

ORIGINAL ARTICLE

Open Access



Improving the diagnosis of acute ischemic stroke on non-contrast CT using deep learning: a multicenter study

Weidao Chen^{1,3†}, Jiangfen Wu^{2,3†}, Ren Wei³, Shuang Wu³, Chen Xia³, Dawei Wang³, Daliang Liu⁴, Longmei Zheng⁵, Tianyu Zou⁶, Ruijiang Li⁷, Xianrong Qi^{8,9*} and Xiaotong Zhang^{1,10,11*} 

Abstract

Objective: This study aimed to develop a deep learning (DL) model to improve the diagnostic performance of EIC and ASPECTS in acute ischemic stroke (AIS).

Methods: Acute ischemic stroke patients were retrospectively enrolled from 5 hospitals. We proposed a deep learning model to simultaneously segment the infarct and estimate ASPECTS automatically using baseline CT. The model performance of segmentation and ASPECTS scoring was evaluated using dice similarity coefficient (DSC) and ROC, respectively. Four raters participated in the multi-reader and multicenter (MRMC) experiment to fulfill the region-based ASPECTS reading under the assistance of the model or not. At last, sensitivity, specificity, interpretation time and interrater agreement were used to evaluate the raters' reading performance.

Results: In total, 1391 patients were enrolled for model development and 85 patients for external validation with onset to CT scanning time of 176.4 ± 93.6 min and NIHSS of 5 (IQR 2–10). The model achieved a DSC of 0.600 and 0.762 and an AUC of 0.876 (CI 0.846–0.907) and 0.729 (CI 0.679–0.779), in the internal and external validation set, respectively. The assistance of the DL model improved the raters' average sensitivities and specificities from 0.254 (CI 0.22–0.26) and 0.896 (CI 0.884–0.907), to 0.333 (CI 0.301–0.345) and 0.915 (CI 0.904–0.926), respectively. The average interpretation time of the raters was reduced from 219.0 to 175.7 s ($p = 0.035$). Meanwhile, the interrater agreement increased from 0.741 to 0.980.

Conclusions: With the assistance of our proposed DL model, radiologists got better performance in the detection of AIS lesions on NCCT.

Key points

- The model simultaneously segments infarcts and estimates ASPECTS by using baseline CT.
- A mirror-assembly module plus dual-path DCNN model improved the segmentation efficiency.

[†]Weidao Chen and Jiangfen Wu contributed equally to this work

*Correspondence: qixr@bjmu.edu.cn; zhangxiaotong@zju.edu.cn

¹ Interdisciplinary Institute of Neuroscience and Technology, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou 310027, Zhejiang, China

⁸ School of Pharmaceutical Sciences, Peking University, Beijing 100191, China

Full list of author information is available at the end of the article

- We evaluated the model in a multi-reader and multicenter (MRMC) setting.

Keywords: Deep learning, Non-contrast computed tomography, Early ischemic changes, Alberta Stroke Program Early Computed Tomography Score, Multi-reader and multi-center study, Acute ischemic stroke

Introduction

Stroke is one of the major threats to human health, and it is the third leading cause of death in the world with high mortality and disability rate [1–3]. Non-contrast computed tomography (NCCT) is deemed as the first choice for all stroke diagnosis due to its relative high speed, broad accessibility and cost-effectiveness compared with magnetic resonance imaging (MRI) and CT perfusion (CTP) [3, 4], especially in the setting of emergency department [5]. NCCT not only has high sensitivity to the detection of intracranial hemorrhage, but also is a widely used tool to select patients for endovascular therapy [6–8]. The Alberta Stroke Program Early Computed Tomography Score (ASPECTS) is a quantitative score method for early ischemic changes (EIC) evaluation based on NCCT [9] and has been used in several randomized controlled trials (RCTs) for patient selection and exclusion [10–12]. Until now, NCCT-ASPECTS is still the most widely used modality for early ischemic triage and thrombolytic outcome prediction in emergency department [13–19].

However, EIC detection using NCCT is yet challenging and the ASPECTS can only be roughly determined in practice. The early signs of ischemia and their translation into ASPECTS suffer from considerable missed diagnosis and interrater variability due to the rater's experience difference [20–22], since the mild infarction on NCCT is difficult to be recognized by naked eyes, and there is no obvious boundary among brain regions involved in ASPECTS scoring. As reported in the previous studies, only 10% acute ischemic stroke (AIS) and 7% hyperacute ischemic stroke patients could be detected by using NCCT only [5]. Moreover, the EIC detection is also experience-dependent and suffers from the limited inter-observer consistency [23, 24]. More endeavor needs to be devoted to improve the EIC detection sensitivity and ASPECTS assessment consistency.

Several software applications using artificial intelligence, including classical machine learning and deep learning, have been designed for automated EIC detection and ASPECTS scoring [25–29]. The classical machine learning approach usually uses the image gray-scale-based segmentation and pre-defined features to define the lesion, e.g., e-ASPECTS [29] (Brainomix, UK), RAPID-ASPECTS [24] (iSchemaView, USA) and Frontier ASPECTS [30] (Siemens Healthcare, Germany). These

methods are limited by hand-crafted texture patterns or geometric shapes that rely on data scientists' expertise, and thus the ASPECTS results vary among studies [30, 31]. Deep learning (DL) has emerged to be a powerful technique in medical imaging diagnosis, which can discover abstract task-specific features and further uses these features to produce accurate clinical interpretations in an end-to-end manner [32–35]. The application of DL methods in the EIC detection and ASPECTS interpretation has just emerged [27, 28] and showed good performance. However, there are few well-established DL approaches for NCCT-ASPECTS scoring and most of the existing studies are single-centered; in addition, they only focus on the evaluation of the model efficiency rather than the model performance in clinical emergency scenarios. Therefore, the DL model performance in clinical scenarios needs further investigations.

