Stroke

TOPICAL REVIEW

Section Editors: David S. Liebeskind, MD, and Joanna M. Wardlaw, MD

Accuracy of Automated Computer-Aided Diagnosis for Stroke Imaging: A Critical Evaluation of Current Evidence

Joanna M. Wardlaw, MD; Grant Mair, MD; Rüdiger von Kummer, MD; Michelle C. Williams, PhD; Wenwen Li, PhD; Amos J. Storkey, PhD; Emanuel Trucco, PhD; David S. Liebeskind, MD; Andrew Farrall, PhD; Philip M. Bath, DSc; Philip White, MD

ABSTRACT: There is increasing interest in computer applications, using artificial intelligence methodologies, to perform health care tasks previously performed by humans, particularly in medical imaging for diagnosis. In stroke, there are now commercial artificial intelligence software for use with computed tomography or MR imaging to identify acute ischemic brain tissue pathology, arterial obstruction on computed tomography angiography or as hyperattenuated arteries on computed tomography, brain hemorrhage, or size of perfusion defects. A rapid, accurate diagnosis may aid treatment decisions for individual patients and could improve outcome if it leads to effective and safe treatment; or conversely, to disaster if a delayed or incorrect diagnosis results in inappropriate treatment. Despite this potential clinical impact, diagnostic tools including artificial intelligence methods are not subjected to the same clinical evaluation standards as are mandatory for drugs. Here, we provide an evidence-based review of the pros and cons of commercially available automated methods for medical imaging diagnosis, including those based on artificial intelligence, to diagnose acute brain pathology on computed tomography or magnetic resonance imaging in patients with stroke.

Key Words: artificial intelligence ■ brain ■ machine learning ■ perfusion ■ stroke

Stroke is common, hospitals are busy, delays equal lost brain; diagnosis of the cause should be rapid so that appropriate treatment can start and give the patient the best chance of independent survival. Brain imaging is essential to differentiate ischemic from hemorrhagic stroke and stroke mimics. Furthermore, with advances in treatment options for specific patient subgroups, it is not enough just to identify ischemia or hemorrhage, since the size of the acute lesion, presence of other features (obstructed arteries, mass effect), and prestroke changes (leukoaraiosis, old infarcts, brain atrophy) also influence management. Most acute general hospitals assess several patients with suspected stroke each day, meaning that all steps in the process, including diagnosis, should be rapid, timely, efficient, and accurate.

It takes many years to train a neuroradiologist, they are scarce in many countries and serve many disease areas in addition to stroke. Vascular neurologists and stroke physicians learn to recognize early features of ischemic brain on scanning and major contraindications to reperfusion treatment. However, early ischemic changes on noncontrast computed tomography (CT) can be subtle and complex, with serious implications hanging on their correct identification, fueling interest in ways to improve their recognition and quantification. Methods such as perfusion CT require post-processing to generate a diagnostic image, highlighting abnormalities such as thresholded tissue blood flow. Such cleaned images may seem more user friendly and may facilitate rapid interpretation.

Alongside these longstanding pressures to reduce time and increase diagnostic accuracy, there have been substantial advances in computer vision and artificial intelligence

Correspondence to: Joanna M. Wardlaw, MD, Centre for Clinical Brain Sciences, University of Edinburgh, Chancellor's Building, 49 Little France Crescent, Edinburgh EH16 4SB, United Kingdom. Email joanna.wardlaw@ed.ac.uk

For Sources of Funding and Disclosures, see page 2401.

Stroke is available at www.ahajournals.org/journal/str

(AI) technologies across all walks of life. At its most general, AI refers to use of computers to solve problems in ways that mimic human behavior; machine learning (ML) is the key technology behind AI where computer algorithms learn from examples (ground truth) without explicit programming, which properties of the data are relevant for a given problem (feature selection); and deep learning (DL) is a subset of ML that uses biologically inspired neural networks to learn abstract high-order features in any type of data without requiring predetermined inputs.[1]

Medical imaging has been an obvious target for these developments.[2–4] Several commercial CT and MR scan diagnostic software for stroke are now in use in many hospitals. Nonetheless, the major demand for accelerated diagnosis in acute stroke, the fascination with the latest imaging tools, and huge potential financial gains for industry,[5] should not cloud the essential need to demonstrate that AI tools are accurate, are improving not impeding health care, and that the benefits outweigh the potential harms for patients.[6]

We assess the current evidence for AI diagnostic imaging tools in stroke, commercially available or in clinical use, the motivations, expectations, and work needed to underpin their implementation into clinical practice.

REVIEW OF EVIDENCE

What Could Automated AI Imaging Technologies Achieve in Acute Stroke Diagnosis?

AI technologies could improve accuracy, speed, and standardization of stroke diagnosis,[7–11] particularly in low throughput Centers,[12] and improve prognostication using quantitative measures of acute and chronic brain injury. AI tools could also improve workflows and communication along thrombectomy referral pathways, reduce time to treatment, improve clinical outcomes, and save clinicians and radiologists time.[2]

AI software could be most useful to assess the cause of stroke and its likely pathology:

1. Brain focal ischemia versus intracranial hemorrhage versus stroke mimics (migraine, seizure), the latter 2 requiring different management to ischemic stroke. Intracranial hemorrhages (brain hemorrhage, subarachnoid hemorrhage, subdural hematoma) can be visualized on CT and magnetic resonance imaging (MRI) with high accuracy. Stroke mimic is a clinical diagnosis based on the exclusion of brain ischemia and hemorrhage, noting that in patients with ischemic stroke symptoms, CT and MRI may seem normal, or show various degrees of reduced attenuation/altered signal intensity and/or swelling in the affected tissue.
2. Arterial pathology: site of acute embolic or thrombotic arterial occlusion, secondary features (eg, collateral blood flow), and underlying pathologies (atherosclerosis, dissection, inflammation, vasospasm). This is, perhaps, the most important determinant of acute therapy decisions and related prognosis assessment. Thin slice nonenhanced CT, angiography (CTA, MRA, DSA) and arterial wall imaging may differentiate between embolic or local arterial disease and have an impact on secondary stroke prophylaxis.
3. Early ischemic brain tissue alterations: isolated tissue swelling due to autoregulatory vasodilation,[13] tissue swelling with reduced attenuation (indicating early focal net water uptake or ionic edema triggered by critically low CBF <15 mL/100 g per minute, which cannot be tolerated by brain tissue for more than about 30 minutes), ion pump failure triggered by a CBF <30 mL/100 g per minute with consequent neuronal dysfunction and cellular edema indicated by ADC decrease/increase of DWI signal intensity. The still limited understanding of ischemia-effected brain tissue alterations makes it a potential mistake that algorithms are trained to detect brain infarction on nonenhanced CT or DWI within 6 hours of symptom onset,[14,15] since focal brain ischemia of up to 6-hour duration does not induce coagulation necrosis.[16,17]
4. Prestroke changes: leukoaraiosis (or white matter hyperintensities), prior infarcts or hemorrhages, and brain atrophy are all associated with worse short-term (hemorrhagic transformation, dependency, death)[18,19] and long-term (dependency, recurrent stroke, death, cognitive impairment, dementia) outcomes.[19]

Commonly used approaches for developing AI tools in stroke start with identifying attenuation or signal change that indicates ischemic tissue and its extent, adapted from visual rating tools. The Alberta Stroke Program Early CT Score (ASPECTS) is a widely used visual rating tool used to help diagnose acute ischemic stroke and select patients for thrombolysis and thrombectomy. Table 1 lists 19 vendors that currently provide 32 different commercially available AI software packages for ischemic stroke to assess a range of features including: ischemic tissue change (CT or MRI); hyperattenuated arteries (a surrogate for arterial thrombus); hemorrhage on CT; large artery occlusion on CTA; and salvageable tissue from CT or MR perfusion imaging. ASPECTS is used by 7 of 18 companies currently offering commercial stroke AI software,[7,20] despite several drawbacks of ASPECTS (not an independent outcome predictor,[18] variable cut points, validity[21]). Four companies provide comprehensive packages for handling nonenhanced, angiographic and perfusion imaging in one workflow (Brainomix, NICO.Lab, RapidAI, and VizAI), while 3 companies combine assessment of nonenhanced and angiographic imaging only (Aidoc, Avicenna, Circle Neurovascular Imaging). Other companies assess different combinations or individual components of ischemic stroke or hemorrhagic stroke (Table 1).

