Automated ASPECTS on Noncontrast CT Scans in Patients with Acute Ischemic Stroke Using Machine Learning


*AJNR Am J Neuroradiol* 2019, 40 (1) 33-38
doi: https://doi.org/10.3174/ajnr.A5889
http://www.ajnr.org/content/40/1/33
Automated ASPECTS on Noncontrast CT Scans in Patients with Acute Ischemic Stroke Using Machine Learning


ABSTRACT

BACKGROUND AND PURPOSE: Alberta Stroke Program Early CT Score (ASPECTS) was devised as a systematic method to assess the extent of early ischemic change on noncontrast CT (NCCT) in patients with acute ischemic stroke (AIS). Our aim was to automate ASPECTS to objectively score NCCT of AIS patients.

MATERIALS AND METHODS: We collected NCCT images with a 5-mm thickness of 257 patients with acute ischemic stroke (<8 hours from onset to scans) followed by a diffusion-weighted imaging acquisition within 1 hour. Expert ASPECTS readings on DWI were used as ground truth. Texture features were extracted from each ASPECTS region of the 157 training patient images to train a random forest classifier. The unseen 100 testing patient images were used to evaluate the performance of the trained classifier. Statistical analyses on the total ASPECTS and region-level ASPECTS were conducted.

RESULTS: For the total ASPECTS of the unseen 100 patients, the intraclass correlation coefficient between the automated ASPECTS method and DWI ASPECTS scores of expert readings was 0.76 (95% confidence interval, 0.67–0.83) and the mean ASPECTS difference in the Bland-Altman plot was 0.3 (limits of agreement, −3.3, 2.6). Individual ASPECTS region-level analysis showed that our method yielded κ = 0.60, sensitivity of 66.2%, specificity of 91.8%, and area under curve of 0.79 for 100 × 10 ASPECTS regions. Additionally, when ASPECTS was dichotomized (>4 and ≤4), κ = 0.78, sensitivity of 97.8%, specificity of 80%, and area under the curve of 0.89 were generated between the proposed method and expert readings on DWI.

CONCLUSIONS: The proposed automated ASPECTS scoring approach shows reasonable ability to determine ASPECTS on NCCT images in patients presenting with acute ischemic stroke.

ABBREVIATIONS: AIS = acute ischemic stroke; AUC = area under curve; ICC = intraclass correlation coefficient; IQR = interquartile range

Management of patients with acute ischemic stroke (AIS) relies heavily on an assessment of the extent of irreversibly injured brain at baseline. Patients with extensive early ischemic changes at presentation are unlikely to benefit from thrombolysis or thrombectomy procedures. Moreover, such patients may also be at higher risk of developing complications of treatment such as intracerebral hemorrhage. The Alberta Stroke Program Early CT Score was devised as a systematic method of assessing the extent of early ischemic change on noncontrast CT in patients with AIS.1,2 Across the years, ASPECTS has gained credence and is now used the world over for this purpose,3–7 though it has not been proved useful for selecting patients for treatment.8,9

Although conceptually, the ASPECTS is a simple method, scoring early ischemic change on NCCT scans continues to be a challenge, especially for readers with less experience.10–12 Technical factors such as peak x-ray energy (kiloelectron volt/megaelectron volt) image processing and display procedures; patient factors such as old infarcts, brain atrophy, and leukoaraiosis; and reader factors such as experience, training, and specialty, all potentially affect ASPECTS interpretation.11,12 A solution to improve ASPECTS reading is training readers to recognize these issues while providing them with strategies that can help improve the reliability and validity of these reads. Another solution is to use novel technologies such as machine
learning and feature extraction to develop automated solutions to ASPECTS interpretation.\textsuperscript{13-17}

In recent years, evidence that automated ASPECTS scoring methods based on machine learning are comparable with expert reading of ASPECTS is accumulating.\textsuperscript{18-24} In this study, we developed an automated ASPECTS scoring system based on machine learning and feature engineering and compared it with expert ASPECTS readings on acute DWI. We introduced multiple high-order computational textural features into our machine learning model and hypothesized that this automated method can determine ASPECTS scores accurately and reliably compared with expert ASPECTS readings on acute DWI.

