PRACTICAL CONSIDERATIONS WHEN IMPLEMENTING AN ARTIFICIAL INTELLIGENCE PLATFORM AS AN ADJUNCT FOR ACUTE STROKE IMAGING

Paul Couto MD[1], Kunal Patel MD[1], Daniel Braga MD[1], Arijit Chakraborty DO[1], Naseem Kashani BS, Vinay Bhatia MD[1]

Mount Sinai MEDICAL CENTER Diagnostic Radiology Residency

[1]*Department of Diagnostic Radiology, Mount Sinai Medical Center, Miami Beach FL*

Correspondence: paul.couto@msmc.com

FRS Florida Radiological Society

BACKGROUND & PURPOSE

Acute stroke is one of the most emergent and time-sensitive examinations radiologists face. It is therefore logical that in recent years many stroke-focused artificial intelligence platforms have been developed, often reporting workflow advantages such as improved efficiency and accuracy. In fact, from 2018 - 2020, the number of AI-based medical imaging companies tripled, and investments doubled to $1.17 billion[1]. The list of FDA-approved imaging algorithms is expanding rapidly, and in 2018 the FDA published a fast-track approval plan for AI medical algorithms[2]. It is therefore no surprise that AI technology in recent years is increasingly becoming a significant adjunct in clinical practice. AI-based stroke analysis can assist with each aspect of the standard stroke protocol and provide additional automated decision support features, for example: ASPECTS scoring, collateral vasculature assessment, novel post-processing, and volumetric quantification of penumbra versus core infarct.

There is indeed plenty of excitement surrounding AI in imaging, especially within the realm of acute stroke. **The limitations of stroke-based AI, however, may not be as apparent on first inspection and warrant further investigation.**

OBJECTIVES

The major goals of this project are:

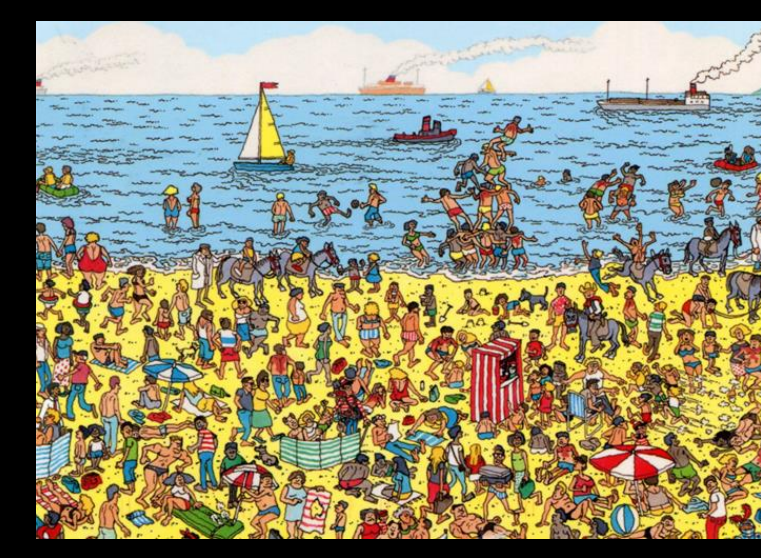
(1) To provide an educational resource for those interested in potentially incorporating AI platforms in clinical practice.

(2) To provide a fundamental understanding of Convolutional Neural Networks (CNNs) – a widely used deep learning algorithm in AI imaging.

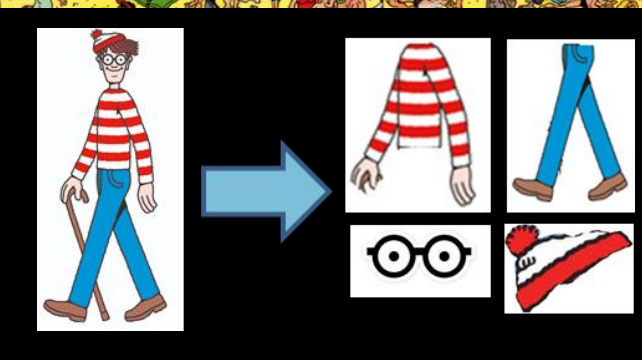
(3) To provide practical considerations regarding stroke AI applications, so their limitations are more thoroughly understood.