Table 1. Current Commercially Available AI-Based Software for Stroke

Company (country)	Product	Automated functions (imaging modality)	CE approval (class) for EU marketing	FDA approval (Class) for US marketing
Aidoc (Israel)	Briefcase	Detect hemorrhage (CT)	Yes (I)	Yes (II)
	Aidoc LVO	Detect LVO (CTA)	Yes (I)	Yes (II)
Avicenna.AI (France)	Cina-ASPECTS	Provide ASPECTS (CT)	Yes (I)	No
	Cina-ICH	Detect hemorrhage (CT)	Yes (I)	Yes (II)
	Cina-LVO	Detect LVO (CTA)	Yes (I)	Yes (II)
Brainomix (United Kingdom)	e-ASPECTS	Provide ASPECTS, detect dense MCA, detect hemorrhage (CT)	Yes (IIa)	No
	e-CTA	Detect LVO, provide collateral scoring (CTA)	Yes (IIa)	Yes (II)
	e-CTP	Process perfusion data (CTP)	Yes (IIa)	No
Cercare Medical (Denmark)	Cercare Stroke	Process perfusion data (CTP, MRP), detect ischemic lesions (MRI)	Yes (IIa)	No
Circle Neurovascular Imaging (Canada)	StrokeSENS	Provide ASPECTS (CT), provide collateral assessment, detect LVO (CTA)	Yes (II), does not include collaterals	Yes (I)*
Deep01 (Taiwan)	DeepCT	Detect hemorrhage (CT)	Yes (I)	Yes (II)
General Electric (United States)	Stroke VCAR	Detect hemorrhage (CT)	Yes (not declared)	Yes (II)
	CT Perfusion 4D	Process perfusion data (CTP)	Yes (not declared)	Yes (II)
Icometrix (Belgium)	Icobrain CVA	Process perfusion data (CTP)	Yes (I)	Yes (II)
Infervision Med Tech (China)	InferRead CT Stroke.AI	Provide ASPECTS, detect hemorrhage (CT)	Yes (IIa)	Yes (II)
JLK, Inc (South Korea)	JBS-01K	Detect ischemic lesions (MRI), provide stroke classification	Yes (I)	No
	JBS-04K	Detect hemorrhage (CT)	Yes (I)	No
Keya Medical (China)	CuraRad-ICH	Detect hemorrhage (CT)	No	Yes (II)
MaxQ.ai (Israel)	Accipio IX	Detect hemorrhage (CT)	Yes (not declared)	Yes (II)
NICO.Lab (Netherlands)	StrokeViewer	Provide ASPECTS, detect hemorrhage (CT), LVO detection, provide collateral assessment (CTA), process perfusion data (CTP)	Yes (I)	Yes (II)
Olea Medical (France)	Olea Sphere	Process perfusion data (CTP)	Yes (IIa)	Yes (II)
Qure.ai (India)	qER	Detect hemorrhage (CT)	Yes (IIa)	Yes (II)
RapidAI (United States)	Rapid	Process perfusion data and provide flow dynamics (CTA, CTP), detect DWI lesions (MRI)	Yes (I)	Yes (II)
	Rapid ASPECTS	Provide ASPECTS (CT)	Yes (I)	Yes (II)
	Rapid ICH	Detect hemorrhage (CT)	Yes (I)	Yes (II)
	Rapid LVO	Detect LVO (CTA)	Yes (I)	Yes (II)
Siemens (Germany)	syngo.CT ASPECTS	Provide ASPECTS (CT)	Yes (not declared)	No
	syngo.CT Neuro Perfusion	Process perfusion data (CTP)	Yes (not declared)	Yes (II)
Vizai (United States)	Viz ICH	Detect hemorrhage (CT)	Yes (not declared)	Yes (II)
	Viz LVO (ContaCT)	Detect LVO (CTA)	Yes (not declared)	Yes (II)
	Viz CTP	Process perfusion data (CTP)	Yes (not declared)	Yes (II)
Zebra Medical Vision (Israel)	HealthICH	Detect hemorrhage (CT)	Yes (not declared)	Yes (II)

Both CE and FDA classification use 3 classes depending on risk to the patient and/or user according to the intended use of the device: I=low risk, II=medium risk, III=high risk. Both classification systems have a more stringent process for classifying higher risk devices. For CE, I can be self-certified by manufacturer, II&III require audit of validation results by a notified body. For FDA, I and II require 501 k (prove equivalence to device already approved for marketing, or de novo assessment if novel) while III requires premarket approval (PMA) including software validation results. Details extracted from publicly available EU and FDA data and vendor websites, correct to August 31, 2021, modified February 21, 2022. AI indicates artificial intelligence; ASPECTS, Alberta Stroke Program Early CT Score; CT, computed tomography; CTA, computed tomography angiography; ICH, intracerebral hemorrhage; LVO, large vessel occlusion; and MRI, magnetic resonance imaging.

*Approved for data transfer only, not automated processing.

What Drives the Increasing Use of Automated Analysis Technologies?

There is undoubtedly a real need to aid busy doctors at the hospital front door, and radiologists are also increasingly pressured. However, 2 decades after early thrombolysis trials, clinical awareness of acute stroke imaging features is much more established. Therefore, any AI diagnostic tools have to be exceptionally fast, easy to use, and accurate on 'real world' data (see section

"What evidence is there that AI technologies will meet the needs of users, including community practitioner, neurologist, radiologist, and the patient?") to add value. AI tools also require users to be trained in their proper use and interpretation.

Commercially, data and AI offer massive financial gains for successful products.[5] Worldwide spending on AI was estimated at US$38 billion in 2019 and is predicted to rise to $98 billion by 2023.[22] Investments in AI-based medical imaging companies in the United States reached US$1.17 billion between January 2014 and January 2019, doubling since 2012 to 2017, while the number of companies in the AI market had trebled,[5] including new industries devoted to image classification (see section "What could improve translation of AI technologies into benefits for patients and health care systems?"). Major medical imaging manufacturers are incorporating AI tools into consoles to retain a marketing edge. AI requires data storage capacity and computing power: between Jan 2014 and Jan 2019, there were over US$435 billion-worth of cloud-based medical imaging transactions, indicating massive investments in these areas.[5]

What Clinically Relevant Testing Should Automated Analysis Undergo in Acute Stroke?