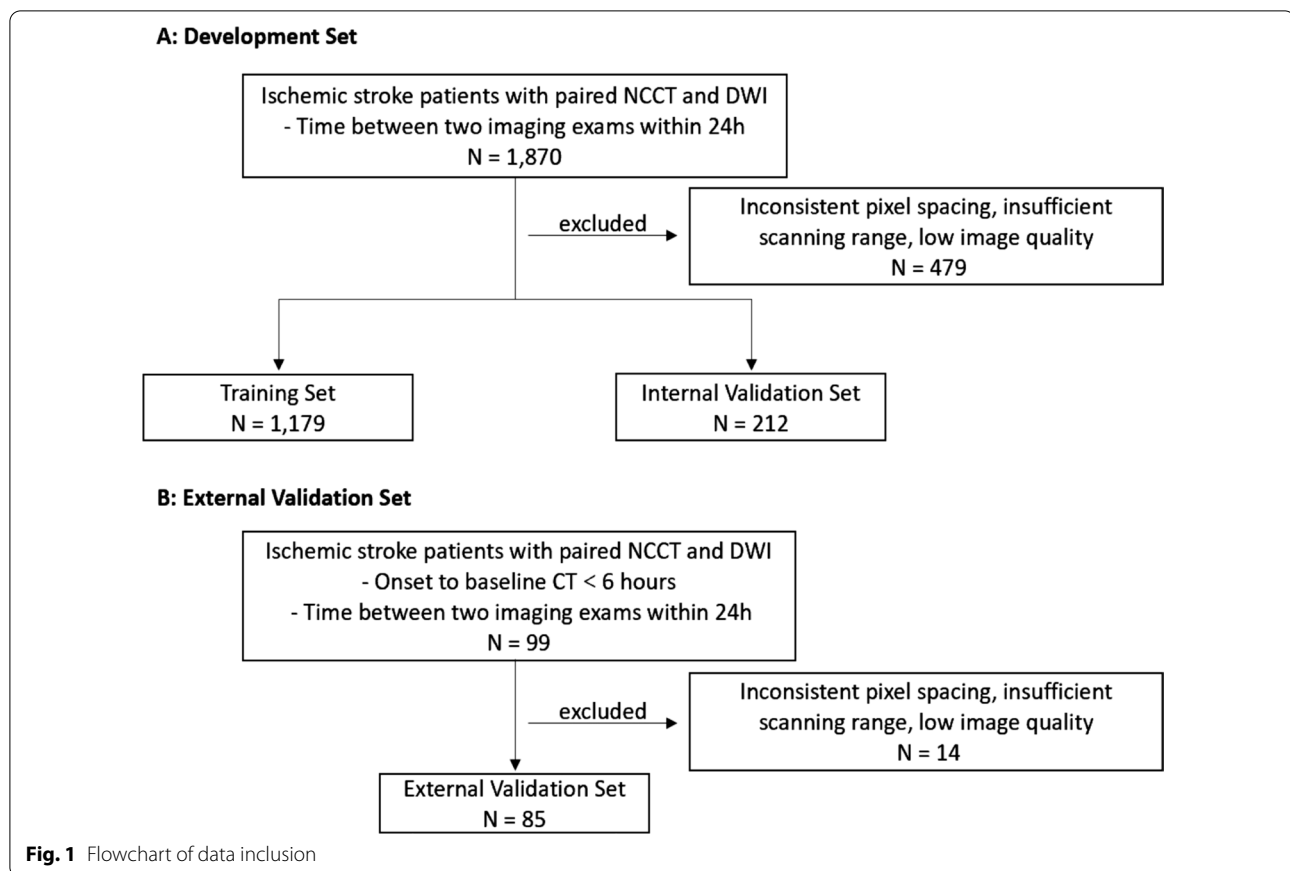
To address the above issues, we propose a novel DL model for the automatic AIS lesions detection and ASPECTS scoring and further demonstrate the value of the model in clinical assistance. Firstly, the model is developed based upon a reasonably large NCCT dataset across 5 stroke centers. Secondly, the model uses a mirror assembly module and a dual-path DCNN model to enhance the lesion detection ability. Besides, to validate the model clinical performance, we utilize a multi-reader and multicenter (MRMC) experiment to evaluate the radiologists' diagnosis under the aid of the proposed model.

Methods and materials

This retrospective study was approved by the Ethics Review Board of all participating hospitals. Patients' private information in the Digital Imaging and Communications in Medicine (DICOM) header file was desensitized. Owing to the retrospective nature, the requirement for informed consent was waived in this study.

Participants

The Development Set was acquired by searching for the keyword of "acute ischemic stroke" in the radiology information system (RIS) between 2013 and 2017 from two participating hospitals, and AIS patients with NCCT-to-DWI time < 24 h were included in this dataset (1870 patients); then, images with inconsistent pixel



spacing, insufficient scanning range or low image quality were excluded. The development set was divided into a training set and an internal validation set at a ratio of 5:1, as illustrated in Fig. 1.

The external validation set was curated from the other three hospitals. AIS patients with onset-to-NCCT time < 6 h and NCCT-to-DWI time < 24 h were included (99 patients). The unqualified images were excluded, and the clinical characteristics of the enrolled patients were also collected.

Specifically, only MCA territory infarcts were included to be analyzed in both the development set and the external validation set. The image acquisition parameters are provided in Additional file 1.

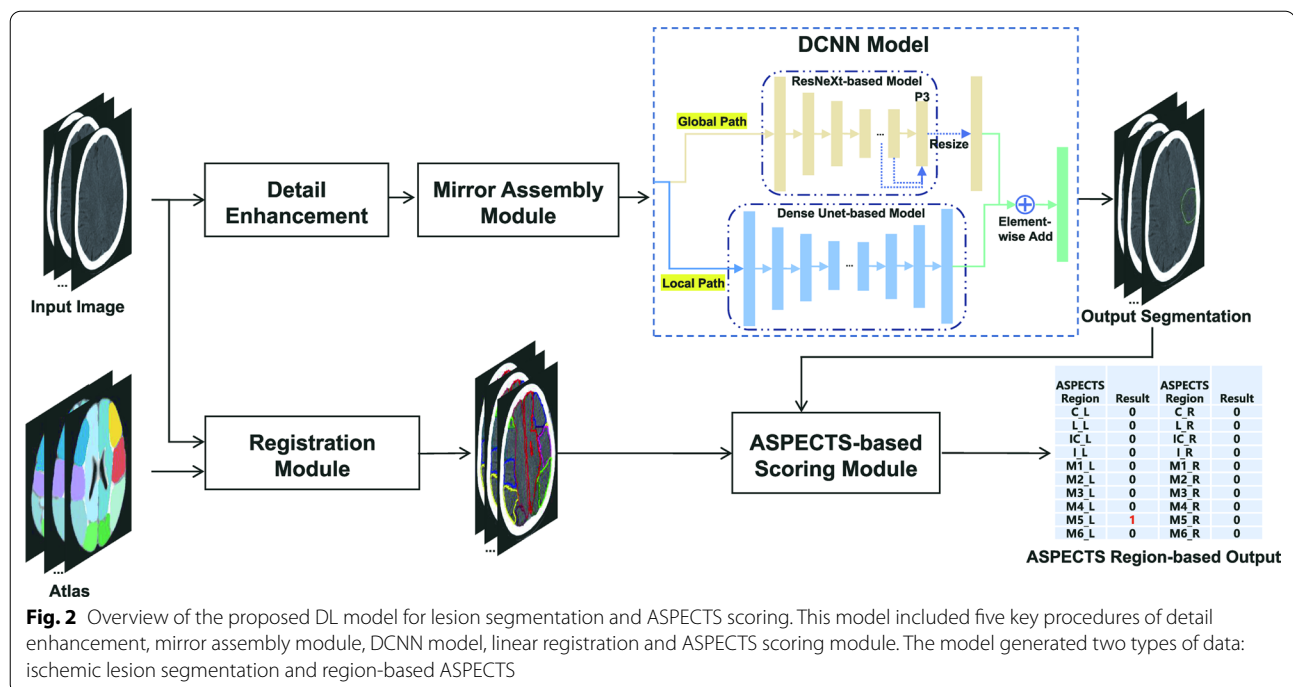
Ground truth determination

Ground truth of ischemic region and ASPECTS for both the development set and the external validation set were established for each NCCT scan via the consensus of three board-certified radiologists (Z.T.Z., L.C.F., P.F.), and all the radiologists had at least 10-year experience in emergency neuroradiology (these experts received guiding instructions on www.aspectsinstroke.com prior to labeling). The ischemic regions were manually drawn

and scored by one radiologist and validated by the other two radiologists independently. All the NCCT ischemic regions were delineated using 3D Slicer (version 4.8.1, www.slicer.org) with reference to the paired DWI images and radiology reports. ASPECTS scores for each region were also rated according to ASPECTS guidelines.