AN INTENTIONAL OVERSIMPLIFICATION OF CONVOLUTIONAL NEURAL NETWORKS

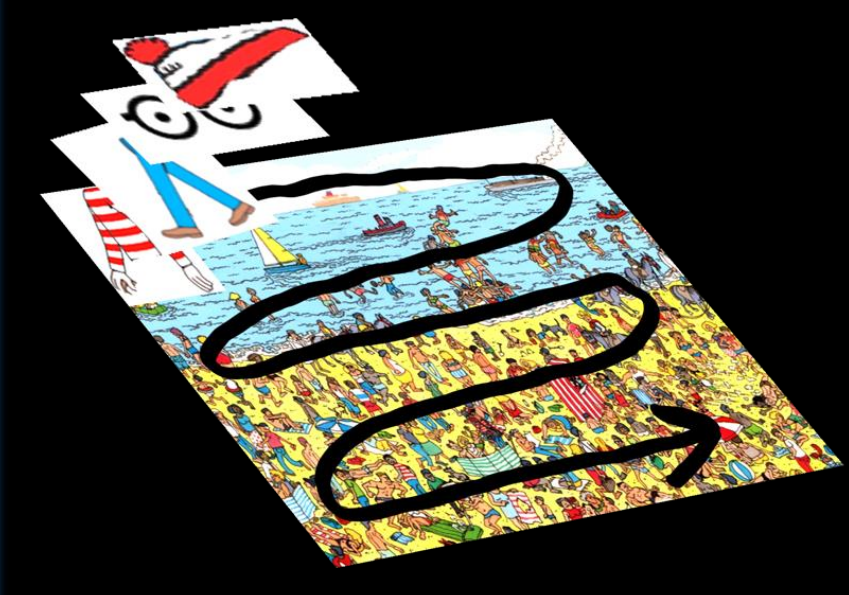
It is important to have a fundamental understanding of AI mechanisms, however **explanations are often overly complex** and can lead to more questions than answers. The majority of AI algorithms use **Convolutional Neural Networks (CNN)**. Below is an intentional oversimplification of CNNs.



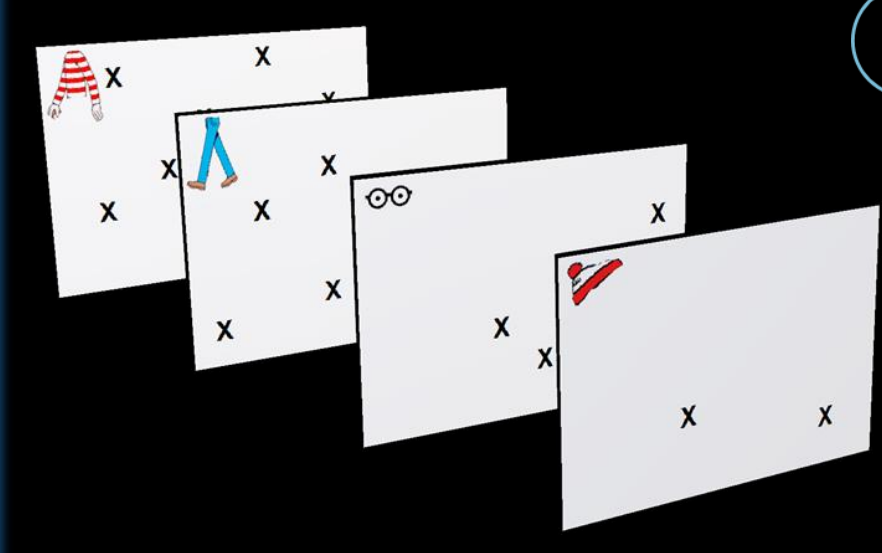
① Consider the game **Where's Waldo**. A CNN designed to find Waldo would break him into **simpler parts**.



