Original Article

Reliability and radiologists' concordance of artificial intelligence (AI)-calculated Alberta Stroke Program Early CT Score (ASPECTS)

Warissara Kiththiworaphongkich, M.D.
Nuttamon Khamwongsa, B.Sc.
Pranruethai Chaimongkol, B.Sc.
From Department of Radiology, Phayao Hospital, Phayao, Thailand.
Address correspondence to W.K. (email: bow_wk@hotmail.com)

Received 19 March 2024; revised 10 December 2024; accepted 10 December 2024
doi:10.46475/asean-jr.v25i3.901

Abstract

Background: ASPECTS was developed for the semi-quantitative assessment of early ischemic changes (EIC) on non-contrast computed tomography (NCCT) in acute ischemic stroke (AIS). Artificial intelligence (AI)-based automated tools for the ASPECT scoring system were developed to automate the diagnosis and improve the agreement with radiologists of AIS. The performance of the automated software compared to physicians should be tested before the software is further used in clinical practice as a tool for clinicians.

Objective: To evaluate the agreement with radiologists of an AI-based automated post-processing software for detecting EIC and calculating ASPECTS on NCCT images in AIS patients using a radiologist's assessment as a reference.

Materials and Methods: NCCT of AIS patients were retrospectively reviewed (Stroke Fast Track Service July 2022 - December 2023). The complete set of clinical data and imaging data from both baseline and follow-up were analyzed by a radiologist as a reference. Two additional observers provided individual ASPECTS from the baseline NCCT only (observer 1 was a radiologist who independently reviewed only the baseline NCCT with stroke window setting. Observer 2 was

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

a radiologist on service which was from the pool of 20 radiologists onsite and online). Recon&GO Inline ASPECTS software (Somaris X, VA40A, Siemens Healthineers AG, Erlangen, Germany) was applied. Both ASPECT score analysis and ASPECTS region analysis were evaluated. Positive percent agreement (PPA) and negative percent agreement (NPA) were calculated. Interobserver agreement was assessed using the Cohen's kappa coefficient and the intraclass correlation coefficient (ICC).

Results: 111 patients with a mean age of 67.8 years (±11.9), 56 (50.5%) females, a mean National Institute of Health Stroke Scale (NIHSS) score of 14.2 (±8.8), and a mean time to baseline NCCT of 123.9 minutes (±58.7) were included. For dichotomized ASPECTS, the automated software showed lower PPA (14.6% vs. 27.1%) but higher NPA (100.0% vs. 93.7%) than observer 2. For the region-based analysis, both the automated software and observer 2 differed in terms of regional contribution. The automated software showed low PPA but rather high NPA with perfect (100%) NPA in lentiform nucleus and M2. The automated software showed higher agreement with the reference and two observers in deep/central regions than cortical regions. For total ASPECTS, the automated software showed a moderate agreement of total ASPECTS with the reference and observer 1 (ICC = 0.545 and 0.545). Observer 2 showed a poor agreement of total ASPECTS with the reference, observer 1, and the automated software (ICC = 0.349, 0.422, and 0.301, respectively).

Conclusion: For total ASPECT score, the agreement of the tested AI software is lower compared to observer 1 obtained by a radiologist using the stroke window on NCCT, but better compared to a pool of radiologists on service with a time limit of 30 minutes to interpret the ASPECT score. When analyzing the ASPECTS regions, there are different advantages for the assessment of the deep regions and the cortical regions. The tested AI software shows higher agreement in deep/central regions than cortical regions. From the result, the tested AI software retains its potential for use in emergency situations, particularly for radiologists with limited experience and limited time to report.

Keywords: Acute ischemic stroke, Alberta Stroke Program Early CT Score, Artificial intelligence, Computed tomography.

Introduction

Ischemic stroke is one of the acute cerebrovascular diseases that is the leading cause of death and severe disabilities worldwide and places an enormous burden on global and national healthcare systems [1, 2]. An acute ischemic stroke (AIS) that is not appropriately interpreted and treated carries a high risk of mortality and disability. Management of an ischemic stroke should be prompt so that appropriate treatment can be initiated and the patient has the best chance of survival. Time is of the essence: the faster the diagnosis is made and the appropriate treatment is initiated, the better the outcome for the patient [3].

In Thailand, the National Health Security Office (NHSO) has supported the free treatment of patients with AIS with thrombolytics since 2008 [4]. Subsequently, the above-mentioned treatment services were expanded into a "Stroke Fast Track" as a stroke service network in all areas of Thailand. This service network aims to ensure that the time from the onset of the patient's neurological symptoms to intravenous thrombolytic treatment is as fast as possible within 270 minutes (4.5 hours) [4]. Recently, the treatment of AIS has developed rapidly due to a highly effective endovascular therapy [5-11]. Endovascular therapy with mechanical thrombectomy (MT) is the current standard of care for patients with AIS and large vessel occlusions (LVO) [12, 13]. Nowadays, the time window for treatment of AIS is generally 6 hours after the onset of the patient's neurological symptoms.

The diagnosis of AIS is based on clinical examination and neuroimaging techniques. Neuroimaging techniques (including CT and MRI) have become an integral approach for the detection and characterization of AIS and the prediction of prognosis. The severity and extent of an ischemic stroke lesion could be used as one of the parameters for the selection of patients for MT [14]. Non-contrast computed tomography (NCCT) of the brain is the most widely used method for assessing the early signs of ischemic brain change in stroke patients. The Alberta Stroke Program Early CT Score (ASPECTS) is a tool for semi-quantitative assessment of early ischemic changes (EIC) on NCCT of patients with AIS in the anterior circulation [15, 16]. Since 2000, the ASPECTS has been widely used to detect the extent of EIC, assess the impact of treatment, and predict the prognosis [17]. The

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

guidelines for early management of AIS recommend the use of the ASPECTS to identify patients suitable for MT. A score of ≥ 6 is an important cut-off value and one of the criteria that qualify patients for MT [13] and thrombolytic treatment. Nevertheless, the early brain imaging changes, including loss of the normal gray-white matter interface and effacement of the cortical sulci, can be subtle and make it particularly difficult for clinicians to recognize them and determine the optimal management in an emergency scenario. ASPECT scoring requires a high level of expertise to detect subtle changes in the NCCT in the early phase of brain ischemia [18]. Expertise is not available in every center where stroke patients are presented.

Recently, artificial intelligence (AI) technology has greatly impacted the field of stroke image analysis by automating diagnosis and improving the diagnostic accuracy of AIS [19]. Many tools for stroke image analysis have been developed to reduce the time needed to detect abnormalities and to provide more accurate results [20]. Several studies reported the use of AI-based automated software versus radiologists for the diagnosis of AIS on CT or MR imaging, such as detection of subtle changes in attenuation on NCCT [21], the automatic ASPECTS scoring system for the AIS area on CT and MRI [22, 23], detection of the hyperdense sign of the middle cerebral artery (MCA) on CT [24] and the automatic quantification of cerebral edema on CT [25]. Recon&GO Inline ASPECTS software, the AI-based automated post-processing ASPECTS software in this study was developed by Siemens. Several studies reported on the use of automated ASPECTS software compared to radiologists. Any automated software may lead to different diagnostic outcomes. However, no study has investigated this software.

The AI-based automated post-processing ASPECTS software has been used in Phayao Hospital since July 2022. However, the performance of the AI-based automated post-processing software should be tested with physicians before this software is further used in clinical practice as a tool for clinicians. Therefore, this study was designed to evaluate the agreement with radiologists of the AI-based automated post-processing software for detecting EIC and calculating the ASPECTS on the NCCT in AIS patients (including patients in the stroke fast track service), using a radiologist's assessment as a reference. Imaging data from the baseline and the follow-up were evaluated as a reference.

Materials and methods

Study participants

Data from consecutive patients who were presented to the stroke fast track service in Phayao Hospital between July 2022 and December 2023 and underwent multimodality brain CT for suspected AIS and met the inclusion criteria were retrospectively identified using the Radiologic Information System and the Picture Archiving and Communication System. The inclusion criteria were: (1) the patient was ≥18 years old, (2) the baseline multimodal NCCT was performed within 270 minutes (4.5 hours) of symptom onset or last seen normal, and (3) the AIS was located in the middle cerebral artery (MCA) territory. Exclusion criteria were: (1) patients with intracranial hemorrhage or other diagnoses explaining the symptoms, such as tumor, infection, ischemic stroke within the anterior cerebral artery (ACA) or posterior circulation, and (2) patients with a previous brain surgery or bilateral old large territorial infarction, and (3) no follow-up CT or CTA performed within 7 days of the baseline NCCT. The detailed flowchart of the patient selection process is shown in Figure 1.

