Journal of Stroke and Cerebrovascular Diseases 33 (2024) 107512

Contents lists available at ScienceDirect

Journal of Stroke and Cerebrovascular Diseases

journal homepage: www.elsevier.com/locate/jstroke

Evaluating artificial intelligence software for delineating hemorrhage extent on CT brain imaging in stroke
AI delineation of ICH on CT

Adam Vacek [a], Grant Mair, MD [a,*], Philip White, MD [b], Philip M Bath, DSc [c], Keith W Muir, MD [d], Rustam Al-Shahi Salman, PhD [a], Chloe Martin [a], David Dye [a], Francesca M Chappell, PhD [a], Rüdiger von Kummer [e], Malcolm Macleod, PhD [a], Nikola Sprigg, DM [c], Joanna M Wardlaw, MD [a]

[a] Centre for Clinical Brain Sciences & UK Dementia Research Institute Centre, University of Edinburgh, UK
[b] Translational and Clinical Research Institute, Newcastle University, UK
[c] Stroke Trials Unit, Mental Health & Clinical Neuroscience, University of Nottingham, UK
[d] School of Psychology & Neuroscience, University of Glasgow, UK
[e] Department of Neuroradiology, University Hospital, Technische Universität Dresden, Germany

ARTICLE INFO

Keywords:
Artificial intelligence
CT
Stroke
Hemorrhage

ABSTRACT

Background: : The extent and distribution of intracranial hemorrhage (ICH) directly affects clinical management. Artificial intelligence (AI) software can detect and may delineate ICH extent on brain CT. We evaluated e-ASPECTS software (Brainomix Ltd.) performance for ICH delineation.
Methods: : We qualitatively assessed software delineation of ICH on CT using patients from six stroke trials. We assessed hemorrhage delineation in five compartments: lobar, deep, posterior fossa, intraventricular, extra-axial. We categorized delineation as excellent, good, moderate, or poor. We assessed quality of software delineation with number of affected compartments in univariate analysis (Kruskall-Wallis test) and ICH location using logistic regression (dependent variable: dichotomous delineation categories 'excellent-good' versus 'moderate-poor'), and report odds ratios (OR) and 95 % confidence intervals (95 %CI).
Results: : From 651 patients with ICH (median age 75 years, 53 % male), we included 628 with assessable CTs. Software delineation of ICH extent was 'excellent' in 189/628 (30 %), 'good' in 255/628 (41 %), 'moderate' in 127/628 (20 %), and 'poor' in 57/628 cases (9 %). The quality of software delineation of ICH was better when fewer compartments were affected (Z = 3.61-6.27; p = 0.0063). Software delineation of ICH extent was more likely to be 'excellent-good' quality when lobar alone (OR = 1.56, 95 %CI = 0.97-2.53) but 'moderate-poor' with any intraventricular (OR = 0.56, 95 %CI = 0.39-0.81, p = 0.002) or any extra-axial (OR = 0.41, 95 %CI = 0.27-0.62, p<0.001) extension.
Conclusions: : Delineation of ICH extent on stroke CT scans by AI software was excellent or good in 71 % of cases but was more likely to over- or under-estimate extent when ICH was either more extensive, intraventricular, or extra-axial.

Introduction

In patients with intracranial hemorrhage (ICH), higher ICH volume has poorer prognosis, and intraventricular extension may require neurosurgery.[1] Stroke imaging interpretation should be performed by trained practitioners (usually radiologists), who may not be immediately available. Artificial intelligence (AI) software may help less experienced clinicians detect ICH but also quantify its extent and distribution.[2,3] Our aim was to assess the delineation of ICH by one AI software.

