for acute ischemic stroke," *N. Engl. J. Med.* **372**(1), 11–20 (2015).
6. P. Barber et al., "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* **355**(9216), 1670–1674 (2000).
7. R. von Kummer et al., "Detectability of cerebral hemisphere ischaemic infarcts by CT within 6 h of stroke," *Neuroradiology* **38**(1), 31–33 (1996).
8. M. E. Mullins et al., "Influence of availability of clinical history on detection of early stroke using unenhanced CT and diffusion-weighted MR imaging," *Am. J. Roentgenol.* **179**(1), 223–228 (2002).
9. M. H. Lev et al., "Acute stroke: improved nonenhanced CT detection—benefits of soft-copy interpretation by using variable window width and center level settings," *Radiology* **213**(1), 150–155 (1999).
10. D. Vu and M. H. Lev, "Noncontrast CT in acute stroke," *Semin. Ultrasound CT MRI* **26**(6), 380–386 (2005).
11. D. W. Dippel et al., "The validity and reliability of signs of early infarction on CT in acute ischaemic stroke," *Neuroradiology* **42**(9), 629–633 (2000).
12. J. Grotta et al., "Agreement and variability in the interpretation of early CT changes in stroke patients qualifying for intravenous rtPA therapy," *Stroke* **30**(8), 1528–1533 (1999).
13. J. M. Wardlaw et al., "Can stroke physicians and neuroradiologists identify signs of early cerebral infarction on CT?," *J. Neurol. Neurosurg. Psychiatry* **67**(5), 651–653 (1999).
14. A. C. Gupta et al., "Interobserver reliability of baseline noncontrast CT Alberta Stroke Program Early CT Score for intra-arterial stroke treatment selection," *Am. J. Neuroradiol.* **33**(6), 1046–1049 (2012).
15. K. Nedelchev et al., "Predictors of early mortality after acute ischaemic stroke," *Swiss Med. Wkly* **140**(17–18), 254–259 (2010).
16. I. Dzialowski et al., "Brain tissue water uptake after middle cerebral artery occlusion assessed with CT," *J. Neuroimaging* **14**(1), 42–48 (2004).
17. M. Bendszus et al., "Improved CT diagnosis of acute middle cerebral artery territory infarcts with density-difference analysis," *Neuroradiology* **39**(2), 127–131 (1997).
18. J. Maldjian et al., "Automated CT segmentation and analysis for acute middle cerebral artery stroke," *Am. J. Neuroradiol.* **22**(6), 1050–1055 (2001).
19. R. K. Kosior et al., "Atlas-based topographical scoring for magnetic resonance imaging of acute stroke," *Stroke* **41**(3), 455–460 (2010).
20. W. L. Nowinski et al., "Analysis of ischemic stroke MR images by means of brain atlases of anatomy and blood supply territories," *Acad. Radiol.* **13**(8), 1025–1034 (2006).
21. W. L. Nowinski et al., "Fast Talairach Transformation for magnetic resonance neuroimages," *J. Comput. Assist. Tomogr.* **30**(4), 629–641 (2006).
22. S. M. Zinkstok and Y. B. Roos, "Early administration of aspirin in patients treated with alteplase for acute ischaemic stroke: a randomised controlled trial," *Lancet* **380**(9843), 731–737 (2012).

23. T. Brott et al., "Measurements of acute cerebral infarction: a clinical examination scale," *Stroke* **20**(7), 864–870 (1989).
24. J. C. van Swieten et al., "Interobserver agreement for the assessment of handicap in stroke patients," *Stroke* **19**(5), 604–607 (1988).
25. MeVisLab—development environment for medical image processing and visualization, MeVis Medical Solutions AG and Fraunhofer MEVIS, Bremen, Germany (2011).
26. C. Gyldensted and M. Kosteljanetz, "Measurements of the normal ventricular system with computer tomography of the brain. A preliminary study on 44 adults," *Neuroradiology* **10**(4), 205–213 (1976).
27. S. Klein et al., "Elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging* **29**(1), 196–205 (2010).
<http://elastix.isi.uu.nl>.
29. S. Klein et al., "Adaptive stochastic gradient descent optimisation for image registration," *Int. J. Comput. Vision* **81**(3), 227–239 (2009).
30. K. H. Höhne and W. A. Hanson, "Interactive 3D segmentation of MRI and CT volumes using morphological operations," *J. Comput. Assist. Tomogr.* **16**(2), 285–294 (1992).
31. R. I. Aviv et al., "Alberta Stroke Program Early CT Scoring of CT perfusion in early stroke visualization and assessment," *Am. J. Neuroradiol.* **28**(10), 1975–1980 (2007).
32. D. P. Shamonin et al., "Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease," *Front. Neuroinform.* **7**, 50 (2013).
33. J. H. Pexman et al., "Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke," *Am. J. Neuroradiol.* **22**(8), 1534–1542 (2001).
34. F. J. Schuier and K. A. Hossmann, "Experimental brain infarcts in cats. II. Ischemic brain edema," *Stroke* **11**(6), 593–601 (1980).
35. E. Unger, J. Littlefield, and M. Gado, "Water content and water structure in CT and MR signal changes: possible influence in detection of early stroke," *Am. J. Neuroradiol.* **9**(4), 687–691 (1988).
36. O. Watanabe, C. R. West, and A. Bremer, "Experimental regional cerebral ischemia in the middle cerebral artery territory in primates. Part 2: Effects on brain water and electrolytes in the early phase of MCA stroke," *Stroke* **8**(1), 71–76 (1977).
37. J. L. Lancaster et al., "Automated Talairach atlas labels for functional brain mapping," *Hum. Brain Mapp.* **10**(3), 120–131 (2000).
38. E. C. Jauch et al., "Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association," *Stroke* **44**(3), 870–947 (2013).
39. A. Przelaskowski et al., "Improved early stroke detection: wavelet-based perception enhancement of computerized tomography exams," *Comput. Biol. Med.* **37**(4), 524–533 (2007).
40. N. Takahashi et al., "Improvement of detection of hypoattenuation in acute ischemic stroke in unenhanced computed tomography using an adaptive smoothing filter," *Acta Radiol.* **49**(7), 816–826 (2008).
41. F. H. Tang, D. K. Ng, and D. H. Chow, "An image feature approach for computer-aided detection of ischemic stroke," *Comput. Biol. Med.* **41**(7), 529–536 (2011).
42. D. P. Davis, T. Robertson, and S. G. Imbesi, "Diffusion-weighted magnetic resonance imaging versus computed tomography in the diagnosis of acute ischemic stroke," *J. Emerg. Med.* **31**(3), 269–277 (2006).
43. R. Dhamija and G. Donnan, "The role of neuroimaging in acute stroke," *Ann. Indian Acad. Neurol.* **11**, 12–23 (2008).

Biographies for the authors are not available.