Currently, AI software for radiology can only be marketed in the European Economic Area (EEA) after achieving a CE mark (Conformitè Europëenne), indicating that the technology conforms with European health, safety, and environmental protection standards, and in the United States with FDA (Food and Drug Administration) approval. However, the clinical standards for achieving these certifications are low as compared with licensing a new drug. Both CE and FDA systems have different classes of approval depending on the perceived risk to the patient and to software users. Since radiology software have to date been designed to support rather than replace physicians, they require usually only Class I or II approval denoting low to medium risk. The decision support labeling raises important questions about what happens when the clinician disagrees with the AI diagnoses, and who is to blame when one or other gets it wrong. In both territories, Class I approval is awarded without external scrutiny and companies self-declare these products as compliant. While Class II approval usually includes the submission of evidence assessed by an independent body, companies can provide this evidence from their own internal testing without peer-review or publication. Indeed, a recent independent review of all CE-marked AI software for radiology in Europe found that 64 of 100 products had no published peer-reviewed evidence of efficacy.[3] It is beyond the scope of this article to assess the depth of peer-reviewed evidence for every available software, so we focus on commercial products with the most citations according to recent reviews (Brainomix, RapidAI, and Vizai).[8–11]

Reporting guidelines and methodological standards for developing AI in medical imaging are available[23] including SPIRIT-AI,[24] CONSORT-AI,[25] and checklists.[26] Several societies have released their own guidelines and position statements.[26,27] The FDA (https://www.fda.gov/media/145022/download), and British Standards Institute (BSI) and Medicines and Healthcare Regulatory Agency (MHRA) (https://standardsdevelopment.bsigroup.com/committees/50299208; pending public consultation) provide standards. Few if any evaluations of AI software are meeting these standards, generally or in stroke.[6,8,9,23,24,28,29] A recent systematic review of reporting quality of studies of ML in medical diagnosis found 28 includable studies but none mentioned a reporting guideline, few mentioned the distribution of disease severity or alternative diagnoses, most studies had a long delay between the reference standard and ML diagnoses, and in half of studies the population was of uncertain relevance to the clinical setting.[28] Five of the 28 included studies concerned brain imaging none of which addressed stroke.[28]

Accuracy of AI in Stroke

1. How accurate is AI software for differentiating acute ischemic from hemorrhagic stroke and stroke mimics, and on which imaging modalities?

Recent systematic reviews,[8,9] narrative reviews,[7,10–12,14] a review protocol,[30] and a pending combined analysis of 9 large stroke trials[31] show that most studies of AI have focused on ischemic not hemorrhagic stroke, and CT not MRI. The systematic reviews identified 20 (tissue and arterial changes[8]) and 68 (noncontrast CT only[9]) includable studies, but most studies were small, provided little documentation of the patient characteristics, recruitment or CT characteristics, and focused on comparing AI against feature detection by humans, not on clinical outcomes. Rates of failed scan processing were often omitted. Many methodological differences between studies precluded formal meta-analyses. All published studies and reviews focused on AI detection of ischemic stroke features including arterial obstruction,[7–9,12] with only one pending study[31] assessing if an AI software can differentiate ischemic from hemorrhagic stroke or mimics.

The extent to which AI software may be affected by patient-related (background brain changes or alternative brain pathology, movement and position during scanning, heart failure, metallic artifacts[32–35]) and imaging-related (scanner manufacturer, acquisition parameters such as slice thickness, or CT gantry position[34,36]) factors is beginning to emerge. These affect the likelihood of successful image processing, agreement with a reference standard, and strength of associations with clinical outcome. However, most published evidence excluded

Stroke. 2022;53:2393–2403. DOI: 10.1161/STROKEAHA.121.036204

difficult cases before analysis, while results of studies that retained difficult cases were often less positive.[32,37,38]

2. How good are these technologies for identifying key features of acute ischemic stroke that are of prognostic importance?

Few studies assessed AI software-based diagnoses and clinical outcomes. Among commercially available AI tools, we reviewed the published literature for the 3 most established providers offering comprehensive imaging packages: Brainomix, RapidAI, and Vizai (Table 2). We used published reviews[7–9,20] updated by searching Pubmed for company and software names, and the vendor's websites. We focus on studies with the largest test datasets (ideally >100 patients) and report diagnostic accuracy statistics for stroke feature detection.

Three studies assessed detection of tissue ischemia (all Brainomix, all retrospective, 2 independent of the company,[32,39] total patients n=367),[32,39,40] with sensitivities of 44% to 83% and specificities of 57% to 93% (Table 2). Six studies assessed detection of large vessel occlusion (LVO; 3 vendors, all retrospective, 1 independent of the vendor,[41] total n=2635)[38,41–45] with sensitivities of 80% to 96% and specificities of 90% to 98%. Only 1 study each assessed hemorrhage detection[46] and MRI diffusion or perfusion imaging.[47] Compared with optimal circumstances, the agreement between each of these software and experts was poorer in patients with leukoaraiosis, old infarcts, or other parenchymal defects.[7] This underscores the importance of evaluating AI tools in realistic and common clinical settings where patients are often older, have multiple conditions or delayed presentations,[48] and not relying on results when tested in the simplified training scenarios common in public datasets.[5,49]

3. On a population level, how many false positives or negatives might arise per 100 typical suspected strokes and what are the implications for patient outcomes?

Given the range of published sensitivity and specificity results for stroke feature detection by AI software above, we translate the results as follows (Figure). For every 100 patients assessed using these software:

- With ischemic stroke, ischemia will be correctly detected in 44 to 83 but missed in 17 to 56.
- Without ischemic stroke, ischemia will be incorrectly detected in 7 to 43.
- With LVO, occlusion will be correctly detected in 80 to 96 but missed in 4 to 20.
- Without LVO, occlusion will be incorrectly detected in 2 to 10.

4. Have (any) automated, including AI, technologies that are proposed for use in stroke, undergone proper prospective randomized blinded outcome clinical trial assessment to determine impact on clinical outcomes or health economics?

Randomized controlled trials of AI technologies are scarce and mostly ongoing: a recent survey of trials' registries and the literature identified only one RCT comparing DL with clinicians in medical imaging (on breast ultrasound).[6,29] We identified one ongoing multicenter RCT testing the impact of Viz LVO on stroke workflow and 90 day clinical outcome in 500 participants admitted with stroke and suspected LVO in the United States (ALERT [Automated Detection and Triage of Large Vessel Occlusions Using AI for Early and Rapid Treatment], URL: https://www.clinicaltrials.gov; Unique identifier: NCT04142879). Diagnostic accuracy is not a primary or specified secondary outcome in ALERT. Another ongoing multicenter RCT in China (GOLDEN BRIDGE II, URL: https://www.clinicaltrials.gov; Unique identifier: NCT04524624) is testing AI identification of stroke on diffusion imaging plus decision support versus usual care in 21 689 patients requiring secondary stroke prevention.

Three studies compared times to thrombectomy before and after introduction of RAPID[41,50] or Viz LVO,[51] reporting average reductions of 30 minutes to groin puncture; however, all were retrospective, 2 only report the small numbers of patients who all received thrombectomy, and before-after comparisons are unable to address many sources of bias.

How Do Stroke AI Technologies Compare With Other Medical AI Technologies, Particularly Medical Imaging AI, in Terms of Stage of Development and Quality and Thoroughness of Assessment?