Imaging data and evaluation by radiologists

All patients in this study received the same CT scan protocol, including NCCT scans of the brain on a 32-row detector/192-slice multidetector CT scanner (SOMATOM go.All, Siemens Healthineers AG, Forchheim, Germany). The NCCT scans were acquired in the helical mode with the following parameters: 0.7 mm slice collimation, spiral pitch factor of 0.55, tube voltage of 120 kV, and image matrix 512 x 512. The images were reconstructed in 1 mm overlapping sections. Axial images were reconstructed with a slice thickness of 1 mm and 5 mm. Coronal and sagittal multiplanar reconstructions were performed with a slice thickness of 3 mm.

For the manual ASPECTS assessment, a radiologist (with 9 years of experience) independently reviewed only the baseline NCCT with stroke window setting (window level of 35 Hounsfield units (HU) and width of 30 HU [26]) to create the

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

ASPECT score, hereafter referred to as observer 1. For this assessment, the reader was only aware of the sidedness of neurological deficit and did not have access to the follow-up NCCT.

In addition, each baseline NCCT was regularly assessed by a pool of twenty radiologists on service (two radiologists at Phayao Hospital and eighteen teleradiologists) who were prospectively assigned to read NCCT in the stroke fast-track procedure (hereafter referred to as observer 2). Their results were recorded for analysis, which corresponded to the actual situation.

Eight weeks later, the same radiologist (observer 1) independently performed an ASPECTS assessment using the completed set of available baselines NCCT, follow-up NCCT and/or CTA, and digital subtraction angiography (DSA) as the reference. The reader was blinded to the result of the automated software evaluation. Meeting criteria 1 + 2 or 2 +3 or 1 + 2 + 3 was defined as EIC on baseline NCCT: (1) presence of hypodensity and/or loss of gray-white differentiation, (2) ischemic changeable on follow-up NCCT, (3) Any individual region of hypodensity occupying ≥20% of that region. Existing hyperdense MCA signs were also additionally recorded. The reader was informed of the sidedness of neurological deficit and clinical data.

Automated calculation of ASPECTS

For comparison, the Recon&GO Inline ASPECTS software (Somaris X, VA40A, Siemens Healthineers AG, Erlangen, Germany), was used to analyze the NCCT for EIC in acute stroke in the MCA territory as well as the automatically calculated ASPECTS. The data sets of the post-processed axial NCCT reconstruction of the brain with a slice thickness of 5 mm, a matrix size of 512 x 512 and a smooth kernel reconstruction are prerequisites for the evaluation.

An artificial intelligence (AI) recognizes the patient's landmarks and anatomy. A probabilistic atlas was automatically created to outline the ASPECTS regions (caudate nucleus (C), internal capsule (IC), insula (INS), lentiform nucleus (LN), and 6 regions in the vascular territory of the MCA (M1–M6)). This brain atlas

consists of ten volumes of interest for each hemisphere, representing ten ASPECTS regions. The regions from the atlas are resized and both rigidly and non-rigidly registered to the current dataset. Registration may be incorrect for unusual brain anatomies, such as young patients, large ventricles, old infarcts, open skulls, midline shifts or tumors.

After automatically fitting the atlas to an NCCT of the brain, each voxel is assigned a probability of belonging to a particular ASPECTS region. The average Hounsfield unit (HU) values of the brain tissue voxels of each region are calculated. Voxels that are either too dark or too bright are excluded from the calculation. Cerebrospinal fluid, old infarcts, bones, and calcifications as well as voxels that are either too dark (below 10 HU) or too bright (above 55 HU) are excluded. For regions M1 to M6, the average HU value is mainly based on gray matter. The relative difference in average HU between the single ASPECTS region in the affected hemisphere and the contralateral hemisphere is calculated and presented as a percentage HU difference. The affected cerebral hemisphere is automatically selected by the software. Based on a predefined threshold for the relative HU difference, each ASPECTS region in the affected hemisphere is classified as affected (ischemic changes detected by the software) or unaffected. The affected regions are marked (X). Any calculation result marked with an X indicates a region with lower HU values compared to the other hemisphere. Bold text indicates regions that are visible in the layer. In regions M1 to M6, the white matter may be affected in regions marked with WM. The number of affected regions is used to calculate ASPECTS (without the regions marked with WM). If a region is marked, the ASPECT score is reduced. Below the affected hemisphere, the automated ASPECTS provides a score from 0 (most severe) to 10 (least severe) (Figure 2).

Statistical analysis

All statistical analyses were performed with SPSS (version 26.0.0; IBM, Armonk, NY, USA). Continuous variables are presented as mean ± standard deviation (SD), while categorical data were presented as frequencies (percentages). The performance of the automated ASPECTS and the individual observers (1 and 2) compared to the reference was taken into account: (1) positive percent agreement

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

(PPA) and negative percent agreement (NPA) for each individual ASPECTS region on the correct side and dichotomized ASPECTS (≥6 and <6), (2) interobserver agreement for each individual ASPECTS region and dichotomized ASPECTS (≥6 and <6) were assessed using the Cohen's kappa coefficient. For the Cohen's kappa coefficient, values < 0 indicate poor agreement, values between 0.00 and 0.20 indicate slight agreement, values between 0.21 and 0.40 indicate fair agreement, values between 0.41 and 0.60 indicate moderate agreement, values between 0.61 and 0.80 indicate substantial agreement, and values between 0.81 and 1.00 indicate almost perfect agreement [27], (3) interobserver agreement for total ASPECTS was assessed using the intraclass correlation coefficient (ICC). For the ICC, values < 0.5 indicate poor agreement, values between 0.5 and < 0.75 indicate moderate agreement, values between 0.75 and < 0.9 indicate good agreement, and values between 0.9 and 1.0 indicate excellent agreement [28].