DL model development

The proposed DL model is shown in Fig. 2, consisting of 5 key procedures to score the ASPECTS automatically. To reduce the influence of noise, we firstly pre-processed the NCCT images with median filter and the dark image enhancement algorithm [36] (“Detail Enhancement”). The manifestations of EIC on NCCT of AIS patients are extremely insignificant, and experts can only make decisions based on the slight gray-level difference between the left and right brains in NCCT images. For this reason, we designed a “Mirror Assembly Module,” which obtained the mirror image by flipping the original NCCT image around the midline of the brain, and used the original NCCT image to subtract the mirrored one to obtain the difference image. Then, we spliced the original images, the mirror images and the difference images to obtain the input samples of our DL model.



Specifically, the boundaries of acute ischemic lesions were not obvious in the NCCT images and hard to be delineated accurately, thus they were easy to be misdiagnosed. Therefore, if such noisy annotations were used for model training, it would lead to model learning bias and damage model efficiency, especially when the traditional panoptic segmentation methods, like UNet [37] and panoptic FPN [38], were directly applied for model development. In order to tackle this problem and being inspired by the idea of ensemble learning, we proposed a novel deep convolutional neural network (DCNN) model with two pathways—"Global Path" and "Local Path." "Global Path" was built based on ResNeXt-model [39, 40] to achieve good global localization and coarse segmentation of ischemic lesions and reduce the lesion-level missed diagnosis rate. However, "Global Path" could only output the coarse segmentation images from the P3 layer of ResNeXt-50 and ignored the boundary details of ischemic lesions, thus we developed a "Local Path" as a supplement to "Global Path" to perform careful pixel-level segmentation under full-resolution images based on the Dense-UNet model [41]. Then, we resized the output segmentations of "Global path" and "Local Path" to the size of origin NCCT images with bilinear interpolation and subsequently fused the resized two-pathway output segmentation by element-wise addition to obtain the final output segmentation. Details of the "Global Path" and "Local Path" are provided in Additional file 1.

In order to score the ASPECTS, we generated blood supply maps based on the MNI brain template [42, 43], as shown in Additional file 1: Fig. S3. Then, by linearly registering the blood supply region maps to the NCCT images ("Linearly Registration"), the blood supply regions of the corresponding individuals were obtained. At last, according to the ASPECTS scoring criteria [9] of calculating the overlap proportion of the lesion segmentation and the blood supply regions, a score for each ASPECTS region was obtained ("ASPECTS Scoring Module"). The proposed model was implemented by using Python 2.7 and the MxNet framework (version 1.4.0, mxnet.apache.org) with two Nvidia GTX 1080Ti GPUs for computation accelerations. The model was trained with the stochastic gradient descent optimizer and a stepped power-decaying learning rate scheduler starting from 0.002. The deep learning model was trained with an input image size of 512×512 and 100 epochs. When the model loss of the internal validation set began to increase, the training process was stopped. In order to make the model focus on the details as well as the overall segmentation contour of the lesion, we constructed a loss function integrating the cross-entropy and DICE loss functions. The loss function is introduced in Additional file 1.

Multi-reader and multicenter (MRMC) experiment

We conducted a MRMC experiment on the external validation set to evaluate the clinical efficacy of the proposed DL model in the help of four radiologist raters from

different tertiary teaching hospitals; each rater had over 10 years of clinical experience in emergency radiology. Before the experiment, all raters were asked to participate in an educational session on how to use ASPECTS according to the instructions on www.aspectsinstroke.com, and then each of them was provided with apparatus like the ones equipped at the emergency department. They were also instructed to interpret each scan in line with the emergency clinical practice. The raters were blind with the patients' medical histories and onset symptoms. Two sessions were operated in this MRMC experiment: in the first session, only original scans were available for rater review. And the second session was performed after a washout period of 2 weeks and the model segmentation and region-based ASPECTS were provided to the raters. All raters were asked to score the region-based ASPECTS on a spreadsheet, and their reading time was also recorded in the two sessions. Interrater agreement among the four raters was also calculated and compared between the two sessions.

Statistical analysis

Baseline demographic characteristics are demonstrated in Table 1. DL models were evaluated in a region-based manner according to the ASPECTS criteria. Twenty binomial outcomes (20 ASPECTS regions) were generated for each case. The ROC curve was plotted in the following manner: we enumerated the probability thresholds for the DCNN model between 0 and 1, and for each threshold, the corresponding TPR and FPR were calculated based on the model outputs for each ASPECTS region; these TPR-FPR pairs were plotted and formed into ROC curves. AUC (area under curve) was calculated based on the ROC curve, and its confidence interval (CI) was calculated referenced to Dai Feng's method [44]. Dice similarity coefficient (DSC), precision and recall were used for segmentation performance evaluation.

In MRMC experiment, we used the region-based sensitivity and specificity for clinical efficacy evaluation. The total time was recorded for each session, and the averaged case-reading time was calculated through averaging the total duration by the total number of cases. The averaged time was compared with paired Student's *t* test. The sensitivity and specificity were compared with the McNemar test. The interrater agreement was measured using ICC (intraclass correlation coefficient) of two-way mixed single absolute agreement [45]. All statistical analysis was performed by R programming language (version 3.6.2), and a two-sided $\alpha < 0.05$ was considered statistically significant.

Results

Patient data

At last, a total of 1,391 patients (1,179 patients in the training set and 212 patients in the internal validation set) were enrolled for model development and 85 patients were eligible for model external validation (see Fig. 1). All eligible patients' baseline characteristics are summarized in Table 1. The age difference among these three sets was statistically significant ($p < 0.001$), and there was no gender difference ($p = 0.120$); furthermore, post hoc analysis of age showed that there was no difference between the training set and the internal validation set, and a statistical age difference was found between the two groups and the external validation set. It was observed that CT scanners made by GE Healthcare (Boston, MA, USA) and Siemens Healthineers (Erlangen, Germany) were in a majority for development set (Training Set: GE 57.0%, SIEMENS 41.7%; Internal Validation Set: GE 61.8%, SIEMENS: 37.3%), while PHILIPS Healthcare (Best, The Netherlands) was the major vendor in the external validation set (58.8%), and there was a significant manufacturer difference among the three groups ($p < 0.001$).