**MATERIALS AND METHODS**

Data are from the Keimyung Stroke Registry, an ongoing single-center prospective cohort study of patients with acute ischemic stroke presenting to the Keimyung University Hospital in Daegu, South Korea. Two hundred fifty-seven patients with acute ischemic stroke presenting within 8 hours of last known well who had baseline NCCT (slice thickness, ≤5 mm) followed by DWI performed within 1 hour of NCCT were included in the study. An expert scored ASPECTS on DWI; any individual region with diffusion affected. To assess the reliability of expert-reading DWI ASPECTS, another expert was asked to score 60 DWI scans randomly selected from the 257 patients with AIS.

Of the 257 patient images, 157 were randomly selected for training a machine learning model, while the remaining 100 images were used to evaluate the trained model. Specifically, a NCCT template with ASPECTS regions manually contoured was non-linearly registered onto all NCCT images (Fig 1). During the training stage, 376 texture features (details of texture features are shown in the On-line Appendix) such as high-order statistics and image textural features were extracted from each ASPECTS region from the 157 patient images bilaterally after median filtering. Note that the feature extraction and classification for each ASPECTS region were performed in 3D. Information on the side of the brain affected by ischemic stroke was used as an additional input to compute difference features between ischemic and normal brain tissue. Specifically, observers first determined the ischemic hemisphere based on imaging and clinical parameters. Feature differences were then obtained by subtracting regional level values on the ischemic side from those on the contralateral side. Sixty patients (38 in the training dataset and 22 in the testing dataset) with posterior circulation strokes were included intentionally to reflect clinical reality. In patients with posterior circulation strokes, the left side was regarded as the default ischemic side.

The computed features were first ranked using linear discriminant analysis. The ranked features were input into a random forest model using the expert-assessed ASPECTS on DWI as a class label. We used 5-fold cross-validation on the training samples to select training hyperparameters including the number of trees in the forest, the maximum depth of trees, and also the number of ranked features. Class weight was set to deal with the imbalanced data distribution on the basis of the ratio of abnormal and normal samples in the training data. The detailed parameter settings are shown in the On-line Table. We trained a classifier for each ASPECTS region. The random forest training and testing were implemented using Scikit-learn in Python (http://scikit-learn.org/stable/). The trained random forest classifier was then validated on the remaining 100 test patient images. A flowchart of the training and testing process of each ASPECTS region is shown in Fig 2.

**Statistical Analysis**

Expert ASPECTS readings on DWI of the 100 test images were used as the ground truth to evaluate the automated ASPECTS obtained by our method. Agreement on the total ASPECTS score was measured using the intraclass correlation coefficient (ICC). Boxplots and Bland-Altman plots were used to illustrate differences in the assessment of total ASPECTS between the automated method and the ground-truth (expert-read DWI ASPECTS). The ICC analysis was also stratified by stroke onset-to-CT time (≤90 minutes, n = 69; 90–270 minutes, n = 21; >270 minutes, n = 10). Because physicians use the presence or absence of extensive early ischemic changes to make clinical decisions on treatment in patients with acute ischemic stroke, we also assessed agreement on the ASPECTS interpretation between the automated method and DWI using κ statistics on a dichotomized ASPECTS threshold (>4 versus ≤4).\textsuperscript{25} κ statistics were also used to assess agreement between the automated method and expert-read DWI at each individual ASPECTS region.