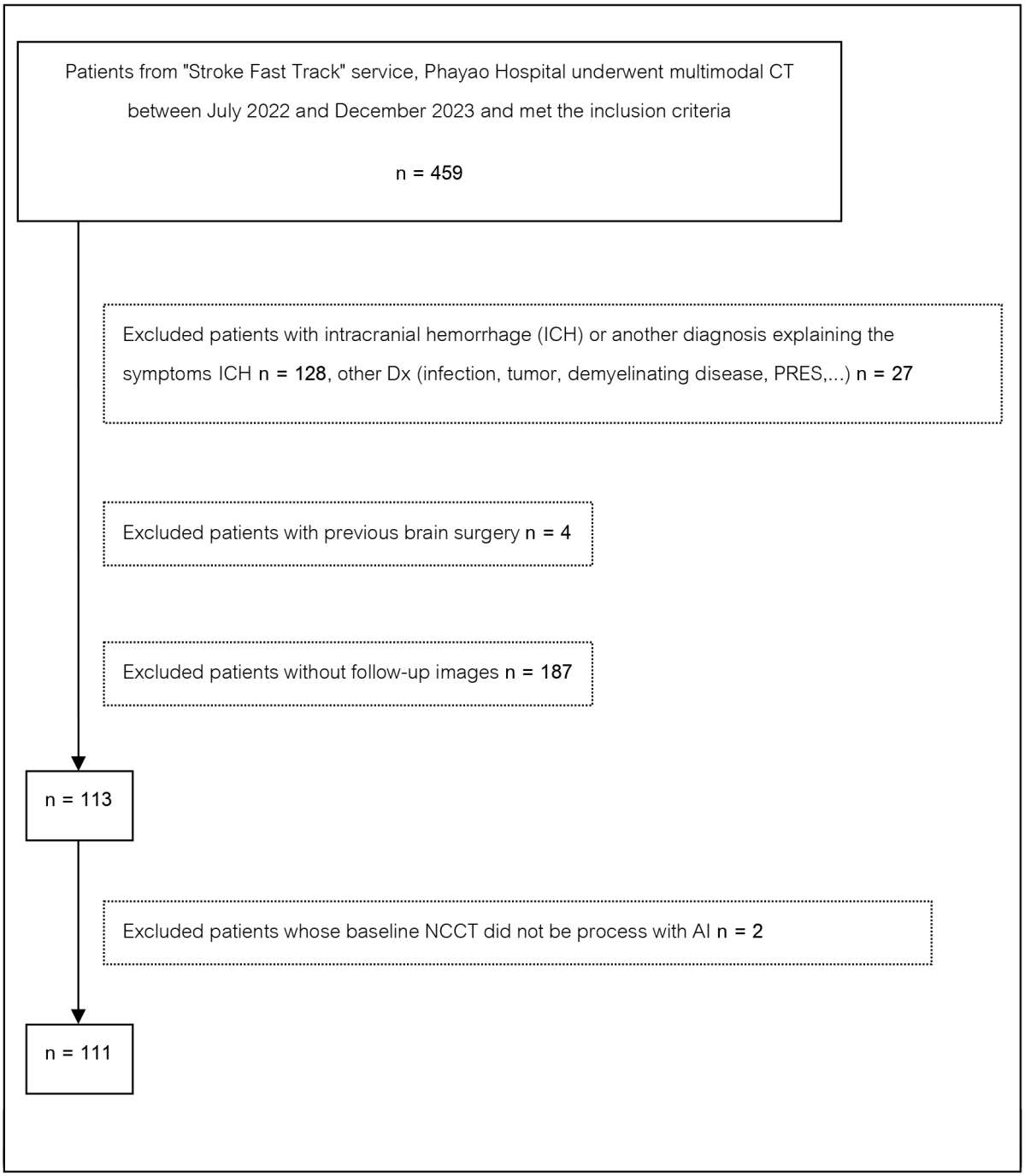

Figure 1. *Flowchart of patients included in the study.*

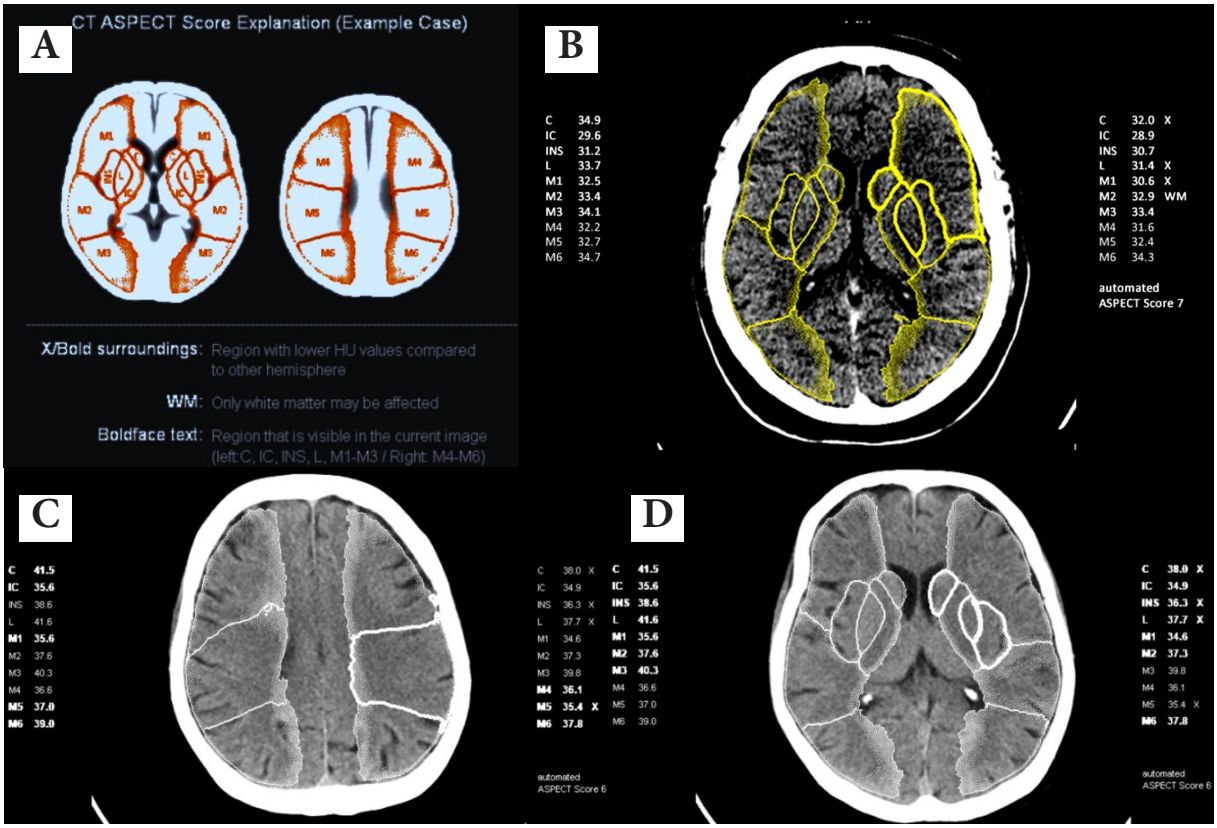

Figure 2. *(A) Example of CT-ASPECT score explanation in the automated software, showing the outlines of ten ASPECTS regions as an overlay on the axial NCCT images based on a quantitative topographic comparison of the left and right hemispheres. (B) Example of axial NCCT images with automatically computed EIC regions calculated with automatic software; the affected hemisphere is automatically selected by the software on the left side. The three affected regions are marked with X. Bold text indicates regions that are visible in the slice image (both hemispheres). Regions marked with WM, white matter, may be affected (are not used to calculate the ASPECTS). The automatically calculated ASPECTS (below the affected hemisphere) was 7. (C and D) Axial NCCT images of the same patient show areas with automatically calculated EIC, visualized within an overlaid white line using automated software. Bold text indicates regions visible in the different slices (both hemispheres). The automatically calculated ASPECTS was 6 in the left cerebral hemisphere. ASPECT scores were based on standard NCCT in the Picture Archiving and Communication System.*

To calculate the ASPECTS, 1 point is subtracted from 10 for each detection of an EIC for each of the 10 defined middle cerebral artery (MCA) vascular territories. C = caudate, IC = internal capsule, INS = insula, L = lentiform nucleus, M1 = frontal operculum, M2 = anterior temporal lobe, M3 = posterior temporal lobe, M4 = anterior MCA, M5 = lateral MCA, M6 = posterior MCA.

Results

General information

Finally, 111 patients were identified retrospectively. Their mean age was 67.8 years (±11.9 years), and 56 (50.5%) were female. The mean NIHSS (National Institute of Health Stroke Scale) score at the presentation was 14.2 (±8.8). Baseline NCCT was performed on average 123.9 minutes after the symptom onset or last seen normal. The clinical characteristics are summarized in Table 1.

Table 1. *Clinical characteristics of the study population.*

Characteristics	N (%) or Mean ± standard deviation
No. of patients	111
Age (years)	67.8 (±11.9)
Gender	
Male	55 (49.5)
Female	56 (50.5)
National Institute of Health Stroke Scale (NIHSS) score at presentation	14.2 (±8.8)
Time from symptom onset to baseline NCCT (minutes)	123.9 (±58.7)
Side of neurological symptoms	
Right	62 (55.9)
Left	49 (44.1)

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

ASPECTS region analysis

The affected cerebral hemisphere in the automated software, observer 1 and observer 2 matched the reference in 53%, 97%, and 57% of the patients, respectively. Nine of the 52 misclassifications in the automated software involved a unilateral cerebral hemisphere and sixteen involved a bilateral cerebral hemisphere. These patients had an unusual brain anatomy; thus, the registration could be incorrect, such as asymmetric ventricle sizes, old infarcts, and asymmetric brain atrophy. Examples of NCCT scans are shown in Figure 3. The remaining scans (27) were classified as normal. Three of the 3 misclassifications in observer1 were classified as normal. Two of the 48 misclassifications in observer 2 were clinically reported on the wrong side of the neurologic symptom at the time of presentation. The remaining scans (46) were classified as normal.

For all ASPECTS regions, observer 1 (radiologist with stroke window setting) showed higher positive percent agreement (PPA) than observer 2 (radiologists on service) and the automated software. However, the performance was slightly lower in the internal capsule, M3, and M6 regions. In the lentiform nucleus and M3, the automated software showed a tendency towards higher PPA and NPA than observer 2. In the caudate, the automated software showed a tendency towards higher PPA, but lower NPA than observer 2. In the internal capsule, M2, M4, and M5, the automated software showed a tendency towards lower PPA, but higher NPA than observer 2. In the insula and M6, the automated software showed a tendency towards lower PPA and NPA than observer 2. In M1, the automated software showed a tendency towards lower PPA but similar NPA compared to observer 2. In the internal capsule, M3 and M6, the automated software and two observers showed a tendency towards low PPA but high NPA. The regions with the lowest PPA of observer 1, observer 2, and the automated software were the internal capsule (61%), M3 (11.8%), and the internal capsule (10.2%), respectively. The full results are shown in Table 2.