Detailed clinical information was only recorded for the external validation set. The mean time from onset to baseline CT was 176.4 ± 93.6 min, and the median NIH Stroke Scale (NIHSS) was 5 (IQR 2–10). The DWI-ASPECTS distribution and the ASPECTS region of the external validation set are illustrated in Fig. 3 with a bar graph (median DWI-ASPECTS = 9, IQR 8–10). In total, 1700 (85×20) ASPECTS regions in the external validation set were scored.

The DL model efficiency

ROC curves of the internal and external validation sets are shown in Fig. 4a, b). For the internal validation set, the model achieved an AUC of 0.876 (95% CI 0.846–0.907), while on the external validation set, the model achieved an AUC of 0.729 (95% CI 0.679–0.779). The external validation showed that the trained model provided higher sensitivity and specificity in identifying individual ischemia ASPECTS regions than all expert raters. The DSC, precision and recall was 0.600, 0.528 and 0.694 for the internal validation set and 0.762, 1.000 and 0.616 for the external validation set, respectively.

To visualize the model performance, NCCT segmentation result of two patients is presented in Fig. 5. Obviously, it is confounded to define the lesion boundary with naked eyes, whereas the model could almost achieve this goal.

Table 1 Baseline demographic and clinical characteristics

	Training set (n = 1179)	Internal validation set (n = 212)	External validation set (n = 85)	p value
<i>Patient characteristics</i>				
Age, mean \pm SD, years	61.3 \pm 12.5	62.0 \pm 11.1	66.6 \pm 12.4	< .001
Male	70.80% (835/1179)	72.20% (153/212)	64.70% (55/85)	0.12
<i>Manufacturer</i>				
GE	57.00% (672/1179)	61.80% (131/212)	28.20% (24/85)	< .001
PHILIPS	1.10% (13/1179)	0.90% (2/212)	58.80% (50/85)	
SIEMENS	41.70% (492/1179)	37.3% (79/212)	12.90% (11/85)	
Other	0.20% (2/1179)	0.00% (0/212)	0.00% (0/85)	
Time from onset to CT (min)	–	–	176.4 \pm 93.6	–
Admission NIHSS score	–	–	5(2–10) (n = 76/85)	–
Admission systolic BP (mmHg)	–	–	153 \pm 22 (n = 84/85)	–
<i>Risk factors</i>				
Previous stroke	–	–	14.10% (12/85)	–
Hypertension	–	–	54.10% (46/85)	–
Diabetes	–	–	16.50% (14/85)	–
Smoking	–	–	24.70% (21/85)	–
Atrial fibrillation	–	–	15.30% (13/85)	–
CVD	–	–	17.60% (15/85)	–

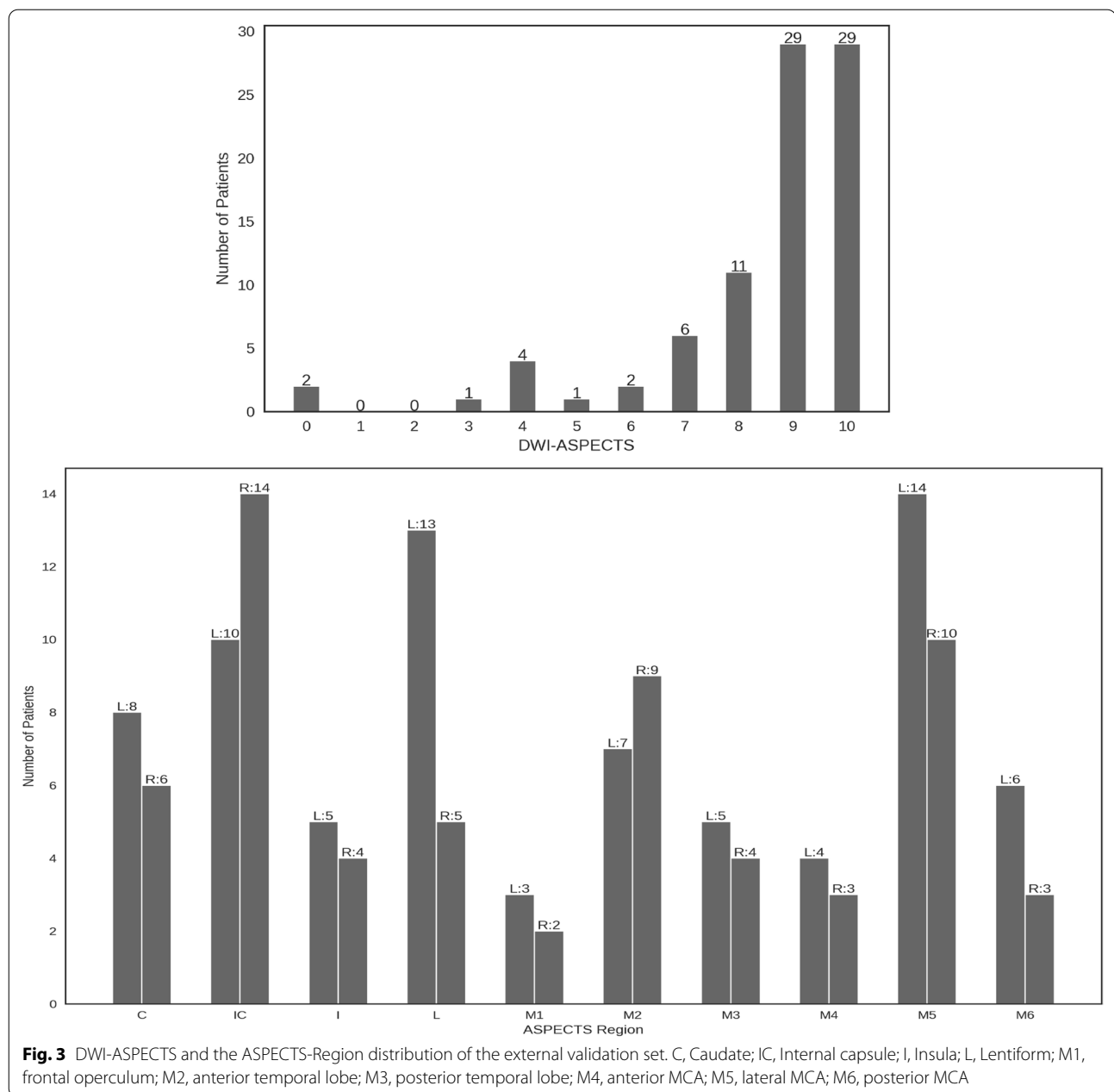
Data are percentages with numbers in parentheses for categorical variables. Continuous variables are reported as mean \pm standard deviation with number of entries in parentheses. Ordinal variables (NIHSS score) are reported as median with interquartile range (IQR) and number of entries in parentheses. ANOVA was used as the test for the difference of mean for continuous variables. Fisher's exact test was used to test for the difference of proportion for categorical variables. The post hoc analysis was performed by Tukey's HSD method. Other manufacturers were omitted from hypothetical tests due to the absence of data