Receiver operating characteristics based on the MedCalc for Windows software (MedCalc Software, Mariakerke, Belgium) were used to report the area under the curve (AUC) for the dichotomized ASPECTS (>4 versus ≤4), and individual region-level ASPECTS analysis using automated ASPECTS, as an independent variable and expert-read DWI ASPECTS, as a dependent variable. A clustered receiver oper-
The agreement between the expert-rated CT ASPECTS and the automated ASPECTS method was 0.61 (95% CI, 0.47–0.72). The AUC (0.85: 95% CI, 0.77–0.92) were obtained.

Sensitivity (90%: 95% CI, 55.5%–99.7%), specificity (80%: 95% CI, 34.8%–93.3%), F1 measure (0.98), and AUC (0.89: 95% CI, 0.81–0.94) were reasonably good. When ASPECTS was trichotomized (0–4, 5–7, 8–10), the agreement between the automated method and DWI was good as well (linear weighted $\kappa$ = 0.66; 95% CI, 0.53–0.80). Agreement on ASPECTS between the automated method and DWI at the individual ASPECTS region level is reported in Table 1. $\kappa$ for agreement between the 2 methods ranged from 0.36 to 0.64. The automated ASPECTS method demonstrated high specificity but modest sensitivity compared with DWI at the regional level. F1 measures are also shown in Table 1.

Sensitivity analysis was attempted by varying the threshold involvement of each ASPECTS region on expert-read DWI as >0% and >50% involvement compared with the >20% involvement used for primary analyses. Additionally, to demonstrate the efficacy of the developed automated ASPECTS method, we compared the ASPECTS reading of a stroke expert on DWI at the individual ASPECTS region level. F1 measures are also shown in Table 1.

The $\kappa$ values for regional and dichotomized ASPECTS between the 2 expert-read DWI ASPECTSs were 0.86 (95% CI, 0.81–0.91) and 0.85 (95% CI, 0.56–1), respectively. The ICC for total ASPECTS between the 2 expert-read DWI ASPECTSs was 0.90 (95% CI, 0.84–0.94).

When patients were stratified by stroke onset-to-CT time ($\leq$90 minutes, $n = 69$; 90–270 minutes, $n = 21$; and >270 minutes, $n = 10$), the ICCs between the automated CT ASPECTS and the DWI ASPECTS for these 3 subgroups were 0.80 (95% CI, 0.69–0.87), 0.77 (95% CI, 0.52–0.90), and 0.26 (95% CI, −0.40–0.75), respectively. No evidence was found that the ICC increased in patients with longer onset-to-CT time using the automated ASPECTS method.

The median baseline ASPECTS generated by the automated method on test data ($n = 100$) was 8 (IQR, 7–9) versus a score of 7 (IQR, 6–9) on the ground truth DWI. Figure 3A shows a boxplot overlaid with a scatterplot showing the distribution of the automated CT ASPECTS at each individual ASPECTS on DWI. The intraclass correlation coefficient for total ASPECTS between the automated method and DWI was 0.76 (95% CI, 0.67–0.83). Figure 3B illustrates Bland-Altman agreement plots between the automated method and DWI for total ASPECTS. The mean difference in total ASPECTS between the automated method and DWI was minimal (0.3; limit of agreement, −3.3, 2.6).

Agreement on ASPECTS between the automated method and DWI using a dichotomized ASPECTS threshold of >4 versus $\leq$4 was good ($\kappa = 0.78$: 95% CI, 0.57–0.99). Sensitivity (97.8%: 95% CI, 92.2%–99.7%), specificity (80%: 95% CI, 34.8%–93.3%), F1 measure (0.98), and AUC (0.89: 95% CI, 0.81–0.94) were reasonably good. When ASPECTS was trichotomized (0–4, 5–7, 8–10), the agreement between the automated method and DWI was good as well (linear weighted $\kappa = 0.66$; 95% CI, 0.53–0.80). Agreement on ASPECTS between the automated method and DWI at the individual ASPECTS region level is reported in Table 1. $\kappa$ for agreement between the 2 methods ranged from 0.36 to 0.64. The automated ASPECTS method demonstrated high specificity but modest sensitivity compared with DWI at the regional level. F1 measures are also shown in Table 1.