For the agreement per region, observer 1 showed the highest agreement with the reference at M1 (0.94; almost perfect agreement) and the lowest agreement at M3 (0.07; slight agreement). The automated software showed the highest agreement

with the reference at caudate (0.50; moderate agreement) and the lowest agreement at M1 (0.05; slight agreement). Observer 2 showed the highest agreement with the reference at M1 (0.35; fair agreement) and the lowest agreement at M3 (0.09; slight agreement). The highest agreement between the automated software and observer 1 was at caudate (0.59; moderate agreement) and the lowest agreement was at M4 (0.06; slight agreement). The highest agreement between the automated software and observer 2 was at lentiform nucleus (0.28; fair agreement) and the lowest agreement was at M6 (-0.02; poor agreement). The highest agreement between observer 1 and observer 2 was at M2 (0.40; fair agreement) and the lowest agreement was at M3 (0.11; slight agreement). The automated software showed a tendency towards higher agreement with the reference and two observers in deep/central regions than cortical regions (slight agreement to moderate agreement in deep/central regions and poor agreement to fair agreement in cortical regions). The regions showed the highest level of overall interobserver agreement were caudate and lentiform nucleus (fair agreement to almost perfect agreement). The region showed the lowest level of overall interobserver agreement was M3 (slight agreement to fair agreement). The complete results are shown in Table 3.

Table 2. *Diagnostic performance of the two observers and automated software for each individual ASPECTS region; the positive percent agreement (PPA) and negative percent agreement (NPA).*

Regions	Positive percent agreement (PPA) % (95% CI)	Negative percent agreement (NPA) % (95% CI)
Deep/Central regions		
Caudate (C)		
observer 1	81.1 (68.6-89.4%)	100.0 (93.8-100.0%)
observer 2	26.4 (16.4-39.6%)	96.6 (88.3-99.1%)
automated software	69.8 (56.5-80.4%)	79.3 (67.2-87.8%)
Internal Capsule (IC)		
observer 1	61.0 (48.3-72.4%)	96.2 (87.0-98.9%)
observer 2	16.9 (9.5-28.5%)	96.2 (87.0-98.9%)
automated software	10.2 (4.7-20.5%)	98.1 (89.9-99.7%)
Insula (INS)		
observer 1	94.4 (86.4-97.8%)	82.5 (68.1-91.3%)
observer 2	40.8 (30.2-52.5%)	97.5 (87.1-99.6%)
automated software	36.6 (26.4-48.2%)	92.5 (80.1-97.4%)
Lentiform nucleus (L)		
observer 1	89.7 (80.2-94.9%)	95.3 (84.5-98.7%)
observer 2	32.4 (22.4-44.2%)	95.3 (84.5-98.7%)
automated software	42.6 (31.6-54.5%)	100.0 (91.8-100.0%)

Cortical regions		
M1		
observer 1	95.5 (84.9-98.7%)	98.5 (92.0-99.7%)
observer 2	40.9 (27.7-55.6%)	91.0 (81.8-95.8%)
automated software	13.6 (6.4-23.8%)	91.0 (81.8-95.8%)
M2		
observer 1	84.9 (73.0-92.2%)	93.1 (83.6-97.3%)
observer 2	39.6 (27.6-53.1%)	91.4 (81.4-96.3%)
automated software	13.2 (6.5-24.8%)	100.0 (93.8-100.0%)
M3		
observer 1	70.6 (46.9-86.7%)	96.8 (91.0-98.9%)
observer 2	11.8 (3.3-34.3%)	94.7 (88.2-97.7%)
automated software	29.4 (13.3-53.1%)	96.8 (91.0-98.9%)
M4		
observer 1	89.5 (78.9-95.1%)	87.0 (75.6-93.6%)
observer 2	35.1 (24.0-48.1%)	85.2 (73.4-92.3%)
automated software	19.3 (11.1-31.3%)	90.7 (80.1-96.0%)
M5		
observer 1	89.2 (79.4-94.7%)	91.3 (79.7-96.6%)
observer 2	33.8 (23.5-46.0%)	87.0 (74.3-93.9%)
automated software	20.0 (12.1-31.3%)	97.8 (88.7-99.6%)
M6		
observer 1	63.0 (44.2-78.5%)	95.2 (88.4-98.1%)
observer 2	22.2 (10.6-40.8%)	92.9 (85.3-96.7%)
automated software	14.8 (5.92-32.5%)	91.7 (83.8-96.0%)

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

Table 3. *Agreement comparison of each individual ASPECTS region between observer 1, observer 2, automated software and reference.*

Regions	Cohen's kappa (95% CI)					
	Software vs. Reference	Software vs. Observer 1	Software vs. Observer 2	Observer 1 vs. Reference	Observer 2 vs. Reference	Observer 1 vs. Observer 2
Deep/Central regions						
Caudate (C)	0.50 (0.33-0.65)	0.59 (0.44-0.74)	0.23 (0.09-0.38)	0.82 (0.71-0.92)	0.24 (0.10-0.37)	0.34 (0.18-0.49)
Internal Capsule (IC)	0.08 (0.00-0.16)	0.18 (0.04-0.32)	0.14 (-0.11-0.40)	0.56 (0.42-0.70)	0.12 (0.02-0.23)	0.33 (0.16-0.50)
Insula (INS)	0.24 (0.11-0.36)	0.18 (0.06-0.30)	0.15 (-0.05-0.34)	0.78 (0.66-0.90)	0.31 (0.19-0.44)	0.25 (0.13-0.37)
Lentiform nucleus (L)	0.34 (0.24-0.49)	0.43 (0.30-0.56)	0.28 (0.08-0.48)	0.83 (0.73-0.94)	0.23 (0.11-0.35)	0.21 (0.08-0.35)
Cortical regions						
M1	0.05 (-.0.08-0.19)	0.10 (-0.04-0.25)	0.09 (-0.10-0.29)	0.94 (0.88-1.00)	0.35 (0.18-0.52)	0.36 (0.19-0.53)
M2	0.14 (0.04-0.23)	0.16 (0.05-0.27)	0.16 (-0.02-0.35)	0.78 (0.67-0.90)	0.32 (0.16-0.47)	0.40 (0.25-0.56)
M3	0.34 (0.08-0.59)	0.38 (0.11-0.64)	0.07 (-0.17-0.32)	0.07 (-0.07-0.09)	0.09 (-0.12-0.29)	0.11 (-0.11-0.33)
M4	0.10 (-0.02-0.23)	0.06 (-0.07-0.18)	0.22 (0.02-0.42)	0.77 (0.65-0.89)	0.20 (0.04-0.36)	0.26 (0.11-0.41)
M5	0.15 (0.06-0.25)	0.21 (0.10-0.31)	0.08 (-0.10-0,27)	0.80 (0.68-0.91)	0.19 (0.05-0.33)	0.22 (0.07-0.36)
M6	0.08 (-0.10-0.26)	0.07 (-0.13-0.26)	-0.02 (-0.19-0.16)	0.63 (0.45-0.80)	0.19 (-0.01-0.38)	0.12 (-0.09-0.33)

ASPECT score analysis

In the score analysis, the mean (±SD) and range of ASPECTS scores of the reference, observer 1, observer 2, and automated software were 5.37 (±3.0) and 0-10, 5.79 (±2.7) and 0-10, 8.14 (±2.8) and 0-10, and 8.36 (±1.7) and 2-10, respectively.

For dichotomized ASPECTS (≥6 and <6), the misclassifications (false high ASPECTS; underestimation) for observer 1, observer 2, and the automated software were eight out of 70, thirty-five out of 94, and forty-one out of 104, respectively. The misclassifications (false low ASPECTS; overestimation) for observer 1, observer 2, and the automated software were one out of 41, four out of 17, and 0 out of 7, respectively.

The PPA and NPA of the observer 1 were 83.3% and 98.4%, respectively. The PPA and NPA of the observer 2 were 27.1% and 93.7% respectively. The PPA and NPA of the automated software were 14.6% and 100.0% respectively. Observer 1 showed higher PPA than observer 2 and the automated software. However, observer 1 showed a tendency towards lower NPA than the automated software. The automated software also showed a tendency towards lower PPA but higher NPA than observer 2. The complete results are shown in Table 4.