NIHSS, National Institutes of Health Stroke Scale; BP, Blood pressure; CVD, cardiovascular disease

Performance comparisons of the MRMC study

The performance of the expert raters on the external validation set is summarized in Table 2. The average sensitivity was 0.254 (95% CI 0.22–0.26) and 0.333 (95% CI 0.301–0.345), and the specificity was 0.896 (95% CI 0.884–0.907) and 0.915 (95% CI 0.904–0.926) without vs. with AI, respectively—there was a statistical difference for average specificity ($p=0.014$) but not for the average sensitivity ($p=0.196$). As shown in Fig. 4b, c, with the assistance of the DL model, the sensitivity and specificity of Expert 1 and 2 were significantly

improved (Expert 1: sensitivity $p<0.001$, specificity $p=0.002$; Expert 2: sensitivity $p=0.006$, specificity $p=0.022$); for Expert 3, the specificity was significant improved ($p=0.019$) and the sensitivity behaved no significant difference ($p=1$); no significant differences were observed in the sensitivity ($p=0.281$) and specificity ($p=0.630$) of Expert 4. In addition, the region-based ICC with model aid among the four raters comes up to 0.980 compared to 0.741 without model aid, and the score-based ASCETS ICC improved from 0.614 to 0.809.



Regarding the reading time, the average reading time of the external validation set is summarized in Table 3. Obviously, with the aid of the DL model, the interpretation time of all expert raters was significantly reduced ($p=0.035$).

Discussion

Diagnosis of EIC and ASPECTS using NCCT is a persistent challenge in emergency department with urgent clinical needs. This study attempts to utilize the DL

technique to detect infarction area and evaluate the DL performance in emergency clinical scenario. Specifically, our proposed DL model demonstrates good efficacy in ASPECTS scoring and achieves an AUC of 0.729 in the external validation set. The DL model using mirror assembly module and dual-path DCNN model obtains a high DSC of 0.762 in occult lesion segmentation. With the model assistance, raters show improved performance in ASPECTS scoring with higher sensitivity and specificity, shorter operating time and good interrater agreement. To the best of our knowledge, it is

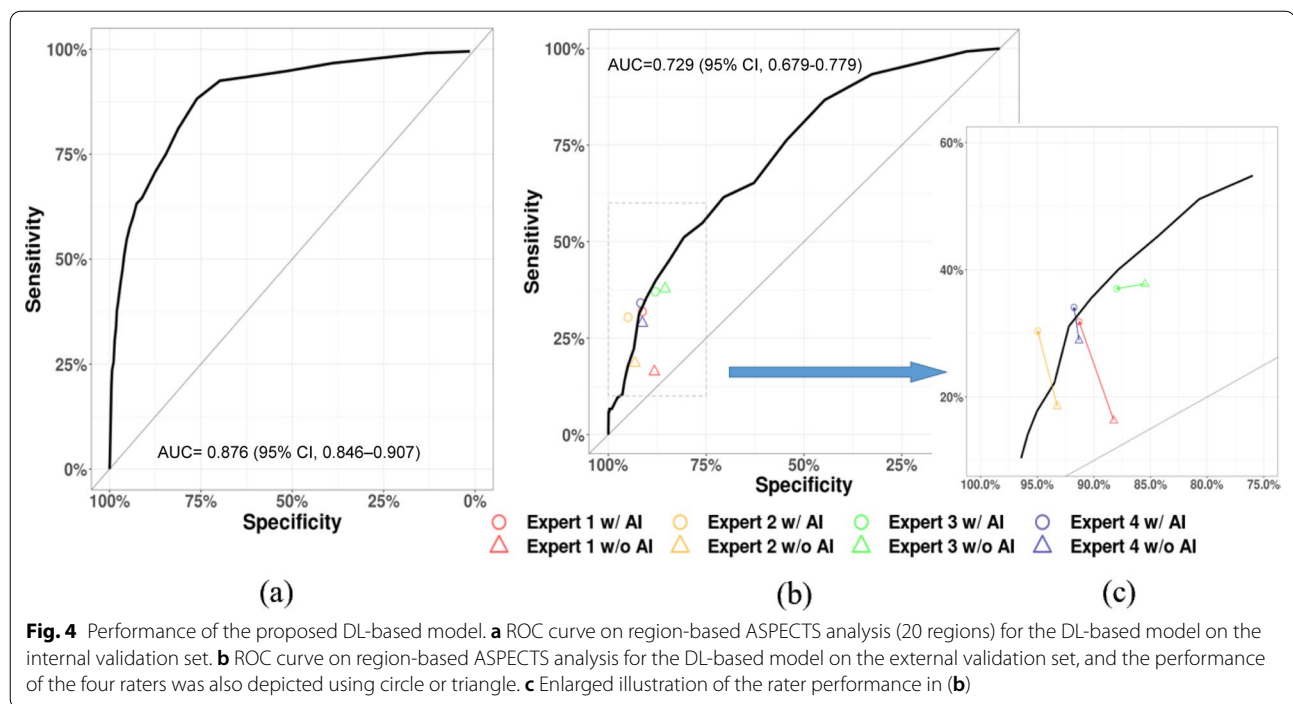


Fig. 4 Performance of the proposed DL-based model. **a** ROC curve on region-based ASPECTS analysis (20 regions) for the DL-based model on the internal validation set. **b** ROC curve on region-based ASPECTS analysis for the DL-based model on the external validation set, and the performance of the four raters was also depicted using circle or triangle. **c** Enlarged illustration of the rater performance in (b)