Stroke AI is similar to other medical imaging AI—great hopes but important challenges for delivery into clinical practice. These challenges reflect data curation, model development, relevance to clinical practice, potential to introduce and amplify biases, the AI tool's transparency, and evidence of accuracy, impact on outcomes and cost effectiveness meeting RCT evidence standards.

The quality, quantity, diversity, and provenance of the data used to train a ML model are critical to its utility in clinical practice. Many current papers describe ML models trained on one small dataset from one hospital,[52,53] insufficient to be useful on the variety seen in clinical practice.[53] Commercial ML models have similar issues, particularly black box models where the training is not described.

Datasets created as part of international challenges, for example, the RSNA 2019 Brain CT Hemorrhage Challenge[54] (874 035 images, multiple institutions) are limited in diversity, accuracy, and reproducibility.[55] Overtuning of software to particular datasets (overfitting) impedes generalization. Sometimes AI models identify confounders rather than target disease, for example, hip fracture detection, where scanner model (emergency department) and priority request marker predicted

Table 2. Accuracy of 3 Commercially Available AI-Based Software in Stroke

Company	Software	Detection of ischemic brain injury: CT	Detection of ischemic brain injury: MRI/CTP	Detection of LVO	Detection of hemorrhage
Brainomix	e-ASPECTS	Retrospective, 132, 44%–45%, 91%–93%, follow-up CT[40]	†	†	†
		Retrospective,* 119, 83%, 57%, follow-up CT[32]			
		Retrospective,* 116, 75%, 73%, experts with all data including follow-up CT or MRI.[39]			
	e-CTA	†	†	Retrospective, 301, 84%, 96%, experts with all data including follow-up imaging.[42]	†
	e-CTP	†	†	†	†
Rapid	Rapid	†	Retrospective, 63, 100%, 91%, experts[47]	†	†
	Rapid ASPECTS	†		†	†
	Rapid ICH	†	†	†	Retrospective, 308, 96%, 95%, expert consensus[46]
	Rapid LVO	†	†	Retrospective,* 310, 80%, NS, experts[41]	†
				Retrospective, 217, 96%, 98%, experts[43]	
				Retrospective, 477, 92%–94%, 97%–98%, experts[44]	
Vizai	Viz ICH	†	†		†
	Viz LVO (ContaCT)	†	†	Retrospective, 1167, 81%–82%, 90%–96%, experts[38]	†
				Retrospective, 163, 96%, 94%, NS[45]	
	Viz CTP	†	†		†
SUMMARY		367 patients, 1 software		2635 patients, 3 different software	
		Sensitivity=44%–83%			
		Specificity=57%–93%		Sensitivity=80%–96%	
				Specificity=90%–=]\98%	

Results are (study design, n, sensitivity, specificity, reference standard), unless otherwise stated. AI indicates artificial intelligence; ASPECTS, Alberta Stroke Program Early CT Score; CT, computed tomography; ICH, intracerebral hemorrhage; LVO, large vessel occlusion; MRI, magnetic resonance imaging; and NS, not specified.

*Study conducted independent of company.

†No suitable papers were identified or expected.

fracture better than the imaging findings.[56] Explainable AI models may help to show the underlying features identified by a DL model to avoid black box problems.

Bias emerging during the development of any automated system is common in AI studies[57] and can reflect issues with the underlying datasets or model development techniques. Numerous examples of biases have emerged, including those related to sex,[58] race,[59] and geography,[60] which could inadvertently exacerbate underlying health care inequalities.[61]

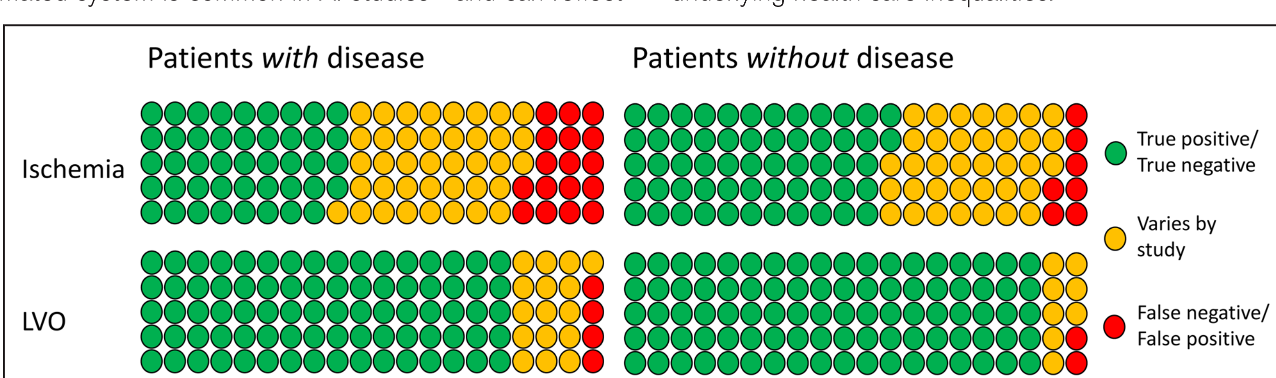

Figure. Potential clinical implications of artificial intelligence software use for stroke feature detection per 100 patients assessed, derived from data in Table 2.

Orange circles indicate the overlapping range of values provided by different studies. LVO indicates large vessel occlusion.

There are increasing suggestions that AI can perform better than humans in medical tasks. A recent systematic review found 10 RCTs testing DL versus clinicians (2 completed, 8 ongoing), only one of which was in medical imaging (breast ultrasound, ongoing).[6] In contrast, they found 81 nonrandomized comparisons (9 prospective, 6 relevant clinical environment). The AI was said to be better than or comparable to the clinician in 47% of the 81 studies. However, development and testing often used the same dataset, had small numbers of human comparators (eg, 5), most studies had high risk of bias, and few adhered to reporting standards.[6]

Another comparison of the diagnostic accuracy of DL versus clinicians identified 82 studies in which it was possible to calculate accuracy in 69. Mean sensitivity was 79.1% (range, 9.7%–100%), specificity was 88.3% (range, 38.9%–100%),[29] but many studies did not compare the DL and clinicians on the same data, did not externally validate their results, or report their methods adequately. Among the 14 studies with external validation that tested DL and clinicians on the same sample, the pooled sensitivity was 87.0% (95% CI, 83.0%–90.2%) versus 86.4% (79.9%–91.0%), and pooled specificity was 92.5% (85.1%–96.4%) versus 90.5% (80.6%–95.7%) for DL versus clinicians, respectively.[29] Of note, there were no studies of AI in stroke among the 82 studies, and only 2 studies concerned brain imaging (1 MRI in dementia, 1 CT in head trauma).

What Evidence Is There That AI Technologies Will Meet the Needs of Users, Including Community Practitioner, Neurologist, Radiologist, and the Patient?