For dichotomized ASPECTS (≥6 and <6), the agreement between the reference and observer 1 was 0.83, indicating almost perfect agreement. Levels of agreement between the reference and the automated software as well as observer 1 and the automated software were both 0.16, indicating slight agreement. Levels of the agreement between observer 2 and the reference, observer 1 and the automated software were 0.23, 0.21, and 0.27, indicating fair agreement. The complete results are shown in Table 5.

The agreement of the overall ASPECTS between the reference and observer 1 was greater than 0.90, indicating excellent agreement. The agreement of the overall ASPECTS between the reference and the automated software as well as observer 1 and the automated software were between 0.5 and 0.75 (ICC 0.545), indicating moderate agreement. The agreements of the overall ASPECTS between observer 2 and the reference, observer 1 and the automated software were all below 0.5, indicating poor agreement. The complete results are shown in Table 6.

Table 4. *Diagnostic performance of the two observers and automated software for dichotomized ASPECTS; the positive percent agreement (PPA) and negative percent agreement (NPA).*

Dichotomized ASPECTS ≥6 and <6	Positive percent agreement (PPA) % (95% CI)	Negative percent agreement (NPA) % (95% CI)
observer 1	83.3 (69.8-92.5%)	98.4 (91.5-99.9%)
observer 2	27.1 (15.3-41.9%)	93.7 (84.5-98.2%)
automated software	14.6 (6.1-27.8%)	100.0 (94.3-100.0%)

Table 5. *Agreement comparison of dichotomized ASPECTS between observer 1, observer 2, automated software and reference.*

	Cohen's kappa (95% CI)					
	Software vs. Reference	Software vs. Observer 1	Software vs. Observer 2	Observer 1 vs. Reference	Observer 2 vs. Reference	Observer 1 vs. Observer 2
Dichotomized ASPECTS ≥6 and <6	0.16 (0.05-0.27)	0.16 (0.03-0.29)	0.27 (0.02-0.52)	0.83 (0.73-0.94)	0.23 (0.07-0.37)	0.21 (0.04-0.38)

Table 6. *Agreement of total ASPECTS between observer 1, observer 2, automated software and reference.*

	With reference ICC (95% CI)	With observer 1 ICC (95% CI)	With observer 2 ICC (95% CI)	With automated software ICC (95% CI)
Observer 1	0.932 (0.902-0.953)		0.422 (0.256-0.564)	0.545 (0.399-0.664)
Observer 2	0.349 (0.174-0.503)	0.422 (0.256-0.564)		0.301 (0.121-0.462)
Automated software	0.545 (0.399-0.664)	0.545 (0.399-0.664)	0.301 (0.121-0.462)	

ICC, Intraclass correlation coefficient

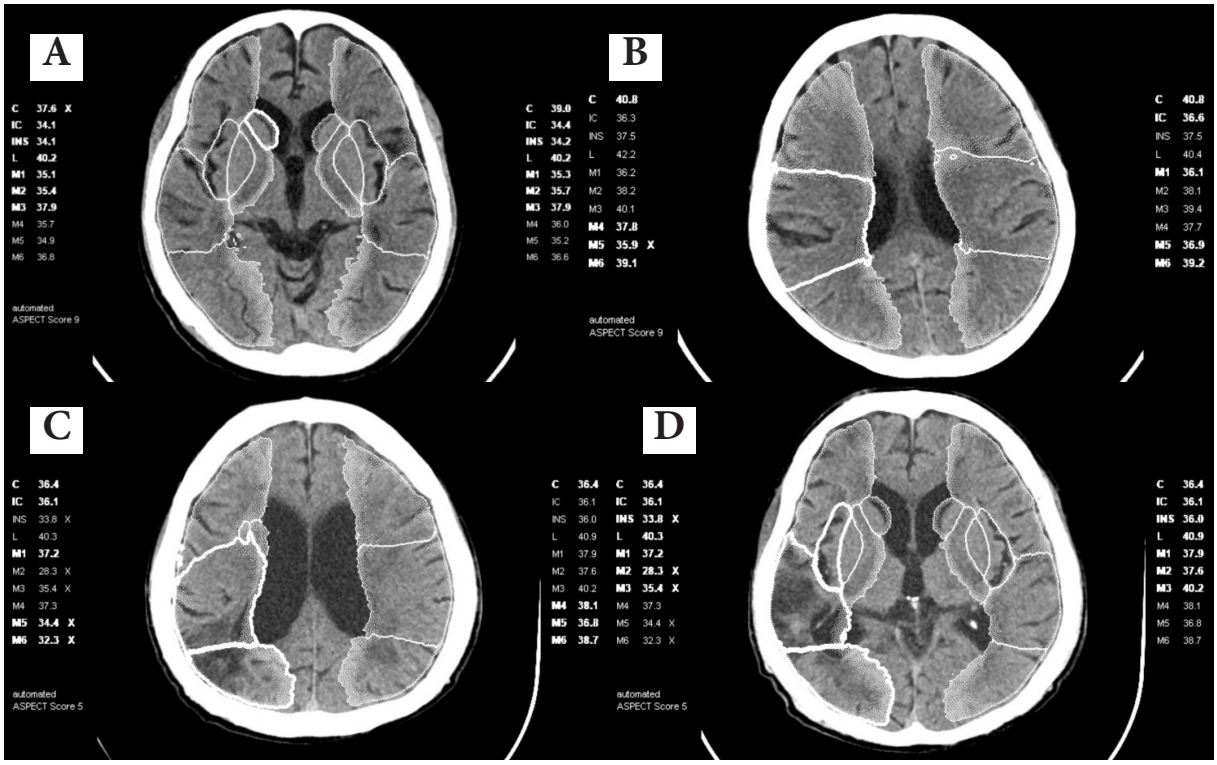

Figure 3. *Example of NCCT scans with unusual brain anatomy leading to misclassification of the affected hemisphere.*
(A) Asymmetric size of bilateral lateral ventricles - the right lateral ventricle was slightly larger than the left lateral ventricle, leading to a false positive result in the right caudate. The final reference for EIC was the left insula (ASPECTS = 9).
(B) Asymmetric gyral atrophy - prominent right frontoparietal sulci, leading to a false positive result at the right M5. The final reference for EIC was at the left insula, left lentiform nucleus, left M4 and left M5 (ASPECTS = 6).
(C and D) Old infarcts in the same patient - chronological change of late subacute infarcts in the right parieto-occipital lobe, leading to false positive findings in M2, M3, M5 and M6 of the right cerebral hemisphere. Prominent sulci of the right insula, leading to a false positive result in the right insula. The final reference for EIC was the left internal capsule (ASPECTS = 9).

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

Discussion

This study evaluated the performance of the automated ASPECTS software for the detection of EIC on the NCCT of the brain. This study showed moderate agreement between the automated software and the reference for total ASPECTS and slight agreement between the automated software and the reference for dichotomized ASPECTS (≥6 and <6). The agreement was similar to the agreement between the automated software and an individual reader (the observer 1).

The automated ASPECTS software has several advantages. The software can exclude voxels that are either too dark or too bright from the calculation before evaluating the mean density of the ASPECTS regions. Excluding these voxels, such as bone, calcifications, CSF or old brain infarcts, improves the accuracy of brain parenchymal density measurements. Automated software can automatically create the lines that divide the MCA territory into ten regions of interest based on quantitative topography as an overlay on each axial image of the NCCT. These lines for the boundary of each ASPECTS region appear when scrolling the NCCT, even if the window settings are changed. Originally, the ASPECTS was calculated within two predefined slices through the level of the basal ganglia and level of the supra-ganglionic structures [15], while several automated software programs now integrate the whole brain scan. These highlighted lines as an overlay on the NCCT could help in the visual evaluation of the NCCT to detect and quantify EIC. The drawback is the discrepancy in the assessment of the affected hemisphere in patients with an unusual brain anatomy is too large to affect the calculation of the average HU in each ASPECTS region, or the HU is not dark enough to be excluded before calculating the average HU. Two studies using other automated software have shown that brain atrophy and atypical brains affect the segmentation accuracy of each ASPECTS region [29, 30]. In this study, the latter cannot be corrected manually, which leads to an incorrect assessment of the ASPECTS.