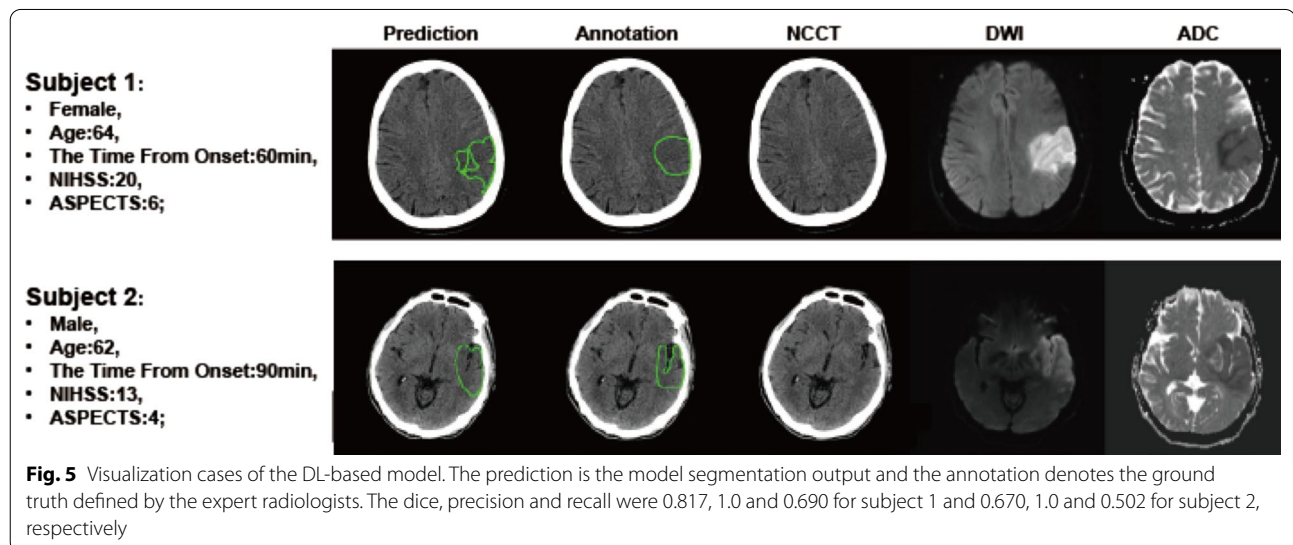


Fig. 5 Visualization cases of the DL-based model. The prediction is the model segmentation output and the annotation denotes the ground truth defined by the expert radiologists. The dice, precision and recall were 0.817, 1.0 and 0.690 for subject 1 and 0.670, 1.0 and 0.502 for subject 2, respectively

the first time to investigate the emergency efficiency of the DL model in a MRMC manner.

Our proposed DL model outperforms the previously reported studies [27, 46] in lesion segmentation, and several factors lay the foundation of the good performance: firstly, the high-quality ground truth provided by stroke experts and follow-up DWI, which assures the trained DL model to learn the task-specific feature effectively; secondly, the DL method is good at simulating and integrating the experience of stroke

experts in detecting AIS lesions; lastly but most importantly, the innovative model pipeline that comprises a mirror assembly module to capture the image difference between left and right brains and a dual-path DCNN model to tackle the problems of indiscernible lesion detection and segmentation. This pipeline could enhance the feature characteristics associated with image segmentation tasks while suppressing redundant features. In addition, to score the ASPECTS accurately, we also develop an ASPECTS atlas and register it to the

Table 2 Performance of the raters on external validation set

	Baseline	Aided	<i>p</i> value
<i>Expert 1</i>			
Sensitivity	0.163 (0.105, 0.236)	0.319 (0.241, 0.404)	< .001
Specificity	0.882 (0.865, 0.898)	0.913 (0.898, 0.927)	0.002
<i>Expert 2</i>			
Sensitivity	0.185 (0.124, 0.261)	0.304 (0.228, 0.389)	0.006
Specificity	0.932 (0.919, 0.944)	0.950 (0.937, 0.960)	0.022
<i>Expert 3</i>			
Sensitivity	0.378 (0.296, 0.465)	0.370 (0.289, 0.458)	1
Specificity	0.855 (0.837, 0.872)	0.880 (0.863, 0.896)	0.019
<i>Expert 4</i>			
Sensitivity	0.289 (0.214, 0.373)	0.341 (0.261, 0.427)	0.281
Specificity	0.913 (0.898, 0.927)	0.918 (0.903, 0.931)	0.63
<i>Average</i>			
Sensitivity	0.254 (0.167, 0.341)	0.333 (0.248, 0.419)	0.196
Specificity	0.896 (0.884, 0.907)	0.915 (0.904, 0.926)	0.014
Region-based ICC	0.741	0.98	–
Score-based ICC	0.614	0.809	–

Sensitivity and specificity were expressed as point estimate with lower and upper bound of 95% confidence interval (CI) in parentheses. CI for average sensitivity and specificity was calculated by the Jackknife method

Table 3 Average time per case in reading experiment

	Average time per case (<i>n</i> = 85)		<i>p</i> value
	Baseline (s)	Aided (s)	
Expert 1	202.1	176.2	–
Expert 2	250.1	176.8	–
Expert 3	197.7	176.2	–
Expert 4	226.1	173.7	–
Average	219.0 ± 24.2	175.7 ± 1.4	0.035

Total time for each reader in each session was recorded and the averaged case-reading time was calculated. Average time for each session was expressed in mean ± standard deviation and was compared with paired Student's *t* test

original NCCT images reversely to reduce image deformation errors; furthermore, in order to suppress the segmentation errors, we use the region-level ASPECTS to determine the segmentation threshold rather than DSC. EIC detection and ASPECTS scoring on NCCT is clinically desired, but the subtle signs of EIC cannot always be captured visually, as shown in Fig. 5. Therefore, the efficacy of deep learning in indiscernible features detection may be an accelerator to the NCCT clinical application in AIS diagnosis.

Performance comparisons between radiologist and automatic software in ASPECTS scoring have been reported previously [25, 26, 29], whereas our study differed from previous ones in using the AI model as first-reader. AI as first-reader has been widely accepted

in pulmonary nodules detection [47], but fewer have been reported in NCCT-ASPECTS interpretation. As shown in Fig. 4b, c and Table 2, all the stroke experts aided with the DL model reached a relatively high sensitivity level (sensitivity ≥ 0.3) with improved specificities ($p = 0.014$), along with the reduced reading time ($p = 0.035$). However, we also note that not all raters' sensitivities or specificities are statistically significant, which may attribute to the raters' acceptance of the model result. Since the signs are too faint to be observed, the confidence for lesion determination may vary with the readers' experience. Various studies have shown that only modest to moderate interrater agreement was achieved for determining NCCT-ASPECTS, and the ICC of radiologists ranged from 0.579–0.936 [21, 23, 24, 28, 48, 49]. In contrast, in our study, the ASPECTS reliability (ICC 0.980) of the radiologists with AI assistance is significantly increased and help to improve the diagnostic confidence and medical quality consistency. In a word, the AI model can be a valuable supplement and/or confirmation to the expert interpretation in ASPECTS scoring with improved performance and reliability.

The greatly varied expert diagnosis and software performance in NCCT-ASPECTS evaluation remained to be a concern, as reported sensitivity and specificity ranged from 0.26 to 0.8 and 0.87 to 0.97, respectively [5, 25, 26, 28, 29]. In our study, the resulting sensitivity and specificity are not superior to previous reported methods, which can be attributed to the experiment setting of the external validation set, including the severity of the infarction, the distribution of the ASPECTS and the invaded-region, the radiologists' experience and the varied scan vendors. Firstly, compared to previous studies, enrolled patients in our external validation set have lower NIHSS (5, 2–10) and age (66.6 ± 12.3), making the lesion more blurred to be detected compared to that of Masaki Naganuma' study [28]. Secondly, the lesion detection sensitivity is greatly affected by the ASPECTS distribution, especially for the super acute stroke with ASPECTS > 7 , while the median ASPECTS of this study is only 9 (IQR 8–10) with 29 patients of ASPECTS 9 and 29 patients of ASPECTS 10, respectively, raising the difficulty of lesion detection compared to that of Hulin Kuang' study [50]. In addition, the ASPECTS accuracy could also be affected by brain region distribution [51]. As shown in Fig. 3B, the lesion regions in our study mainly scatter in M5 ($n = 24$), IC ($n = 24$), L ($n = 18$) and M2 ($n = 16$), and it has been demonstrated that IC and M5 behave lower agreement to the ground truth and higher rate of missed diagnosis [52]. Thirdly, to align with the emergency setting, the participant raters in this study are general radiologists

rather than neuroradiologists. Compared to the neuroradiologist, the sensitivity and specificity of general radiologist were lower in ASPECTS scoring. This is in line with our findings that the general radiologists performed low sensitivity. At last, this is a multicenter study that have more complicated image quality than a single-center study, which may also weaken the clinical performance of the raters in NCCT-ASPECTS scoring. In a word, compared to previous studies, the external validation data of our study are collected from scanners of multiple vendors, having lower NIHSS and higher ASPECTS. Thus, the diagnostic performance in the external validation set is reasonably acceptable and consistent with reported ones.