Methods

Patient cohort

We include all patients from the RITeS (Real-world Independent Testing of e-ASPECTS Software) cohort – a clinically representative

* Corresponding author.
E-mail address: grant.mair@ed.ac.uk (G. Mair).

https://doi.org/10.1016/j.jstrokecerebrovasdis.2023.107512
Received 7 June 2023; Received in revised form 25 October 2023; Accepted 21 November 2023
Available online 25 November 2023

sample of patients from 9 acute stroke studies[4] - with ICH on baseline non-enhanced CT brain imaging. The original study population predominantly includes patients with ischemic stroke or spontaneous brain hemorrhage, but also included a few patients with subdural hemorrhage (n = 28) who presented with acute stroke symptoms. RITeS does not include patients with primary subarachnoid hemorrhage or hemorrhage secondary to venous sinus thrombosis or trauma. Six studies in RITeS included patients with baseline ICH: EuroHYP (n = 1)[5], IST-3 (n = 4)[6], LINCHPIN (n = 404)[7], PRACTISE (n = 26)[8], RESTART (n = 203)[9], RIGHT2 (n = 13)[10]. Each study had an expert panel for imaging assessment. Prior to RITeS, ICH was identified on each scan by a single expert, masked to patient details and without AI software. All studies individually obtained ethical approval. Consent was acquired from or on behalf of all participants.

AI-software

All scans were anonymized before upload for cloud-based processing by commercially available AI software: e-ASPECTS (Brainomix Ltd, UK), version 10.0.28. We selected the most relevant CT per patient (first acquired after stroke, thinnest axial slices processed for soft-tissue viewing) for processing by e-ASPECTS without knowing patient or imaging characteristics (including quality). We recorded where software upload/processing was unsuccessful. The software provides an image overlay to indicate hemorrhage location and extent, Fig. 1.

Qualitative assessment of ICH delineation by software

Following software processing, a trained observer – a single medical student who underwent two weeks of additional training in neuroanatomy and brain hemorrhage evaluation on CT (Edinburgh Criteria for Cerebral Amyloid Angiopathy-associated Intracranial Haemorrhage Training) – viewed side-by-side, the original CT and the CT with AI software overlay, and assessed (with neuroradiologist support, as required) the presence of hemorrhage in five compartments by reviewing 11 discrete locations (scoring each +/- for hemorrhage):

- Lobar – Frontal, parietal, temporal, occipital
- Deep – Basal ganglia
- Posterior fossa – cerebellum, brainstem
- Intraventricular
- Extra-axial – subarachnoid, subdural, extradural

Hemorrhage with reduced CT attenuation on visual inspection (i.e. less hyperattenuating than expected for acute hemorrhage, or if iso-attenuating to brain) or with internal layering was marked as subacute. Intra-rater reliability testing is conducted between two independent analyses of 52 randomly selected cases (roughly 10 % of the whole dataset) examined by the student observer, using Cohen's weighted kappa.

We devised a qualitative scale to assess software delineation of ICH extent:

- Excellent – identified all hemorrhage in all affected compartments.
- Good – identified some hemorrhage in all affected compartments.
- Moderate – missed hemorrhage in <2 compartments or misidentified normal structures as hemorrhage in <2 compartments.
- Poor – missed hemorrhage in ≥ 2 compartments or misidentified normal structures as hemorrhage in ≥ 2 compartments.