② A CNN is organized into many different layers called **filters**. Each filter scans the image for Waldo's simpler parts (**scanning = convolution**).



③ For example, the first filter may scan for red-and-white striped items. Each time a red-and-white striped item is detected, a signal is placed onto a **feature map**. This results in a map of all the red-and-white striped items. This process is repeated for blue pants, then eyeglasses, etc. – producing the **first layer of the convolution**.



④ **In the second layer of the convolution, the CNN spatially aligns and stacks the outputs of the first convolution.** It then uses a new set of filters and looks for patterns – for instance, where red-and-white striped shirts are adjacent to blue pants and eyeglasses.

⑤ The output of the second layer would be the feature maps including adjacent clothing items. **Eventually, we would have a convolutional layer which included all the features of Waldo, thereby finding him.**

A CNN would deal with further complexity by having more layers and therefore more precision. For example, with non-con head CTs, a CNN may have many layers looking for specific features such as HUs. Remember, the output of a CNN is the presence/absence of the pathology and is always a probability.

MATERIALS & METHODS

A systematic literature review was performed analyzing published performance data on AI platforms, including: Aidoc, Canon AutoStroke, Brainomix e-Stroke, RapidAI, and Viz.ai. The analysis was guided by the CLAIM tool (Checklist for AI in Medical Imaging) for evaluating AI imaging literature and included: AI algorithm, study purpose, inclusion/exclusion criteria, study type, ground truth, performance data, sample size, and author conclusions. Publications from the past 5 years (2016-2021) were collected via search inputs in PubMed, RSNA and ASNR databases, Google Scholar, and Embase. Search criteria included platform name (e.g. 'Aidoc') and 'stroke', along with modifiers such as 'large vessel occlusion', 'intracranial hemorrhage', and 'CT perfusion'. Studies were excluded if they did not explicitly indicate a ground truth to which the AI algorithm was compared. A total of 44 publications were included. While other stroke AI platforms are available, only AI platforms with substantial available published data were included.

Scan the QR code on the bottom right to view the full literature review table, as well as links to each of the included publications.