Of note, the underestimation of dichotomized ASPECTS by the automated software totaled forty-one of 104, thirty-five of 94 by observer 2 (radiologists on service), and eight of 70 by observer 1 (radiologist with stroke window setting).

The underestimation of ASPECTS was probably caused by the limited calculation of the average HU in each ASPECTS region. Probable causes included changes so subtle or areas so small that they could not affect the calculation of the average HU, and the assessment of cortical regions based mainly on gray matter. On the other hand, the automated software found no overestimation compared to the two observers (one out of 41 by observer 1 and four out of 17 by observer 2). Therefore, if the automated ASPECTS are less than 6, an NCCT should be reviewed. If the NCCT shows no unusual brain anatomy, the findings could be true positives.

The relatively poor positive percent agreement (PPA) for dichotomized ASPECTS and poor agreement for total ASPECTS of observer 2 in this study could be partly explained by a pool of multiple radiologists, as almost all radiologists were teleradiologists. The teleradiologists received the clinical information from clinicians. The interpretation of NCCT by individual radiologists in the diagnosis of AIS might be affected by various practical conditions and limitations. The possible causes included: 1) analyzing the imaging information took time (the report production time for the stroke fast-track at Phayao Hospital was within 30 minutes of examination of the NCCT), 2) not all radiologists were trained to analyze and interpret the ASPECTS, resulting in varying levels of imaging expertise, 3) quantitative measurements of acute infarction on the NCCT were difficult in a routine clinical practice because the signs of infarction were more subtle and human judgment was highly variable. 4) there were some cases where teleradiologists were told the wrong side of the neurological symptoms on presentation, leading to incorrect classifications. The correct NCCT interpretation of a patient with an AIS prior to thrombolysis or MT therefore requires training and experience. Nevertheless, their results were recorded for analysis, which corresponded to the actual situation.

The underestimation of dichotomized ASPECTS by observer 2 with a total of thirty-five out of 94 was compared to eight out of 70 by observer 1. Eight patients received the wrong side of neurological symptoms at the time of presentation. Two out of eight were classified the wrong affected cerebral hemisphere on the NCCT and the remaining six were classified as normal. Four of these eight patients

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

were underestimated. One of the two patients with underestimation missed the hyperdense clot in the proximal middle cerebral artery (MCA). In addition, observer 2 missed the hyperdense MCA sign on the NCCT in a total of twenty-five out of 58 patients. The hyperdense MCA signs were also recorded in the reference. The automated software in this study cannot evaluate this sign. The hyperdense vessel sign (HDVS) on the NCCT is an early marker for AIS caused by occlusion of the intracranial artery. HDVS is most commonly reported in the MCA region, as the MCA territory is usually the brain region most affected by an ischemic stroke [31]. Several studies concluded that HDVS has a high specificity and a high sensitivity for detecting LVO on the NCCT in AIS patients [32-35] with an NIHSS of more than 10 and suspected MCA occlusion (M1 segment) [33]. The density of the pathologic MCA should be more than 43 HU and 1.2 times higher than the contralateral MCA [35]. Due to its early visibility and before the appearance of pathological ischemic changes of the parenchyma, it represents a diagnostic aid in time-critical acute strokes. Automating the process of detecting MCA or intracranial HDVS signs in emergency imaging could speed up the identification of positive cases, especially in centers without regular access to CTA or MRI. This could shorten the time to execute acute treatment such as systemic thrombolysis or MT. Takahashi et al. published a study in 2013 on automated software to detect the HDVS sign on CT [24]. Considering the ongoing improvement of this automated software in the future, the detection of hyperdense MCA signs using HU is needed to improve diagnostic accuracy. However, the correct side of neurological symptoms at the time of presentation from the physician is very important for diagnostic imaging in patients with acute stroke, especially in AIS which shows rather subtle changes in density on the initial NCCT.

In this study, the performance was analyzed for each ASPECTS region. When analyzing the ASPECTS regions, the automated software and observer 2 showed greater differences in PPA than in negative percent agreement (NPA) for the detection of EIC in both cortical and deep regions. The automated software and the observer 2 showed low PPA but high NPA for the detection of EIC in all ASPECTS regions. However, the automated software and the observer 2 did not perform higher PPA than the observer 1 in detecting EIC in all ASPECTS regions.

For the internal capsule (IC), M3, and M6, two observers and the automated software showed lower PPA than the other regions, especially for IC. The remaining regions evaluated by observer 1 showed high PPA and NPA. The automated software showed a tendency towards higher agreement with the reference and the two observers in deep/central regions than cortical regions. The regions showed the highest level of overall interobserver agreement was caudate and lentiform nucleus (gray matter). The regions showed the lowest level of overall interobserver agreement was M3 and M6 (white matter). IC is a white matter and has a small volume and an indistinct boundary, which makes it difficult for humans to recognize the hypodensity and makes the outline of the ASPECTS region inappropriate for the automated software. M3 and M6, as well as the other cortical regions, are limited for EIC detection in the NCCT due to their subtle hypodensity [36]. It is difficult for automated software to precisely outline cortical regions. Furthermore, the average HU value in automated software is mainly based on gray matter. In regions M1 to M6, white matter may be affected in the regions labeled WM, but this is not taken into account in the automated ASPECTS calculation, resulting in an underestimation. In caudate, the automated software showed significantly higher PPA but lower NPA compared to observer 2. From these results, it appeared that the automated software had the highest PPA in caudate, while the overall NPA was good. The fluctuations in the results of the automated software and of observer 2 could be due to an inconsistent definition of the region boundaries [37, 38]. The automated software applied a standardized atlas method to outline ASPECTS regions based on a quantitative topographic comparison of the left and right cerebral hemisphere. Brain atrophy and an unusual brain anatomy also affected segmentation accuracy [29, 30], not only the ASPECTS calculation. Previous studies with other automated ASPECTS software showed variability in the characteristics of the regions scored [36-38]. Even if the automated segmentation of the ASPECTS region was not very precise, the high performance of observer 1 could improve the diagnostic accuracy in this situation. A combination of automated overlay of ASPECTS regions on NCCT images and stroke window settings is recommended for EIC detection. They show that the automated software overlays the boundaries of the regions in real time to detect subtle hypodensity in the same axial NCCT. However, future studies are needed to investigate the ASPECTS segmentation accuracy of the automated software.

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

Another problem in the evaluation of automated software is the choice of reference standard. The reference standard has a direct impact on the accuracy of the ASPECTS assessment. The reference standard used to assess the accuracy and reliability of automated ASPECTS software has varied in previous studies [38-49]. Conclusions may vary based on the reference standard used. Due to the subtle changes in the AIS, using the expert consensus score from the baseline NCCT as the reference standard may lead to an overestimation of the ASPECTS score [45-47]. Due to the time delay between the baseline NCCT examination and reperfusion, the ASPECTS, which was determined using follow-up images as a reference standard, may be underestimated [37, 48, 49]. To define the most accurate reference standard for the ASPECTS, it is advisable to use the baseline of the NCCT and the follow-up images [38, 39, 41-45]. This is consistent with this study, in which all available NCCT baseline examinations, NCCT follow-up examinations and/or CTA and digital subtraction angiography (DSA) were used for assessment. However, both the reference and the blinded ASPECTS assessment analysis (the observer 1) were performed by the same reader at different time points. To minimize a possible mutual influence of the two analysis runs, a long interval (eight weeks) was chosen between the blinded ASPECTS assessment analysis (observer 1) and the reference so that the reader could no longer remember the original measurement. In addition, sensitivity and specificity implied that the new test was compared to the truth or gold standard. Even though the concisely available reference was established, the reference in this study was not definitely proven the gold standard. Instead, the term of PPA, NPA and agreement were used.