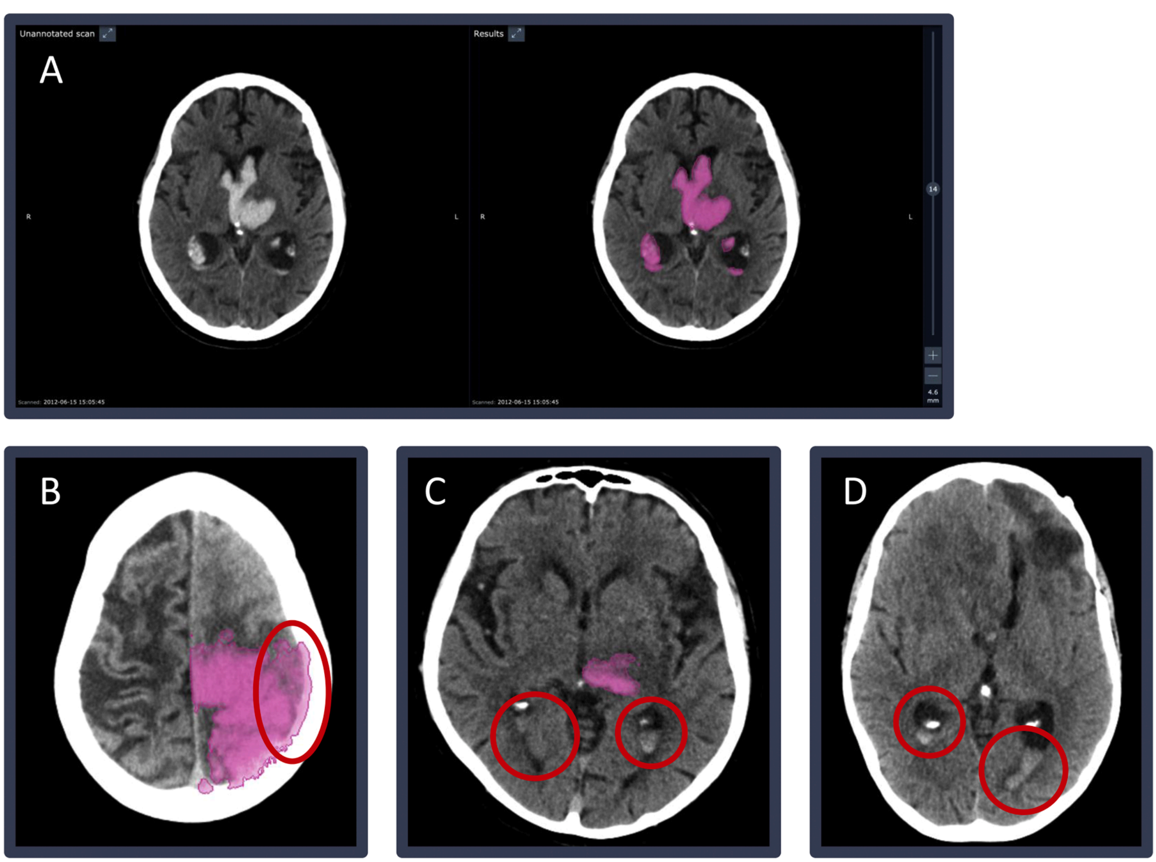

Fig. 1. ICH detection by software and categories used for assessment of extent delineation. A) CT without and with software overlay (pink) for hemorrhage detection, with excellent delineation of extent. B) Good detection (imperfect delineation). C) Moderate detection (some but not all hemorrhage detected). D) Poor detection (hemorrhage not detected).

Fig. 1 provides examples.

Statistical analysis

We used RStudio for MacOS (version 1.4.1106) for all analyses. We used Kruskal-Wallis test with Dunn's post-hoc examination to assess ICH delineation quality by CT slice thickness and number of locations affected (0-[11]), Z-value reported (Z). We sought association between software delineation quality and hemorrhage location using binary logistic regression with 'excellent-good' (software did not miss hemorrhage in any affected compartment) versus 'moderate-poor' delineation as the dependent variable, odds ratios (OR) with 95 % confidence intervals (95 %CI) reported. We report our results using TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis).

Results

We identified 651 patients in RITeS with ICH on CT, 53 % (n = 329) male, median age 75 years. We included 628 patients in the assessment of ICH delineation by software. Reasons for exclusion are presented in Fig. 2. Median CT slice thickness was 1 mm, with 89 % (n = 560/628) $\leq$2 mm. Most patients had hemorrhage in one (249/628, 39.6 %) or two (239/628, 38.1 %) compartments. In total, we identified 393 lobar, 297 deep, 75 posterior fossa, 246 intraventricular, and 153 extra-axial hemorrhages. Subacute hemorrhage was identified in 9 patients. The intra-rater reliability of the medical student was excellent (κ = 0.95, 95 % CI = 0.89-0.99, n = 52).

Software delineation of ICH extent was 'excellent' in 189/628 (30 %), 'good' in 255/628 (41 %), 'moderate' in 127/628 (20 %), and 'poor' in 57/628 cases (9 %). Hemorrhage was completely missed in 30 cases (5 %). The quality of software delineation was better with thinner CT slices (Z = 10.68, p = 0.014), and when fewer compartments/locations were affected by ICH (1-2 compartments mostly associated with 'excellent' delineation (Z = 6.27), 3-4 with 'good' (Z = 3.61), 5+ with 'moderate' (Z = 4.17; p = 0.0063). All 9 subacute hemorrhages were associated with 'poor' delineation.