Do AI tools reduce door to needle time? Or might AI tools worsen treatment delays, or deny some patients effective treatments?[29] There are no completed prospective randomized trials of the impact of stroke AI tools on workflows or clinical outcomes, only before-after evaluations (see "Have (any) automated, including AI, technologies that are proposed for use in stroke, undergone proper prospective randomised blinded outcome clinical trial assessment to determine impact on clinical outcomes or health economics?" above).[41,50,51] It is common in hospitals for new digital systems to slow, not accelerate, workflows.[62] Attention of AI should focus on improving routine medical workflows including image management and electronic case records to reduce time wasted, improve information content and diagnostic utility.[2,63] Some AI tools perform tasks that are not that helpful,[64] provide clinically irrelevant measures, or operate slower than a seasoned user of existing medical computing systems, delaying uptake of AI into general radiology.[5]

Different algorithms may give different results. Comparison of 3 AI tools for ASPECTS scoring[7,20] showed the highest correlation between the expert read and Brainomix (ICC=0.871 [0.818–0.909]; P<0.001) but comparable area-under-curve between the AI applications and expert consensus (Brainomix: AUC, 0.759 (0.670–0.848), P<0.001; Frontier V2: AUC, 0.752 (0.660–0.843), P<0.001; RAPID: AUC, 0.734 (0.634–0.831), P<0.001). AI software may help less experienced doctors interpret acute stroke CT scans, for example, use of Syngo.via Frontier ASPECT Score Prototype improved the correlation between junior radiologists and the reference standard from good (r=0.680) to excellent (r=0.852) in 1 small study.[65] Apparently, good performance at group level may mask important variation at the individual level. Among 12 ML and 7 statistical models to predict cardiovascular disease risk using data from 3.6 million patients, the models had similar population level performance (C statistics of about 0.87) but varied widely in their prediction of individual risks particularly at higher risks[49] and compared with the risk predicted by a reference model, about 60% of patients would have been classed as lower risk by another model.[49]

AI tool evaluation is usually restricted to single technical measures in controlled settings that only indirectly relate to the intended clinical tool use. However, the clinical need is rarely raw classification but rather diagnostic or therapeutic decision support.[66] DL tools are notoriously sensitive to changes in input characteristics and not customarily stress-tested across different scanners, parameters, clinics, patient groups, preprocessing tools, etc, further hindered by the black-box nature of DL methods and commercial confidentiality.[67] The benefits and challenges of introducing computational innovations into existing clinical ecosystems[62] remains sparsely assessed; AI software may require specific tailoring to suit different settings and institutions.[68]

There is little participant involvement in development of AI software[69] despite concerns.[70] Few people want to receive a terminal diagnosis from a robot,[71] or want major treatment decisions for a life-threatening disease (like stroke) to be based primarily on AI-determined scan characteristics, particularly when these differ between AI software.[7,72] Increasing availability of AI diagnostic tools including front line use by less experienced doctors risks subverting medical judgement through inflexible application of easy-to-derive threshold values—for example, treat if ≤70 mL core, not if >70 mL. Treatment decisions with the very powerful reperfusion therapies now available for stroke must consider the whole patient and not place inappropriate weight on perfusion maps that only represent brief snapshots in time of a highly dynamic disorder, particularly when use of a different software is very likely to give different results.[7]

What Could Improve Translation of AI Technologies Into Benefits for Patients and Health Care Systems?

There is a gulf between AI software developers and the intended clinical uses of the software, and a need for consensus-based, interdisciplinary and comprehensive scoping of user needs and constraints to guide effective development of AI tools.[24,25,63] The accuracy of an AI tool depends on its data input. AI developers highlight the need for ground truth for training the software, meaning images where relevant features have been demarcated, often by hand, and in large datasets.[54,55] While large collections of medical imaging data are increasingly available, manual annotation is a massive, time consuming task and must be done by humans, therefore not many large annotated datasets exist. Some companies are outsourcing the work to low paid workers in low income countries.[22] The market for data-labeling services may triple to US$5billion by 2023, with companies like Mechanical Turk (owned by Amazon) providing freelancers ready to perform micro-tasks like tagging images, or Hive, which runs online data labeling games where operators earn money for labeling features,[22] which questions the reliability of the ground truth thereby derived.

Can we accelerate reliable, representative and diverse dataset availability, and, is ground truth really essential, or could correlated variables like clinical outcomes be used instead? AI tools could be even more valuable if they could discover novel features or markers of severity, or predict clinically relevant outcomes and treatment response, and thus improve clinical management. Table 3 lists important factors, including more accessible large-scale data, standardization of preprocessing pipelines, sharing open source codes, adoption of guidelines for reporting of AI development, closer working between ML/DL developers and clinicians, and better standards for evaluating AI tools against relevant clinical outcomes, control of confounders, correlates and colliders that impede AI performance.

DISCUSSION AND CONCLUSIONS

While AI tools hold great promise in stroke, much more work is required to demonstrate their clinical value and cost-effectiveness to patients, doctors, and health care providers. AI development requires more focus on multidisciplinary teams including AI experts, IT experts, radiologists, and strokologists[63] listening to each other carefully. Currently, without a more cohesive multidisciplinary effort, it seems that the stroke-AI is at risk of Verschlimmbesserung, meaning an attempted advance without improvement, or even with worsening. Imaging and AI image analysis should demonstrate clearly through objective, relevant, and reliable evidence that improve patient management and clinical outcomes.

Digital methods complement but cannot replace the human touch in medicine.[70] Patients should be actively involved in the design and evaluation of AI tools usage, since patient groups are clearly very wary

Table 3. Steps Needed to Accelerate AI Software Development

Need	Comment
More accessible large-scale data repositories	RSNA (Radiology Society of North America) Head CT Challenge[54]: ICH detection (>800 000 scans)
	ASFNR (American Society of Functional Neuroradiology) Head CT Challenge: CT pathology at different ages (https://aichallenge.asfnr.org/)
	ISLES (Ischemic Stroke Lesion Segmentation) Challenge (n=75)[73]
	Large trials, for example, Third International Stroke Trial (IST-3) (3035 pts, >10 000 scans)
	UK Biobank (ukbiobank.ac.uk); HDR UK (Health Data Research)
Guidelines to standardize preprocessing pipelines[74]	Preprocessing steps: reading DICOM data, converting DICOM to other formats, brain extraction from skull, defacing, registration, etc
	RSNA 2019 ICH detection challenge[54] shows many common problems that can occur, for example, under-labeling data, human errors, imbalanced classes, inappropriate de-identification and anonymization across data sources.
Guidelines and standards for reporting of machine learning models[6,63]	Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis–statement specific to machine learning (TRIPOD-ML), Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT)-AI and Consolidated Standards of Reporting Trials (CONSORT)-AI.[25]
Collaboration between ML practitioners and all relevant disciplines	Physicians and radiologists benefit from becoming familiar with basic concepts in machine learning.[63] Machine learning methods must be developed by those familiar with details and context of the real clinical settings including decision triage, practicalities of the specific patient population, typical medical imaging, image interpretation, definition of treatment relevant imaging finding, and data preparation.
Testing and validation protocols	Should move beyond technical performance evaluation (eg, ROC, accuracy, sensitivity/specificity) to determine the clinical value of a system, given limited annotator availability, experience, accuracy, and interest. The limits of testing criteria used in international challenges are evident.[75]
Attuned algorithmic development	Move beyond classifiers: methods must control for confounders, correlates and colliders that introduce bias and produce nonrobust methods that collapse with potentially dangerous consequences when deployed in different real settings[76]
Open-source AI code and data	Proprietary code impedes replication, reproducibility, clinical validation: it is difficult and costly.[63] Commercial development based on ill-established methods that the community cannot verify in clinical settings risks reputational damage.