There are only a few studies in which the automated software from Siemens Healthineers has been tested. The first study was published by Goebel J et al. comparing the performance of syngo.via Frontier ASPECTS Prototype V1.2.0 with the e-ASPECTS software [50]. The authors found high agreement in ASPECTS grading between two board-certified radiologists, i.e. consensus reading of NCCT images by experts, and e-ASPECTS. However, it showed only low to moderate agreement to Frontier- ASPECTS by Siemens. The second study was published by Hoelter P et al. evaluating three fully automated software applications for ASPECTS (Syngo.via Frontier ASPECT Score Prototype V2, Brainomix

e-ASPECTS® and RAPID ASPECTS) in AIS patients [51]. The authors found a high correlation between the expert and three software programs. The highest correlation was between the expert and Brainomix (r = 0.871 (0.818, 0.909), p < 0.001), followed by the correlation between the expert and Frontier V2 (r = 0.801 (0.719, 0.859), p < 0.001) and between the expert and RAPID (r = 0.777 (0.568, 0.871), p < 0.001). There was a high correlation between the software tools (Frontier V2 and Brainomix: r = 0.830 (0.760, 0.880), p < 0.001; Frontier V2 and RAPID: r = 0.847 (0.693, 0.913), p < 0.001; Brainomix and RAPID: r = 0.835 (0.512, 0.923), p < 0.001). The third study was published by Wolff L et al. who evaluated the syngo.via Frontier ASPECTS Prototype V2.0.1 [42]. The authors found that the performance of the automated ASPECTS was comparable to that of experts and could assist readers in the detection of EIC. The automated ASPECTS post-processing software in this study was developed based on Syngo via Frontier ASPECTS Prototype for commercial use. Despite the fact that there are many studies on the use of the automated ASPECTS software, each automated software may lead to different diagnostic results, but no study investigated the Recon&GO Inline ASPECTS software. Further studies are needed to compare the performance of this commercial version of the software with that of other software.

Some studies that used automated software to analyze the ASPECTS score level showed similar results. Temmen SE, et al. found that the performance of the automated software as Canon medical system base version 1.3 for LVO detection was lower than that of the radiologist in 100 patients [44]. Sawicki M, et al. found that the performance of e-CTA (a part of e-STROKE, version 10.1p3, Brainomix Ltd.) was acceptable for the detection of proximal LVOs in 108 patients, while it did not seem to be accurate enough for distal LVOs compared to the neuroradiologist [41].

Studies have used other automated software for ASPECTS analysis at the regional level that come to similar results. Austein F, et al. found moderate agreement between two automated software packages and expert consensus for overall ASPECTS in 52 patients and no significant difference in the overall performance. However, the software packages differed in terms of regional contribution. Package

THE ASEAN JOURNAL OF RADIOLOGY
ISSN 2672-9393

A was more sensitive in cortical areas than the other methods but at the expense of specificity. Expert consensus and software package B had higher sensitivity but lower specificity for deep brain structures. The variability of results at the region level may be due to inconsistent definitions of region boundaries, especially for IC and M4-6 [37]. Chen Z, et al, evaluated two automated ASPECTS software (The RAPID (iSchemaView, software version 5.0) and NBC (NeuBrainCARE software version 1.0)) compared to radiologists in 276 patients. The authors found greater differences in sensitivity than in specificity and accuracy. NBC showed higher specificity and lower sensitivity, while RAPID showed lower specificity and higher sensitivity for detecting EIC in deep regions. NBC showed lower specificity and higher sensitivity, while RAPID showed higher specificity and lower sensitivity for the detection of EIC in cortical regions. NBC and RAPID did not perform better than experienced radiologists in detecting EIC in cortical regions [38]. Future studies are needed to investigate the ASPECTS segmentation accuracy of the automated software. In addition, further studies are needed to evaluate the time dependence of the ASPECTS analysis at the regional level.

There are several limitations to this study. First, it was a retrospective study collected from the Radiologic Information System and the Picture Archiving and Communication System. Second, the study group was rather small. Third, this study used imaging data from a single center to test only one vendor-specific AI software. Fourth, this study used the assessment of a single reader as a reference and individual observer. Fifth, the time interval between the baseline NCCT and the follow-up images varied within a week. Sixth, there were different times for interpreting the ASPECTS between the two observers in which observer 2 limited the time to 30 minutes, while the observer 1 did not limit the time for interpreting the NCCT.

Conclusion

This study aimed to evaluate the agreement with radiologists of an AI-based automated post-processing software for the detection of EIC and the calculation of ASPECTS on NCCT images in AIS patients. The study results showed that the tested AI software had lower agreement for the total ASPECTS than the radiologist-determined observer 1 when using the stroke window on the NCCT, but better than a pool of radiologists on service. For dichotomized ASPECTS (≥6 and <6), the tested AI software showed perfect (100%) NPA. When analyzing the ASPECTS regions, there were different advantages for scoring the deep regions and the cortical regions. The tested AI software showed low PPA but rather high NPA with perfect (100%) NPA in lentiform nucleus and M2. The tested AI software also showed higher agreement with the reference and two observers in deep/central regions than cortical regions. From the results, the tested AI software retained its potential for use in emergency situations, particularly for radiologists with limited experience and limited time to report. The high NPA of automated software on dichotomized ASPECTS can help provide immediate results. The future utility of the automated ASPECTS software can be used in conjunction with an experienced reader or radiologist on service to validate the final score and identify artifacts or technical issues that may lead to an over- or underestimation of the actual ASPECT score. Due to the high performance of the observer 1, an experienced reader or radiologist could use the automatically highlighted marker as an overlay on NCCT images from the automated software along with the stroke window setting during the interpretation of fast-track NCCT scans to improve the accuracy of EIC detection and ASPECTS calculation.

Funding: The author states that this work did not receive any financial support.

Institutional Review Board Statement: The study protocol was approved by the research ethics committee of Phayao Hospital (COA number 257). The informed consent was waived because of the retrospective study design.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Soun JE, Chow DS, Nagamine M, Takhtawala RS, Filippi CG, Yu W, et al. Artificial intelligence and acute stroke Imaging. AJNR Am J Neuroradiol 2021;42:2-11. doi: 10.3174/ajnr.A6883.

2. Boehme AK, Esenwa C, Elkind MS. Stroke risk factors, genetics, and prevention. Circ Res 2017;120:472-95. doi: 10.1161/CIRCRESAHA.116.308398.

3. Emberson J, Lees KR, Lyden P, Blackwell L, Albers G, Bluhmki E, et al. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. Lancet 2014;384:1929-35. doi: 10.1016/S0140-6736(14)60584-5.

4. Tiamkao S. [13 years - learning and development of Stroke Fast Track in Thailand]. J Health Syst Res 2023;17:191-99. Thai.

5. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. N Engl J Med 2015;372:11-20. doi: 10.1056/NEJMoa1411587.

6. Campbell BC, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. N Engl J Med 2015;372:1009-18. doi: 10.1056/NEJMoa1414792.

7. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. N Engl J Med 2018;378:708–18. doi: 10.1056/NEJMoa1713973.

8. Saver JL, Goyal M, Bonafe A, Diener HC, Levy EI, Pereira VM, et al. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. N Engl J Med 2015;372:2285-95. doi: 10.1056/NEJMoa1415061.

9. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. N Engl J Med 2015;372:1019-30. doi: 10.1056/NEJMoa1414905.

10. Jovin TG, Chamorro A, Cobo E, de Miquel MA, Molina CA, Rovira A, et al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. N Engl J Med 2015;372:2296-306. doi: 10.1056/NEJMoa1503780.

11. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and Infarct. N Engl J Med 2018;378:11-21. doi: 10.1056/NEJMoa1706442.

12. Turc G, Bhogal P, Fischer U, Khatri P, Lobotesis K, Mazighi M, et al. European Stroke Organisation (ESO) - European Society for Minimally Invasive Neurological Therapy (ESMINT) guidelines on mechanical thrombectomy in acute ischemic stroke. J Neurointerv Surg 2023;15:e8. doi: 10.1136/neurintsurg-2018-014569.

13. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, 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-e418. doi: 10.1161/STR.0000000000000211.

14. Venema E, Mulder MJHL, Roozenbeek B, Broderick JP, Yeatts SD, Khatri P, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: development and validation of a clinical decision tool in two randomized trials. BMJ 2017;357:j1710. doi: 10.1136/bmj.j1710.

15. 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 Programmed Early CT Score. Lancet 2000;355:1670-4. doi: 10.1016/s0140-6736(00)02237-6.

16. Pexman JH, Barber PA, Hill MD, Sevick RJ, Demchuk AM, Hudon ME, et al. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke. AJNR Am J Neuroradiol 2001;22:1534-42.

17. Prakkamakul S, Yoo AJ. ASPECTS CT in acute ischemia: review of current data. Top Magn Reson Imaging 2017;26:103-12. doi: 10.1097/RMR.0000000000000122.

18. Herweh C, Ringleb PA, Rauch G, Gerry S, Behrens L, Möhlenbruch M, 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. doi: 10.1177/1747493016632244.

19. Lee EJ, Kim YH, Kim N, Kang DW. Deep into the brain: artificial intelligence in stroke imaging. J Stroke 2017;19:277-85. doi: 10.5853/jos.2017.02054.