Several limitations merit discussion. Firstly, this study mainly focuses on the stroke lesion, neglecting the presence of other neuroimaging signs. Particularly, the existence of leukoencephalopathy and old infarctions may disturb the calculation of ASPECTS. Secondly, the ground truth is MRI infarct images obtained within 24 h after patients receiving complete reperfusion, causing a time delay between the NCCT infarct and the follow-up acquisitions. The time delay may introduce DWI infarct bias as infarct grows over time. Thirdly, patients' outcome is not taken into consideration yet, and a perspective study ought to be conducted to evaluate the patient's outcome improvement aroused by the proposed model.

In conclusion, the proposed deep learning model can automatically detect EIC and interpret the ASPECTS and demonstrate improved and reliable performance in the clinical scenario. DL ASPECTS model could be a good assistant to the general radiologist, especially in the hospitals with limited expertise and resource, and further guide the AIS therapeutic decision-making.

Abbreviations

AIS: Acute ischemic stroke; ASPECTS: Alberta Stroke Program Early Computed Tomography Score; AUC: Area under curve; CI: Confidence interval; DCNN: Deep convolutional neural network; DICOM: Digital imaging and communications in medicine; DL: Deep learning; DSC: Dice similarity coefficient; EIC: Early ischemic changes; ICC: Intraclass correlation coefficient; MRI: Magnetic resonance imaging; MRMC: Multi-reader and multicenter; NCCT: Non-contrast computed tomography; RCTs: Randomized controlled trials; RIS: Radiology information system.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13244-022-01331-3>.

Additional file 1. This material supplements the details of the image acquisition parameters, model structure, loss function, and self-developed ASPECTS atlas.

Acknowledgements

We thank Drs. Kang Wang and Ziqi Xu for helpful discussions.

Author contributions

WC and JW were responsible for the study design, literature search, data analysis and manuscript drafting. DL, LZ and TZ were responsible for data collection. RW, SW, CX and DW were mainly responsible for administrative, technical or material support. XZ, XQ and RL were responsible for the study concept and critical revision. All authors read and approved the final manuscript.

Funding

This work was supported in part by China Brain Project (2021ZD0200401), National Key Research and Development Program of China (2018YFA0701400), National Natural Science Foundation of China (52277232, 81701774, 61771423), the Fundamental Research Funds for the Central Universities (226-2022-00136), Key R&D Program of Zhejiang Province (2021C03001), Key R&D Program of Jiangsu Province (BE2022049-4), and Key-Area R&D Program of Guangdong Province (2018B030333001).

Availability of data and materials

The data that support the findings of this study are available upon request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

Declarations

Ethics approval and consent to participate

The project was approved by the Institutional review boards.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing financial or non-financial interests.

Author details

¹Interdisciplinary Institute of Neuroscience and Technology, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou 310027, Zhejiang, China. ²Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ³Infervision Institute of Research, Beijing 100025, China. ⁴Liaocheng People's Hospital, Liaocheng 252000, Shandong, China. ⁵Medical Imaging Center, Ankang Central Hospital, Ankang 725000, Shanxi, China. ⁶Weihai Municipal Hospital, Weihai 264200, Shandong, China. ⁷Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA 94304, USA. ⁸School of Pharmaceutical Sciences, Peking University, Beijing 100191, China. ⁹Beijing Key Laboratory of Molecular Pharmaceutics and New Drug Delivery System, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China. ¹⁰College of Electrical Engineering, Zhejiang University, Hangzhou 310000, Zhejiang, China. ¹¹MOE Frontier Science Center for Brain Science & Brain-machine Integration, Zhejiang University, Hangzhou 310000, Zhejiang, China.

Received: 11 September 2022 Accepted: 11 November 2022

Published online: 06 December 2022

References

1. Collaborators GBD (2017) Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 390(10100):1151–1210
2. Latchaw RE, Alberts MJ, Lev MH et al (2009) Recommendations for imaging of acute ischemic stroke: a scientific statement from the American Heart Association. *Stroke* 40(11):3646–3678
3. Powers WJ, Rabinstein AA, Ackerson T et al (2018) 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* 49(3):e46–e110
4. Wardlaw JM, Mielke O (2005) Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment—systematic review. *Radiology* 235(2):444–453