AI indicates artificial intelligence; CT, computed tomography; and ICH, intracerebral hemorrhage.

of too much unregulated use of computers in clinical decision-making.[71,72]

Not all measurement has value. Rather, consider what features would be treatment relevant to detect and quantify. Imaging findings in stroke patients help us chose the most effective treatment. Spontaneous brain hemorrhage requires angiography to find and treat a vascular malformation or aneurysm. Exclusion of hemorrhage but thrombus within the MCA requires immediate thrombectomy. Stroke imaging is complex, not a single feature process, and perhaps a more difficult place to initiate AI tool development than it might seem superficially. Low ASPECTS is an independent predictor of poor outcome, but patients may still benefit from treatment, therefore reducing information to a single binary variable such as ASPECTS score would seem to be a retrograde step.

Can costs of AI tools be realistic? Currently, one typical commercial AI software costs around US$47 868 for 1 hospital for 1 year in the United Kingdom, equivalent to about a third of a hospital consultant's salary, and seems an unreasonable amount of money for something which only identifies a few features in one disease, should only be used by an experienced medic, and thus does not replace anything or anyone, and has a limited evidence base.

The essential next steps are first, to be aware of the limitations where commercial AI tools are in use, and second, to obtain reliable evidence of benefits versus harms of imaging AI tools' performance. This would best be tested in large scale randomised trials, to minimize bias, and in the clinical settings in which they will be used to ensure applicability to real world clinical practice. It is no longer appropriate to show only diagnostic accuracy on selected retrospective datasets, without reporting the failures. We need to see how AI tools work in clinical practice, how they integrate into patient care, and if and how much they are beneficial, through providing evidence that they change management for the better, improving outcomes, and are cost effective.

ARTICLE INFORMATION

Affiliations

Centre for Clinical Brain Sciences, UK Dementia Research Institute Centre at the University of Edinburgh, Little France, United Kingdom (J.M.W., G.M., W.L., A.F.). Institute of Diagnostic and Interventional Neuroradiology, Universitätsklinikum Carl Gustav Carus, Dresden, Germany (R.v.K.). Centre for Cardiovascular Science, University of Edinburgh, Little France, United Kingdom (M.C.W.). School of Informatics, University of Edinburgh (A.J.S.). VAMPIRE project, Computing, School of Science and Engineering, University of Dundee (E.T.). Neurovascular Imaging Res Core, UCLA, Los Angeles, CA (D.S.L.). Stroke Trials Unit, Mental Health & Clinical Neuroscience, University of Nottingham, Queen's Medical Centre campus, United Kingdom (P.M.B.). Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne and Newcastle upon Tyne Hospitals NHS Trust, United Kingdom (P.W.).

Acknowledgments

P.M. Bath is Stroke Association Professor of Stroke Medicine and an Emeritus NIHR Senior Investigator.

Sources of Funding

Dr Wardlaw is supported by the UK Dementia Research Institute, which receives its funding from DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK. Dr Mair is Stroke Association Edith Murphy Foundation Senior Clinical Lecturer (SA L-SMP 18\1000). We acknowledge funding from the Stroke Association (TSA_CR_2017/01) and Medical Research Council Proximity to Discovery fund (MC_PC_17188) for the RITeS Study. Dr von Kummer receives royalties as Editor-in-Chief of Neuroradiology. Dr Li is funded by Health Data Research UK, which receives its funding from the Medical Research Council. DR Williams is supported by the British Heart Foundation (FS/ICRF/20/26002). P.M. Bath is supported by the British Heart Foundation and NIHR Health Technology Programme.

Disclosures

Dr Mair received speaker and consulting fees from Canon Medical Ltd, Europe. PMB provided consultancy for Diamedica, Phagenesis, Moleac, Sanofi, Nestle. MCW received speaker fees from Canon Medical Systems. Dr Liebeskind consultants for imaging core laboratory services to Cerenovus, Genentech, Medtronic, Stryker and Rapid Medical. Dr White reports institutional unrestricted educational grants from Medtronic, Stryker and Penumbra; Institutional research funding from Microvention and IBEX and consulting fees/core laboratory services from Microvention and E-VASC. Dr Farrall received institutional educational grants from Medtronic for the development of training materials in stroke radiology. Dr Storkey has received funding from Huawei, ARM and Microsoft Research, Cambridge not related to the topic of the current article. Dr Wardlaw holds institutional academic grants from the Fondation Leducq, EU H2020 programme, Stroke Association, Health Data Research UK, British Heart Foundation, Alzheimer's Society, Weston Brain Institute, and Row Fogo Charitable Trust for research unrelated to the.

REFERENCES

1. Chen D, Liu S, Kingsbury P, Sohn S, Storlie CB, Habermann EB, Naessens JM, Larson DW, Liu H. Deep learning and alternative learning strategies for retrospective real-world clinical data. *NPJ Digit Med*. 2019;2:43. doi: 10.1038/s41746-019-0122-0
2. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet*. 2020;395:1579–1586. doi: 10.1016/S0140-6736(20)30226-9
3. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol*. 2021;31:3797–3804. doi: 10.1007/s00330-021-07892-z
4. Yao AD, Cheng DL, Pan I, Kitamura F. Deep learning in leuroradiology: a systematic review of current algorithms and approaches for the new wave of imaging technology. *Radiol Artif Intell*. 2020;2:e190026. doi: 10.1148/ryai.2020190026
5. Alexander A, Jiang A, Ferreira C, Zurkiya D. An intelligent future for medical imaging: a market outlook on artificial intelligence for medical imaging. *J Am Coll Radiol*. 2020;17(1 Pt B):165–170. doi: 10.1016/j.jacr.2019.07.019
6. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, Topol EJ, Ioannidis JPA, Collins GS, Maruthappu M. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020;368:m689. doi: 10.1136/bmj.m689
7. Soun JE, Chow DS, Nagamine M, Takhtawala RS, Filippi CG, Yu W, Chang PD. Artificial intelligence and acute stroke imaging. *AJNR Am J Neuroradiol*. 2021;42:2–11. doi: 10.3174/ajnr.A6883
8. Murray NM, Unberath M, Hager GD, Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerv Surg*. 2020;12:156–164. doi: 10.1136/neurintsurg-2019-015135
9. Mikhail P, Le MGD, Mair G. Computational image analysis of nonenhanced computed tomography for acute ilschaemic stroke: a systematic review. *J Stroke Cerebrovasc Dis*. 2020;29:104715. doi: 10.1016/j.jstrokecerebrovasdis.2020.104715
10. Zhu G, Jiang B, Chen H, Tong E, Xie Y, Faizy TD, Heit JJ, Zaharchuk G, Wintermark M. Artificial intelligence and stroke imaging: a west coast perspective. *Neuroimaging Clin N Am*. 2020;30:479–492. doi: 10.1016/j.nic.2020.07.001
11. Yeo M, Tahayori B, Kok HK, Maingard J, Kutaiba N, Russell J, Thijs V, Jhamb A, Chandra RV, Brooks M, et al. Review of deep learning algorithms for the automatic detection of intracranial hemorrhages on computed tomography head imaging. *J Neurointerv Surg*. 2021;13:369–378. doi: 10.1136/neurintsurg-2020-017099