20. Mokli Y, Pfaff J, Dos Santos DP, Herweh C, Nagel S. Computer-aided imaging analysis in acute ischemic stroke – background and clinical applications. Neurol Res Pract 2019;1:23. doi: 10.1186/s42466-019-0028-y.

21. Tang FH, Ng DK, Chow DH. An image feature approach for computer-aided detection of ischemic stroke. Comput Biol Med 2011;41:529-36. doi: 10.1016/j.compbiomed.2011.05.001.

22. Chen L, Bentley P, Rueckert D. Fully Automatic Acute Ischemic Lesion Segmentation in DWI Using Convolutional Neural Networks. NeuroImage Clin 2017;15:633–43. doi: 10.1016/j.nicl.2017.06.016.

23. Kuang H, Najm M, Chakraborty D, Maraj N, Sohn SI, Goyal M, et al. Automated ASPECTS on noncontrast CT scans in patients with acute ischemic stroke using machine learning. AJNR Am J Neuroradiol 2019;40:33-8. doi: 10.3174/ajnr.A5889.

24. Takahashi N, Lee Y, Tsai DY, Matsuyama E, Kinoshita T, Ishii K. An automated detection method for the MCA dot sign of acute stroke in unenhanced CT. Radiol Phys Technol 2014;7:79-88. doi: 10.1007/s12194-013-0234-1.

25. Chen Y, Dhar R, Heitsch L, Ford A, Fernandez-Cadenas I, Carrera C, et al. Automated quantification of cerebral edema following hemispheric Infarction: Application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. NeuroImage Clin 2016;12:673–80. doi: 10.1016/j.nicl.2016.09.018.

26. Mainali S, Wahba M, Elijovich L. Detection of early ischemic changes in noncontrast CT head improved with "Stroke Windows". ISRN Neurosci 2014; 2014:654980. doi: 10.1155/2014/654980.

27. Landis JR, Koch GG. The measurement of Observer agreement for categorical data. Biomentrics 1977; 33:159-74.

28. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63. doi: 10.1016/j.jcm.2016.02.012.

29. Shieh Y, Chang CH, Shieh M, Lee TH, Chang YJ, Wong HF, et al. Computer-aided diagnosis of hyperacute stroke with thrombolysis decision support using a contralateral comparative method of CT image analysis. J Digit Imaging 2014 ;27:392-406. doi: 10.1007/s10278-013-9672-x.

30. Stoel BC, Marquering HA, Staring M, Beenen LF, Slump CH, Roos YB, et al. Automated brain computed tomographic densitometry of early ischemic changes in acute stroke. J Med Imaging (Bellingham) 2015;2:014004. doi: 10.1117/1.JMI.2.1.014004.

31. Ng YS, Stein J, Ning M, Black-Schaffer RM. Comparison of clinical characteristics and functional outcomes of ischemic stroke in different vascular territories. Stroke 2007;38:2309-14. doi: 10.1161/STROKEAHA.106.475483.

32. Mair G, Boyd EV, Chappell FM, von Kummer R, Lindley RI, Sandercock P, et al. Sensitivity and specificity of the hyperdense artery sign for arterial obstruction in acute ischemic stroke. Stroke 2015;46:102-7. doi:10.1161/STROKEAHA.114.007036.

33. Lim J, Magarik JA, Froehler MT. The CT-defined hyperdense arterial sign as a marker for acute intracerebral large vessel occlusion. J Neuroimaging 2018;28:212-6. doi: 10.1111/jon.12484.

34. Jensen-Kondering U, Riedel C, Jansen O. Hyperdense artery sign on computed tomography in acute ischemic stroke. World J Radiol 2010;2:354-7. doi: 10.4329/wjr.v2.i9.354.

35. Koo CK, Teasdale E, Muir KW. What constitutes a true hyperdense middle cerebral artery sign? Cerebrovasc Dis 2000;10:419-23. doi: 10.1159/000016101.

36. Chiang PL, Lin SY, Chen MH, Chen YS, Wang CK, Wu MC, et al. Deep learning-based automatic detection of ASPECTS in acute ischemic stroke: improving stroke assessment on CT scans. J Clin Med 2022;11:5159. doi: 10.3390/jcm11175159.

37. Austein F, Wodarg F, Jürgensen N, Huhndorf M, Meyne J, Lindner T, 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. doi: 10.1007/s00330-019-06252-2.

38. Chen Z, Shi Z, Lu F, Li L, Li M, Wang S, et al. Validation of two automated ASPECTS software on non-contrast computed tomography scans of patients with acute ischemic stroke. Front Neurol 2023;14:1170955. doi: 10.3389/fneur.2023.1170955.

39. Neuhaus A, Seyedsaadat SM, Mihal D, Benson JC, Mark I, Kallmes DF, et al. Region-specific agreement in ASPECTS estimation between neuroradiologists and e-ASPECTS software. J Neurointerv Surg 2020;12:720-3. doi: 10.1136/neurintsurg-2019-015442.

40. Naganuma M, Tachibana A, Fuchigami T, Akahori S, Okumura S, Yi K, et al. Alberta Stroke Program Early CT Score calculation using the deep learning-based brain hemisphere comparison algorithm. J Stroke Cerebrovasc Dis 2021;30:105791. doi: 10.1016/j.jstrokecerebrovasdis.2021.105791.

41. Sawicki M, Safranow K, Wiska L, Pasek I, Gajdziel A, Gruszewski M, et al. Diagnostic value of artificial intelligence—based software in detection of large vessel occlusion in acute ischemic stroke. Appl Sci 2021;11:10017. https://doi.org/10.3390/app112110017.

42. Wolff L, Berkhemer OA, van Es ACGM, van Zwam WH, Dippel DWJ, Majoie CBLM, et al. Validation of automated Alberta Stroke Program Early CT Score (ASPECTS) software for detection of early ischemic changes on non-contrast brain CT scans. Neuroradiology 2021;63:491-8. doi: 10.1007/s00234-020-02533-6.

43. Maegerlein C, Fischer J, Mönch S, Berndt M, Wunderlich S, Seifert CL, et al. Automated calculation of the Alberta Stroke Program Early CT Score: feasibility and reliability. Radiology 2019;291:141-8. doi: 10.1148/radiol.2019181228.

44. Temmen SE, Becks MJ, Schalekamp S, van Leeuwen KG, Meijer FJA. Duration and accuracy of automated stroke CT workflow with AI-supported intracranial large vessel occlusion detection. Sci Rep 2023;13:12551. doi: 10.1038/s41598-023-39831-x.

45. Löffler MT, Sollmann N, Mönch S, Friedrich B, Zimmer C, Baum T, et al. Improved reliability of automated ASPECTS evaluation using iterative model reconstruction from head CT scans. J Neuroimaging 2021;31:341-7. doi: 10.1111/jon.12810.

46. McTaggart RA, Jovin TG, Lansberg MG, Mlynash M, Jayaraman MV, Choudhri OA, 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. doi: 10.1161/STROKEAHA.114.006564.

47. Kuang H, Najm M, Chakraborty D, Maraj N, Sohn SI, Goyal M, et al. Automated ASPECTS on noncontrast CT scans in patients with acute ischemic stroke using machine learning. AJNR Am J Neuroradiol 2019;40:33-8. doi: 10.3174/ajnr.A5889.

48. Nagel S, Sinha D, Day D, Reith W, Chapot R, Papanagiotou P, 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. doi: 10.1177/1747493016681020.

49. Guberina N, Dietrich U, Radbruch A, Goebel J, Deuschl C, Ringelstein A, et al. Detection of early infarction signs with machine learning-based diagnosis by means of the Alberta Stroke Program Early CT score (ASPECTS) in the clinical routine. Neuroradiology 2018;60:889-901. doi: 10.1007/s00234-018-2066-5.

50. Goebel J, Stenzel E, Guberina N, Wanke I, Koehrmann M, Kleinschnitz C, et al. Automated ASPECT rating: comparison between the Frontier ASPECT Score software and the Brainomix software. Neuroradiology 2018;60:1267-72. doi: 10.1007/s00234-018-2098-x.

51. Hoelter P, Muehlen I, Goelitz P, Beuscher V, Schwab S, Doerfler A. Automated ASPECT scoring in acute ischemic stroke: comparison of three software tools. Neuroradiology 2020;62:1231-8. doi: 10.1007/s00234-020-02439-3.