5. Chalela JA, Kidwell CS, Nentwich LM et al (2007) Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet* 369(9558):293–298
6. Hirano T, Sasaki M, Tomura N et al (2012) Low Alberta stroke program early computed tomography score within 3 hours of onset predicts subsequent symptomatic intracranial hemorrhage in patients treated with 06 mg/kg Alteplase. *J Stroke Cerebrovasc Dis* 21(8):898–902
7. Von Kummer R, Allen KL, Holle R et al (1997) Acute stroke: Usefulness of early CT findings before thrombolytic therapy. *Radiology* 205(2):327–333
8. Wardlaw JM, Sandercock P, Cohen G et al (2015) Association between brain imaging signs, early and late outcomes, and response to intravenous alteplase after acute ischaemic stroke in the third International Stroke Trial (IST-3): secondary analysis of a randomised controlled trial. *Lancet Neurol* 14(5):485–496
9. Barber PA, Demchuk AM, Zhang JJ et al (2000) Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *Lancet* 355(9216):1670–1674
10. Goyal M, Demchuk AM, Menon BK et al (2015) Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med* 372(11):1019–1030
11. Jovin TG, Chamorro A, Cobo E et al (2015) Thrombectomy within 8 hours after symptom onset in ischemic Stroke. *N Engl J Med* 372(24):2296–2306
12. Saver JL, Goyal M, Bonafe A et al (2015) Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in Stroke. *N Engl J Med* 372(24):2285–2295
13. Demchuk AM, Hill MD, Barber PA et al (2005) Importance of early ischemic computed tomography changes using ASPECTS in NINDS rtPA Stroke study. *Stroke* 36(10):2110–2115
14. Dzialowski I, Hill MD, Coutts SB et al (2006) Extent of early ischemic changes on computed tomography (CT) before thrombolysis prognostic value of the Alberta Stroke program early CT score in ECASS II. *Stroke* 37(4):973–978
15. Gupta A, Schaefer P, Chaudhry ZA et al (2012) Interobserver reliability of baseline noncontrast CT Alberta Stroke program early CT score for intra-arterial stroke treatment selection. *AJNR Am J Neuroradiol* 33(6):1046–1049
16. Hill M, Demchuk A, Tomsick T et al (2006) Using the baseline CT scan to select acute stroke patients for IV-IA therapy. *AJNR Am J Neuroradiol* 27(8):1612–1616
17. Liebeskind DS, Jahan R, Nogueira RG et al (2014) Serial Alberta Stroke program early CT score from baseline to 24 hours in solitaire flow restoration with the intention for thrombectomy study: a novel surrogate end point for revascularization in acute Stroke. *Stroke* 45(3):723–727
18. Lin K, Rapalino O, Law M et al (2008) Accuracy of the Alberta Stroke Program Early CT Score during the first 3 hours of middle cerebral artery stroke: comparison of noncontrast CT, CT angiography source images, and CT perfusion. *AJNR Am J Neuroradiol* 29(5):931–936
19. Powers WJ, Derdeyn CP, Biller J et al (2015) American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 46(10):3020–3035
20. Demaerschalk BM, Silver B, Wong E et al (2006) ASPECT scoring to estimate > 1/3 middle cerebral artery territory infarction. *Can J Neurol Sci* 33(2):200–204
21. Finlayson O, John V, Yeung R et al (2013) Interobserver agreement of ASPECT score distribution for noncontrast CT, CT angiography, and CT perfusion in acute Stroke. *Stroke* 44(1):234–236
22. Mak HK, Yau KK, Khong P-L et al (2003) Hypodensity of > 1/3 middle cerebral artery territory versus Alberta stroke programme early CT score (ASPECTS) comparison of two methods of quantitative evaluation of early CT changes in hyperacute ischemic stroke in the community setting. *Stroke* 34(5):1194–1196
23. Farzin B, Fahed R, Guilbert F et al (2016) Early CT changes in patients admitted for thrombectomy intrarater and interrater agreement. *Neurol* 87(3):249–256
24. Maegerlein C, Fischer J, Monch S et al (2019) Automated calculation of the Alberta stroke program early CT score: feasibility and reliability. *Radiology* 291(1):140–147
25. Herweh C, Ringleb PA, Rauch G et al (2016) Performance of e-ASPECTS software in comparison to that of stroke physicians on assessing CT scans of acute ischemic stroke patients. *Int J Stroke* 11(4):438–445
26. Kuang H, Najm M, Chakraborty D et al (2019) Automated ASPECTS on noncontrast CT scans in patients with acute ischemic stroke using machine learning. *AJNR Am J Neuroradiol* 40(1):33–38
27. Kuang HL, Menon BK, Il Sohn S, Qiu W (2021) EIS-Net: segmenting early infarct and scoring ASPECTS simultaneously on non-contrast CT of patients with acute ischemic stroke. *Med Image Anal* 70:101984
28. Naganuma M, Tachibana A, Fuchigami T et al (2021) Alberta stroke program early CT score calculation using the deep learning-based brain hemisphere comparison algorithm. *J Stroke Cerebrovasc Dis* 30(7):105791
29. Nagel S, Sinha D, Day D et al (2017) 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* 12(6):615–622
30. Goebel J, Stenzel E, Guberina N et al (2018) Automated ASPECT rating: comparison between the frontier ASPECT score software and the Brainomix software. *Neuroradiology* 60(12):1267–1272
31. Mikhail P, Le MGD, Mair G (2020) Computational image analysis of nonenhanced computed tomography for acute ischaemic stroke: a systematic review. *J Stroke Cerebrovasc Dis* 29(5):104715
32. Ardila D, Kiraly AP, Bharadwaj S et al (2019) End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 25(6):954–961
33. Chilamkurthy S, Ghosh R, Tanamala S et al (2018) Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* 392(10162):2388–2396
34. McKinney SM, Sieniek M, Godbole V et al (2020) International evaluation of an AI system for breast cancer screening. *Nature* 577(7788):89–94
35. Zhao W, Chen W, Li G et al (2022) GMILT: a novel transformer network that can noninvasively predict EGFR mutation status. *IEEE Trans Neural Netw Learn Syst*
36. Kim Y, Koh YJ, Lee C, Kim S, Kim CS (2015) Dark image enhancement based on pairwise target contrast and multi-scale detail boosting
37. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Interv* 9351:234–241
38. Kirillov A, Girshick R, He KM, Dollar P (2019) Panoptic feature pyramid networks. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (Cvpr 2019), pp 6392–6401
39. Lin TY, Dollar P, Girshick R et al (2017) Feature pyramid networks for object detection. In: 30th IEEE conference on computer vision and pattern recognition (Cvpr 2017), pp 936–944
40. Xie SN, Girshick R, Dollar P, Tu ZW, He KM (2017) Aggregated residual transformations for deep neural networks. In: 30th IEEE conference on computer vision and pattern recognition (Cvpr 2017), pp 5987–5995
41. Huang G, Liu Z, van der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: 30th IEEE conference on computer vision and pattern recognition (Cvpr 2017), pp 2261–2269
42. Collins DL, Zijdenbos AP, Baare WFC, Evans AC (1999) ANIMAL+INSECT: improved cortical structure segmentation. *Inform Process Med Imag Proc* 1613:210–223
43. Fonov V, Evans AC, Botteron K et al (2011) Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 54(1):313–327
44. Feng D, Cortese G, Baumgartner R (2017) A comparison of confidence/credible interval methods for the area under the ROC curve for continuous diagnostic tests with small sample size. *Stat Methods Med Res* 26(6):2603–2621
45. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15(2):155–163
46. Fuchigami T, Akahori S, Okatani T, Li YZ (2020) A hyperacute stroke segmentation method using 3D U-Net integrated with physicians' knowledge for NCCT. Paper presented at the medical imaging 2020: computer-aided diagnosis
47. Obuchowski NA, Bullen JA (2019) Statistical considerations for testing an AI algorithm used for prescreening lung CT images. *Contemp Clin Trials Commun* 16:100434

48. Kobbitsuksakul C, Tritanon O, Suraratdecha V (2018) Interobserver agreement between senior radiology resident, neuroradiology fellow, and experienced neuroradiologist in the rating of Alberta Stroke Program Early Computed Tomography Score (ASPECTS). *Diagn Interv Radiol* 24(2):104–107
49. McTaggart RA, Jovin TG, Lansberg MG et al (2015) 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* 46(2):407–412
50. Kuang HL, Qiu W, Najm M et al (2020) Validation of an automated ASPECTS method on non-contrast computed tomography scans of acute ischemic stroke patients. *Int J Stroke* 15(5):528–534
51. Neuhaus A, Seyedsaadat SM, Mihal D et al (2020) Region-specific agreement in ASPECTS estimation between neuroradiologists and e-ASPECTS software. *J Neurointerventional Surg* 12(7):720–723
52. Cheng XQ, Su XQ, Shi JQ et al (2021) Comparison of automated and manual DWI-ASPECTS in acute ischemic stroke: total and region-specific assessment. *Eur Radiol* 31(6):4130–4137

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)