TOPICAL REVIEW

12. Bivard A, Churilov L, Parsons M. Artificial intelligence for decision support in acute stroke — current roles and potential. *Nat Rev Neurol*. 2020;16:575–585. doi: 10.1038/s41582-020-0390-y
13. Na DG, Kim EY, Ryoo JW, Lee KH, Roh HG, Kim SS, Song IC, Chang KH. CT sign of brain swelling without concomitant parenchymal hypoattenuation: comparison with diffusion- and perfusion-weighted MR imaging. *Radiology*. 2005;235:992–948. doi: 10.1148/radiol.2353040571
14. Mouridsen K, Thurner P, Zaharchuk G. Artificial intelligence applications in stroke. *Stroke*. 2020;51:2573–2579. doi: 10.1161/STROKEAHA.119.027479
15. Qiu W, Kuang H, Teleg E, Ospel JM, Sohn SI, Almekhlafi M, Goyal M, Hill MD, Demchuk AM, Menon BK. Machine learning for detecting early infarction in acute stroke with non-contrast-enhanced CT. *Radiology*. 2020;294:638–644. doi: 10.1148/radiol.2020191193
16. Garcia JH, Yoshida Y, Chen H, Li Y, Zhang ZG, Lian J, Chen S, Chopp M. Progression from ischemic injury to infarct following middle cerebral artery occlusion in the rat. *Am J Pathol*. 1993;142:623–635.
17. Wu S, Mair G, Cohen G, Morris Z, von Heijne A, Bradey N, Cala L, Peeters A, Farrall AJ, Adami A, et al; IST-3 Collaborative Group. Hyperdense artery sign, symptomatic infarct swelling and effect of alteplase in acute ischaemic stroke. *Stroke Vasc Neurol*. 2021;6:238–243. doi: 10.1136/svn-2020-000569
18. The IST-3 Collaborative Group. 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*. 2015;14:485–496. doi: 10.1016/S1474-4422(15)00012-5
19. Georgakis MK, Duering M, Wardlaw JM, Dichgans M. –WMH and long-term outcomes in ischemic stroke: a systematic review and meta-analysis. *Neurology*. 2019;92:e1298–e1308. doi: 10.1212/WNL.0000000000007142
20. 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;62:1231–1238. doi: 10.1007/s00234-020-02439-3
21. Suss RA. Aspects, the mismeasure of stroke: a metrological investigation. *OSF Preprints*. 2019.
22. Editorial. Artificial intelligence: human-machine interface. *Econ*. 2019:70.
23. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res*. 2016;18:e323. doi: 10.2196/jmir.5870
24. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ*. 2020;370:m3210. doi: 10.1136/bmj.m3210
25. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ*. 2020;370:m3164. doi: 10.1136/bmj.m3164
26. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2:e200029. doi: 10.1148/ryai.2020200029
27. Weikert T, Francone M, Abbara S, Baessler B, Choi BW, Gutberlet M, Hecht EM, Loewe C, Mousseaux E, Natale L, et al. Machine learning in cardiovascular radiology: ESCR position statement on design requirements, quality assessment, current applications, opportunities, and challenges. *Eur Radiol*. 2021;31:3909–3922. doi: 10.1007/s00330-020-07417-0
28. Yusuf M, Atal I, Li J, Smith P, Ravaud P, Fergie M, Callaghan M, Selfe J. Reporting quality of studies using machine learning models for medical diagnosis: a systematic review. *BMJ Open*. 2020;10:e034568. doi: 10.1136/bmjopen-2019-034568
29. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1:e271–e297. doi: 10.1016/S2589-7500(19)30123-2
30. Kundeti SR, Vaidyanathan MK, Shivashankar B, Gorthi SP. Systematic review protocol to assess artificial intelligence diagnostic accuracy performance in detecting acute ischaemic stroke and large-vessel occlusions on CT and MR medical imaging. *BMJ Open*. 2021;11:e043665. doi: 10.1136/bmjopen-2020-043665
31. Mair G, Chappell F, Martin C, Dye D, Bath PM, Muir KW, von Kummer R, Al-Shahi Salman R, Sandercock PAG, Macleod M, et al. Real-world independent testing of e-aspects software (rites): Statistical analysis plan. *AMRC Open Res*. 2020;2.
32. Guberina N, Dietrich U, Radbruch A, Goebel J, Deuschl C, Ringelstein A, Köhrmann M, Kleinschnitz C, Forsting M, Mönninghoff C. 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
33. Bulwa Z, Dasenbrock H, Osteraas N, Cherian L, Crowley RW, Chen M. Incidence of unreliable automated computed tomography perfusion maps. *J Stroke Cerebrovasc Dis*. 2019;28:104471. doi: 10.1016/j.jstrokecerebrovasdis.2019.104471
34. Purrucker JC, Mattern N, Herweh C, Möhlenbruch M, Ringleb PA, Nagel S, Gumbinger C. Electronic alberta stroke program early CT score change and functional outcome in a drip-and-ship stroke service. *J Neurointerv Surg*. 2020;12:252–255. doi: 10.1136/neurintsurg-2019-015134
35. Potreck A, Seker F, Mutke MA, Weyland CS, Herweh C, Heiland S, Bendszus M, Möhlenbruch M. What is the impact of head movement on automated ct perfusion mismatch evaluation in acute ischemic stroke? *J Neurointerv Surg*. 2021. doi: 10.1136/neurintsurg-2021-017510
36. Neuberger U, Nagel S, Pfaff J, Ringleb PA, Herweh C, Bendszus M, Möhlenbruch MA, Kickingereder P. Impact of slice thickness on clinical utility of automated Alberta Stroke Program Early Computed Tomography Scores. *Eur Radiol*. 2020;30:3137–3145. doi: 10.1007/s00330-019-06616-8
37. Kral J, Cabal M, Kasickova L, Havelka J, Jonszta T, Volny O, Bar M. Machine learning volumetry of ischemic brain lesions on CT after thrombectomy-prospective diagnostic accuracy study in ischemic stroke patients. *Neuroradiology*. 2020;62:1239–1245. doi: 10.1007/s00234-020-02419-7
38. Yahav-Dovrat A, Saban M, Merhav G, Lankri I, Abergel E, Eran A, Tanne D, Nogueira RG, Sivan-Hoffmann R. Evaluation of artificial intelligence-powered identification of large-vessel occlusions in a comprehensive Stroke Center. *AJNR Am J Neuroradiol*. 2021;42:247–254. doi: 10.3174/ajnr.A6923
39. Ferreti LA, Leitao CA, Teixeira BCA, Lopes Neto FDN, ZÉtola VF, Lange MC. The use of e-ASPECTS in acute stroke care: validation of method performance compared to the performance of specialists. *Arq Neuropsiquiatr*. 2020;78:757–761. doi: 10.1590/0004-282X20200072
40. Nagel S, Sinha D, Day D, Reith W, Chapot R, Papanagiotou P, Warburton EA, Guyler P, Tysoe S, Fassbender K, 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;12:615–622. doi: 10.1177/1747493016681020
41. Adhya J, Li C, Eisenmenger L, Cerejo R, Tayal A, Goldberg M, Chang W. Positive predictive value and stroke workflow outcomes using automated vessel density (rapid-cta) in stroke patients: one year experience. *Neuroradiol J*. 2021;34:476–481. doi: 10.1177/19714009211012353
42. Seker F, Pfaff JAR, Mokli Y, Berberich A, Namias R, Gerry S, Nagel S, Bendszus M, Herweh C. Diagnostic accuracy of automated occlusion detection in ct angiography using e-CTA. *Int J Stroke*. 2021;17:77–82. doi: 10.1177/1747493021992592
43. Dehkharghani S, Lansberg M, Venkatsubramanian C, Cereda C, Lima F, Coelho H, Rocha F, Qureshi A, Haerian H, Mont'Alverne F, et al. High-performance automated anterior circulation CT angiographic clot detection in acute stroke: a multireader comparison. *Radiology*. 2021;298:665–670. doi: 10.1148/radiol.2021202734
44. Amukotuwa SA, Straka M, Smith H, Chandra RV, Dehkharghani S, Fischbein NJ, Bammer R. Automated detection of intracranial large vessel occlusions on computed tomography angiography: a single center experience. *Stroke*. 2019;50:2790–2798. doi: 10.1161/STROKEAHA.119.026259
45. Golan D, Shalitin O, Sudry N, J M. Ai-powered stroke triage system performance in the wild. *J Exp Stroke Transl Med*. 2020;12:01–04.
46. Heit JJ, Coelho H, Lima FO, Granja M, Aghaebrahim A, Hanel R, Kwok K, Haerian H, Cereda CW, Venkatasubramanian C, et al. Automated cerebral hemorrhage detection using RAPID. *AJNR Am J Neuroradiol*. 2021;42:273–278. doi: 10.3174/ajnr.A6926
47. Straka M, Albers GW, Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke. *J Magn Reson Imaging*. 2010;32:1024–1037. doi: 10.1002/jmri.22338
48. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, Ashrafian H, Darzi A. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med*. 2021;4:65. doi: 10.1038/s41746-021-00438-z
49. Li Y, Sperrin M, Ashcroft DM, van Staa TP. Consistency of variety of machine learning and statistical models in predicting clinical risks of individual patients: longitudinal cohort study using cardiovascular disease as exemplar. *BMJ*. 2020;371:m3919. doi: 10.1136/bmj.m3919