Sensitivity analysis was attempted by varying the threshold involvement of each ASPECTS region on expert-read DWI as >0% and >50% involvement in addition to the >20% involvement used for the primary analyses. Region-level agreement between the automated method and expert-rated DWI ASPECTS for all 3 thresholds is shown in Table 2. Agreement between the 2 methods was best when DWI ASPECTS was rated using the >50% threshold method.

The ICC for total ASPECTS between the expert-rated NCCT and DWI was 0.67 (95% CI, 0.55–0.77). The agreement between the expert-rated CT ASPECTS and DWI using a dichotomized ASPECTS threshold of >4 versus $\leq$4 was modest ($\kappa = 0.42$: 95% CI, 0.22–0.62). Sensitivity (81.1%: 95% CI, 71.5%–88.6%), specificity (90%: 95% CI, 55.5%–99.7%), F1 measure (0.89), and AUC (0.85: 95% CI, 0.77–0.92) were obtained.

The ICC for total ASPECTS between the expert-rated NCCT and the automated CT ASPECTS was 0.61 (95% CI, 0.47–0.72). The agreement between the expert-rated CT ASPECTS and the automated CT ASPECTS using a dichotomized ASPECTS threshold of >4 versus $\leq$4 was modest ($\kappa = 0.48$: 95% CI, 0.28–0.68). An example of expert-rated DWI ASPECTS, our automated CT ASPECTS, and expert-rated CT ASPECTS is shown in Fig 4.
The proposed automated ASPECTS scoring method is based on feature engineering and random forest learning. Random forest is considered one of the most recent and popular boosting methods and has proved classification performance for difficult problems in many medical image–analysis applications compared with other classifiers.27,28 Random forest is an ensemble learning method that combines multiple weak classifiers (decision trees) and lets these decision trees vote for the most popular class. Each tree in the forest relies on a random vector sampled independently, and all trees in the forest have the same distribution. The growth of the tree is governed by feature engineering and random forest learning. Random forest is considered one of the most recent and popular boosting methods.

A commercially available automated ASPECTS scoring system (e-ASPECTS; https://brainomix.com/e-aspects) based on a machine learning algorithm has shown an ability to detect early ischemic changes on NCCT at a level similar to that of junior stroke physicians while being noninferior to neuroradiologists.19 That study used ASPECTS on baseline and follow-up CT scans as the ground truth for comparison. e-ASPECTS was further evaluated in a small study of 34 patients in which baseline CT and DWI scans were obtained <2 hours apart.20 Another automated ASPECTS system combining filtering, bi-level and regional growth, feature selection, and a support vector machine was tested on 40 patients with AIS using DWI scans as the ground truth.24 This method obtained a $\kappa$ of 0.52 for dichotomized ASPECTS (>7 versus ≤7). However, the NCCT and diffusion-weighted imaging time was not reported, making it difficult to evaluate its clinical applicability. Other methods of scoring ASPECTS automatically have mostly been tested against expert-read ASPECTS on NCCT. A major strength of this study is the use of an ASPECTS read on acute DWI by an expert as the ground truth to validate the automated ASPECTS method. This assures that the validity of the automated method was tested to a very high standard.

The proposed automated ASPECTS scoring method is based on feature engineering and random forest learning. Random forest is considered one of the most recent and popular boosting methods and has proved classification performance for difficult problems in many medical image–analysis applications compared with other classifiers.27,28 Random forest is an ensemble learning method that combines multiple weak classifiers (decision trees) and lets these decision trees vote for the most popular class. Each tree in the forest relies on a random vector sampled independently, and all trees in the forest have the same distribution. The growth of the tree is governed by feature engineering and random forest learning. Random forest is considered one of the most recent and popular boosting methods.