The quality of software delineation was significantly reduced (more likely to be 'moderate-poor') in the presence of any intraventricular (OR = 0.56, 95 %CI = 0.39-0.81, p = 0.002) or extra-axial hemorrhage (OR = 0.41, 95 %CI = 0.27-0.62, p<0.001) whereas lobar (OR = 1.56, 95 % CI = 0.97-2.53) and posterior fossa (OR = 1.47, 95 %CI = 0.93-2.35) hemorrhage alone were non-significantly more likely to be 'excellent-good' quality. Posterior fossa hemorrhage was not associated with quality of software delineation (OR = 1.03, 95 %CI = 0.52-2.07), Fig. 3.

Discussion

In an analysis of one commercially available AI software for hemorrhage delineation on stroke CT, we found that ICH delineation was 'excellent' or 'good' in 444/628 (71 %) cases but 'moderate' or 'poor' in 184/628 cases (29 %). ICH distribution/extent, CT slice thickness and ICH acuity all affected software delineation; quality decreased with intraventricular or extra-axial hemorrhage, as more compartments were involved (i.e. it was less effective with more complex hemorrhage), and with thicker CT slices. Subacute hemorrhage was also poorly detected.

Our results have implications for clinical practice as higher volume hemorrhage and hemorrhage extension into the ventricles carry worse prognosis and may require prompt neurosurgical intervention.[1] Thus, we distinguished "excellent-good" software delineation of ICH from

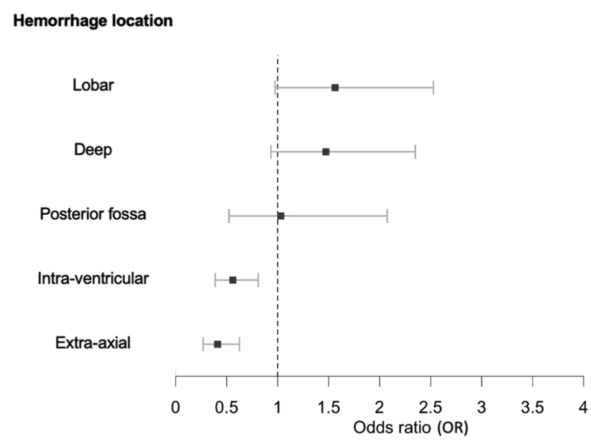

Fig. 3. Positive correlation association between software delineation quality and any hemorrhage location. OR>1 indicate 'excellent-good' delineation, OR<1 indicate 'moderate-poor' delineation.

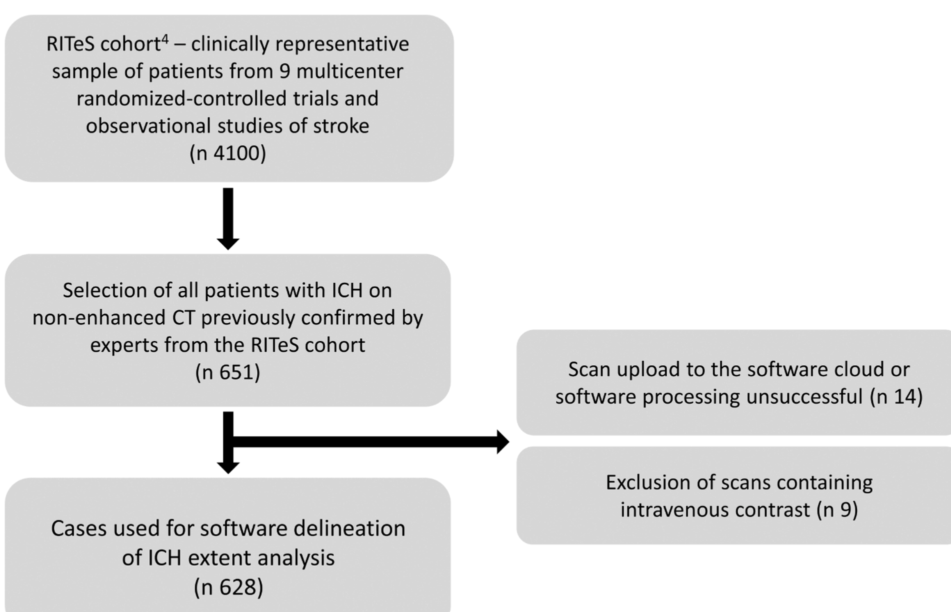

Fig. 2. Patient flow chart.