50. Al-Kawaz M, Primiani C, Urrutia V, Hui F. Impact of RapidAI mobile application on treatment times in patients with large vessel occlusion. *J Neurointerv Surg*. 2022;14:233–236. doi: 10.1136/neurintsurg-2021-017365
51. Morey JR, Zhang X, Yaeger KA, Fiano E, Marayati NF, Kellner CP, De Leacy RA, Doshi A, Tuhrim S, Fifi JT. Real-world experience with artificial intelligence-based triage in transferred large vessel occlusion stroke patients. *Cerebrovasc Dis*. 2021;50:450–455. doi: 10.1159/000515320
52. Domingues I, Pereira G, Martins P, Duarte H, Santos J, Abreu PH. Using deep learning techniques in medical imaging: A systematic review of applications on ct and pet. *Arti Intell Rev*. 2020;53:4093–4160.
53. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, Aviles-Rivero AI, Etmann C, McCague C, Beer L, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans. *Nat Mach Intell*. 2021;3:199–217.
54. Flanders AE, Prevedello LM, Shih G, Halabi SS, Kalpathy-Cramer J, Ball R, Mongan JT, Stein A, Kitamura FC, Lungren MP, et al; RSNA-ASNR 2019 Brain Hemorrhage CT Annotators. Construction of a machine learning dataset through collaboration: the RSNA 2019 brain CT hemorrhage challenge. *Radiol Artif Intell*. 2020;2:e190211. doi: 10.1148/ryai.2020190211
55. Prevedello LM, Halabi SS, Shih G, Wu CC, Kohli MD, Chokshi FH, Erickson BJ, Kalpathy-Cramer J, Andriole KP, Flanders AE. Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. *Radiol Artif Intell*. 2019;1:e180031. doi: 10.1148/ryai.2019180031
56. Badgeley MA, Zech JR, Oakden-Rayner L, Glicksberg BS, Liu M, Gale W, McConnell MV, Percha B, Snyder TM, Dudley JT. Deep learning predicts hip fracture using confounding patient and healthcare variables. *NPJ Digit Med*. 2019;2:31. doi: 10.1038/s41746-019-0105-1
57. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KGM, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ*. 2021;375:n2281. doi: 10.1136/bmj.n2281
58. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci USA*. 2020;117:12592–12594. doi: 10.1073/pnas.1919012117
59. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–453. doi: 10.1126/science.aax2342
60. Kaushal A, Altman R, Langlotz C. Geographic distribution of us cohorts used to train deep learning algorithms. *JAMA*. 2020;324:1212–1213. doi: 10.1001/jama.2020.12067
61. Leslie D, Mazumder A, Peppin A, Wolters MK, Hagerty A. Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? *BMJ*. 2021;372:n304. doi: 10.1136/bmj.n304
62. Overhage JM, McCallie D Jr. Physician time spent using the electronic health record during outpatient encounters: a descriptive study. *Ann Intern Med*. 2020;172:169–174. doi: 10.7326/M18-3684
63. Gilbert FJ, Smye SW, Schönlieb CB. Artificial intelligence in clinical imaging: a health system approach. *Clin Radiol*. 2020;75:3–6. doi: 10.1016/j.crad.2019.09.122
64. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol*. 2019;25:485–495. doi: 10.5152/dir.2019.19321
65. Li L, Chen Y, Bao Y, Jia X, Wang Y, Zuo T, Zhu F. Comparison of the performance between Frontier ASPECTS software and different levels of radiologists on assessing CT examinations of acute ischaemic stroke patients. *Clin Radiol*. 2020;75:358–365. doi: 10.1016/j.crad.2019.12.010
66. Montagnon E, Cerny M, Cadrin-Chênevert A, Hamilton V, Derennes T, Ilinca A, Vandenbroucke-Menu F, Turcotte S, Kadoury S, Tang A. Deep learning workflow in radiology: a primer. *Insights Imaging*. 2020;11:22. doi: 10.1186/s13244-019-0832-5
67. Baselli G, Codari M, Sardanelli F. Opening the black box of machine learning in radiology: can the proximity of annotated cases be a way? *Eur Radiol Exp*. 2020;4:30. doi: 10.1186/s41747-020-00159-0
68. Hickman SE, Baxter GC, Gilbert FJ. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. *Br J Cancer*. 2021;125:15–22. doi: 10.1038/s41416-021-01333-w
69. Gibney E. The battle for ethical AI at the world's biggest machine-learning conference. *Nature*. 2020;577:609. doi: 10.1038/d41586-020-00160-y
70. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, Sharp RR. Patient apprehensions about the use of artificial intelligence in healthcare. *NPJ Digit Med*. 2021;4:140. doi: 10.1038/s41746-021-00509-1
71. Mittelman M, Markham S, Taylor M. Patient commentary: Stop hyping artificial intelligence-patients will always need human doctors. *BMJ*. 2018;363:k4669. doi: 10.1136/bmj.k4669
72. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. *J Consum Res*. 2019;46:629–650
73. Winzeck S, Hakim A, McKinley R, Pinto JAADSR, Alves V, Silva C, Pisov M, Krivov E, Belyaev M, Monteiro M, et al. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. *Front Neurol*. 2018;9:679. doi: 10.3389/fneur.2018.00679
74. Muschelli J. Recommendations for processing head CT data. *Front Neuroinform*. 2019;13:61. doi: 10.3389/fninf.2019.00061
75. Maier-Hein L, Eisenmann M, Reinke A, Onogur S, Stankovic M, Scholz P, Arbel T, Bogunovic H, Bradley AP, Carass A, et al. Why rankings of biomedical image analysis competitions should be interpreted with care. *Nat Commun*. 2018;9:5217. doi: 10.1038/s41467-018-07619-7
76. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:195. doi: 10.1186/s12916-019-1426-2