A commercially available automated ASPECTS scoring system (e-ASPECTS; https://brainomix.com/e-aspects) based on a machine learning algorithm has shown an ability to detect early ischemic changes on NCCT at a level similar to that of junior stroke physicians while being noninferior to neuroradiologists.19 That study used ASPECTS on baseline and follow-up CT scans as the ground truth for comparison. e-ASPECTS was further evaluated in a small study of 34 patients in which baseline CT and DWI scans were obtained <2 hours apart.20 Another automated ASPECTS system combining filtering, bi-level and regional growth, feature selection, and a support vector machine was tested on 40 patients with AIS using DWI scans as the ground truth.24 This method obtained a $\kappa$ of 0.52 for dichotomized ASPECTS (>7 versus ≤7). However, the NCCT and diffusion-weighted imaging time was not reported, making it difficult to evaluate its clinical applicability. Other methods of scoring ASPECTS automatically have mostly been tested against expert-read ASPECTS on NCCT. A major strength of this study is the use of an ASPECTS read on acute DWI by an expert as the ground truth to validate the automated ASPECTS method. This assures that the validity of the automated method was tested to a very high standard.

The proposed automated ASPECTS scoring method is based on feature engineering and random forest learning. Random forest is considered one of the most recent and popular boosting methods and has proved classification performance for difficult problems in many medical image–analysis applications compared with other classifiers.27,28 Random forest is an ensemble learning method that combines multiple weak classifiers (decision trees) and lets these decision trees vote for the most popular class. Each tree in the forest relies on a random vector sampled independently, and all trees in the forest have the same distribution. The growth of the tree is governed by feature engineering and random forest learning. Random forest is considered one of the most recent and popular boosting methods.
random vectors. A measure of randomness is introduced into the training, which can prevent the training classifier from getting stuck at a local minimum, thereby improving the accuracy and reducing the chances of overfitting.

Some previous automated methods have used first-order image features, such as Hounsfield unit (HU) or density and HU difference between the ischemic and contralateral side as features for their algorithms. These first-order image features have limitations in patients with subtle ischemic changes and when images have low signal-to-noise ratios and motion artifacts. The use of multiple higher order computational textural features as part of the machine learning algorithms in the automated ASPECTS method proposed here helps us improve the validity of our technique.

This study has some limitations. First, of the 157 training images randomly selected, only 26.1% (410/1570) of ASPECTS regions had ischemic changes versus 73.9% (1160/1570) normal ASPECTS regions. Improving the performance by tackling the imbalance in data distribution is a goal for our machine learning algorithms. Second, this analysis used imaging data from 1 site. NCCT image acquisition and quality vary across sites; we will, therefore, need to validate the automated ASPECTS method in other data from other sites. Third, only 10 patients had ASPECTS ≤4 of 100 test patients, thus raising some valid concerns about the stability of results in the dichotomized ASPECTS analysis. Validation on a larger dataset is required to demonstrate the robustness of these results. Fourth, only a single atlas was used to localize the ASPECTS regions, which might not be optimal for all patients compared with using a method based on multiple atlases. However, localization based on nonlinear registration using a single atlas can maintain a good trade-off between computational cost and accuracy because saving time is critical in the acute stroke setting. Improving registration accuracy using a single atlas remains an open problem for brain imaging despite some existing atlas-selection techniques.29

CONCLUSIONS

The automated ASPECTS method developed here could accurately and reliably assign ASPECTS on baseline NCCT scans in patients presenting with acute ischemic stroke. This work therefore further validates the utility of machine learning algorithms in developing software that can help and support physicians in interpreting brain scans of patients with acute ischemic stroke.


REFERENCES

19. Nagel S, Sinha D, Day D, et al. e-ASPECTS software is non-inferior to neuroradiologists in applying the ASPECT score to computed
27. Criminisi A, Shotton J. Decision Forests for Computer Vision and Medical Image Analysis. London; Springer-Verlag; 2013