"moderate-poor" delineation since the latter includes all instances where hemorrhage was missed or falsely reported in at least one compartment, i.e. scenarios where a treatment decision may be modified. In the situation where ICH was completely missed (5 %), this could result in serious adverse clinical events (e.g. IV thrombolytic administration occurs because ICH in acute stroke patient missed). A previous study also found that a different AI software was less effective at delineating extra-axial and intraventricular hemorrhage.[11] Therefore, AI software should not be used standalone, but only used as intended to support experienced clinicians in their routine practice rather than to expand the capability of less experienced clinicians (i.e. raising, not lowering, the bar).

Various articles exploring application of AI for ICH evaluation are published. A systematic review of 40 studies looking at detection of ICH on non-enhanced CT reported sensitivity, specificity, and accuracy of 92 %, 94 %, and 93 %, respectively.[12] These results are similar to ours (see below), albeit for a range of software. In another analysis seeking to classify ICH subtypes defined by ICH distribution, an AI model had an area under the curve (AUC) of 0.8 across 5 ICH subtypes. This performance was superior to the average performance of junior radiology trainees, emphasizing the potential of AI to support less experienced clinicians.[13] Lastly, a recent study comparing prognosis prediction methods in patients with ICH, found AI models to have significantly higher AUCs (0.886-0.903) than visually assessed ICH scales (AUCs of 0.747-0.777).[14]

Our analysis has limitations. The retrospective design limits generalizability of our results. The absence of true negative cases for ICH in our cohort precludes the use of diagnostic accuracy statistics, i.e. we were unable to calculate sensitivity and specificity of software for hemorrhage detection. However, we previously tested and reported the diagnostic accuracy of e-ASPECTS in patients with and without hemorrhage[4] (for detection of hemorrhage software sensitivity was 97 %, and specificity was 83 %). The different compartments we assessed are not independent, hemorrhage spreads freely between them which may impact the results of our assessment. However, this design allowed us to assess delineation of ICH extent, as a marker of severity. We did not measure hemorrhage volume or assess ventricular dilatation, features that may also impact both software results and clinical management (guided by ICH severity). Presumed subacute hemorrhage was assessed qualitatively. While quantitative attenuation measurements would have provided a repeatable and more precise assessment, there is no standardized threshold to define subacute hemorrhage and our method was quick and is clinically relevant. At the time of testing, ICH detection by e-ASPECTS software (v10) was marked as experimental, was not commercially available, and may have been upgraded. Future evaluations of e-ASPECTS performance may include more comprehensive quantitative measures. For local software installations, CT protocols can better match software requirements (including slice thickness). Additionally, the software was not designed to be used for delineation of hemorrhage extent alone but rather as part of a wider stroke assessment. Prospective acute clinical studies are required to assess if AI might improve non-expert interpretation of CT in stroke and any clinical impact.

Conclusions

Delineation of ICH extent on stroke CT by e-ASPECTS AI software was excellent or good in 71 % of cases and poor in 9 %. The software was less effective when hemorrhage involved more compartments, extended outside the brain parenchyma or was subacute.

Sources of funding

The Stroke Association principally funded RITeS (TSA_CR_2017/01). The Medical Research Council (MC_PC_17188) supported cost of software licensing. Funders were not involved in the study.

Declaration of Competing Interest

GM: Consultancy fees from Canon Medical for stroke imaging software development. KWM: Institution has research agreement with Brainomix. PMB is Stroke Association Professor of Stroke Medicine and an emeritus NIHR Senior Investigator. PMW has no relevant disclosures. He reports institutional educational grants from Medtronic, Stryker and Penumbra in the last 3 years.

Acknowledgements

Brainomix were not involved in research planning. RITeS collaborators independently set aims and objectives, processed imaging, analyzed/interpreted results and drafted the report. We shared a pre-submission draft with Brainomix to ensure technical accuracy only.

References

1. Greenberg SM, Ziai WC, Cordonnier C, Dowlatshahi D, Francis B, Goldstein JN, Hemphill JC, Johnson R, Keigher KM, Mack WJ, et al. Guideline for the management of patients with spontaneous intracerebral hemorrhage: a guideline from the american heart association/american stroke association. *Stroke*. 2022;53: e282–e361, 2022.
2. Watanabe Y, Tanaka T, Nishida A, Takahashi H, Fujiwara M, Fujiwara T, Arisawa A, Yano H, Tomiyama N, Nakamura H, et al. Improvement of the diagnostic accuracy for intracranial haemorrhage using deep learning–based computer-assisted detection. *Neuroradiology*. 2021;63:713–720.
3. Strub WM, Leach JL, Tomsick T, Vagal A. Overnight preliminary head CT interpretations provided by residents: locations of misidentified intracranial hemorrhage. *AJNR*. 2007;28:1679–1682.
4. Mair G, White P, Bath PM, Muir KW, Al-Shahi Salman R, Martin C, Dye D, Chappell FM, Vacek A, von Kummer R, et al. External validation of e-ASPECTS software for interpreting brain CT in stroke. *Ann Neurol*. 2022;92:943–957.
5. van der Worp HB, Macleod MR, Bath PMW, Demotes J, Durand-Zaleski I, Gebhardt B, Gluud C, Kollmar R, Krieger DW, Lees KR, et al. EuroHYP-1: European multicenter, randomized, phase III clinical trial of therapeutic hypothermia plus best medical treatment vs. best medical treatment alone for acute ischemic stroke. *Int J Stroke*. 2014;9:642–645.
6. The IST-3 collaborative group. The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial. *Lancet North Am Ed*. 2012;379:2352–2363.
7. Samarasekera N, Lerpiniere C, Fonville AF, Farrall AJ, Wardlaw JM, White PM, Torgersen A, Ironside JW, Smith C, Al-Shahi Salman R. Consent for brain tissue donation after intracerebral haemorrhage: a community-based study. *PLoS One*. 2015;10, e0135043.
8. El-Tawil S, Wardlaw J, Ford I, Mair G, Robinson T, Kalra L, Muir KW. Penumbra and re-canalization acute computed tomography in ischemic stroke evaluation: PRACTISE study protocol. *Int J Stroke*. 2017;12:671–678.
9. Al-Shahi Salman R, Dennis M, Sandercock P, Sudlow C, Wardlaw J, Whiteley W, Murray G, Stephen J, Newby D, Sprigg N, et al. Effects of antiplatelet therapy after stroke due to intracerebral haemorrhage (RESTART): a randomised, open-label trial. *Lancet North Am Ed*. 2019;393:2613–2623.
10. Bath PM, Scutt P, Anderson CS, Appleton JP, Berge E, Cala L, Dixon M, England TM, Godolphin PJ, Havard D, et al. Prehospital transdermal glyceryl trinitrate in patients with ultra-acute presumed stroke (RIGHT-2): an ambulance-based, randomised, sham-controlled, blinded, phase 3 trial. *Lancet North Am Ed*. 2019;393:1009–1020.
11. Hssayeni MD, Croock MS, Salman AD, HF Al-khafaji, Yahya ZA, Ghoraani B. Intracranial hemorrhage segmentation using a deep convolutional model. *Data*. 2020;5:14.
12. Matsoukas S, Scaggiante J, Schuldt BR, Smith CJ, Chennareddy S, Kalagara R, Majidi S, Bederson JB, Johanna TF, Mocco J, Kellner CP. Accuracy of artificial intelligence for the detection of intracranial hemorrhage and chronic cerebral microbleeds: a systematic review and pooled analysis. *Radiol Med (Torino)*. 2022; 127:1106–1123.
13. Ye H, Gao F, Yin Y, Guo D, Zhao P, Lu Y, Wang X, Bai J, Cao K, Song Q, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur Radiol*. 2019;29:6191–6201.
14. Chen Y, Jiang C, Chang J, Qin C, Zhang Q, Ye Z, Li Z, Tian F, Ma W, Feng M, et al. An artificial intelligence-based prognostic prediction model for hemorrhagic stroke. *Eur J Radiol*. 2023;167, 111081.