FINDINGS

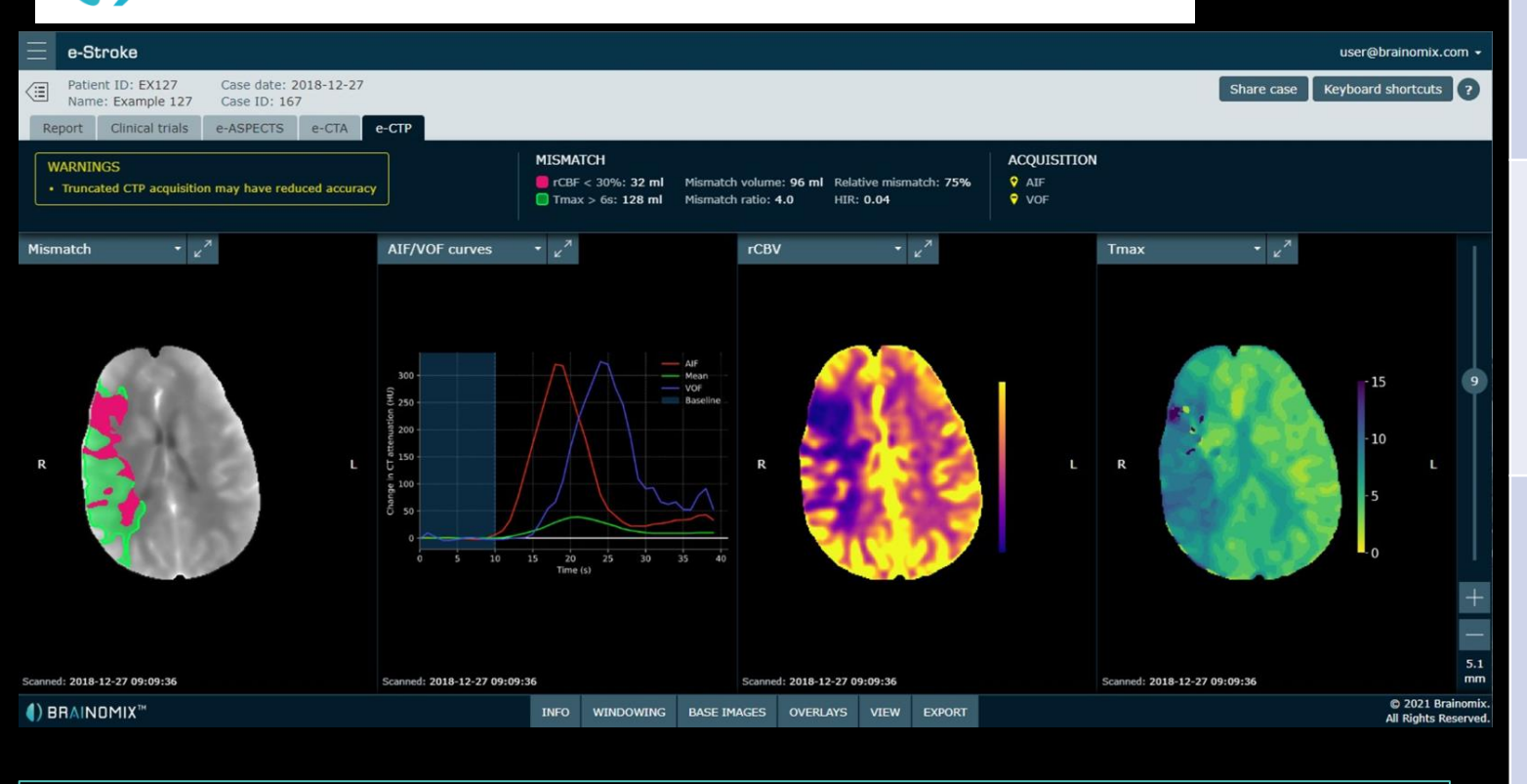
aidoc

Aidoc user interface array demonstrating (from left to right): dynamic NCCT viewer, dynamic CTA viewer, & dynamic CTP viewer. Orange dots on slider (bottom of each dynamic image) demonstrates regions where AI flagged a positive result. Images courtesy of Aidoc.

Paper	Purpose	Study Type	Inclusion Criteria	Ground Truth	n	Performance Data	Author Conclusions
Wismuller et al. 2020	Effect of automated ICH detection on turn around time (TAT)	Single center prospective RCT	Consecutive NCCTs (inpatient & ED) over 14 days	Radiologists	620	TATs for flagged cases (73 min) were lower than TATs for non-flagged (132 mins). 95% sensitivity, 96.7% specificity, 96.4% accuracy.	Automated identification of ICH reduces study TAT for ICH in emergent care settings.
Desbuquoit et al. 2019	Evaluate accuracy of AI for ICH detection on NCCT	Single center retrospective review	Consecutive NCCTs	Senior radiologist	500	Sensitivity 93%, specificity 94%, PPV 86%, NPV 97%	AI for ICH detection shows promise but reevaluation and further optimization is needed for false positives.
Khunte et al. 2020	Assess accuracy of LVO AI for presence and site of LVO.	Single center retrospective review	Patients with known M1 and/or ICA LVO by NIR angio.	Radiologists	348	Sensitivity 92.3% (LVO site), 87.6% (LVO presence). Specificity 94.9% (site), 91% (presence).	AI had high sensitivity and specificity for indicating LVO and localizing occlusion site.

Consistently seen among all LVO algorithms is the limitation of detection to only ICA-terminus to distal M1. Whether this is clinically beneficial is debatable. Additionally, to our knowledge, there are currently no available external validation publications evaluating the performance of Icometrix, the CT perfusion software utilized by Aidoc. The sole publications are internal validations and 510(k) FDA clearance forms. External validation of Icometrix is vital at this time.

BRAINOMIX™ e-Stroke

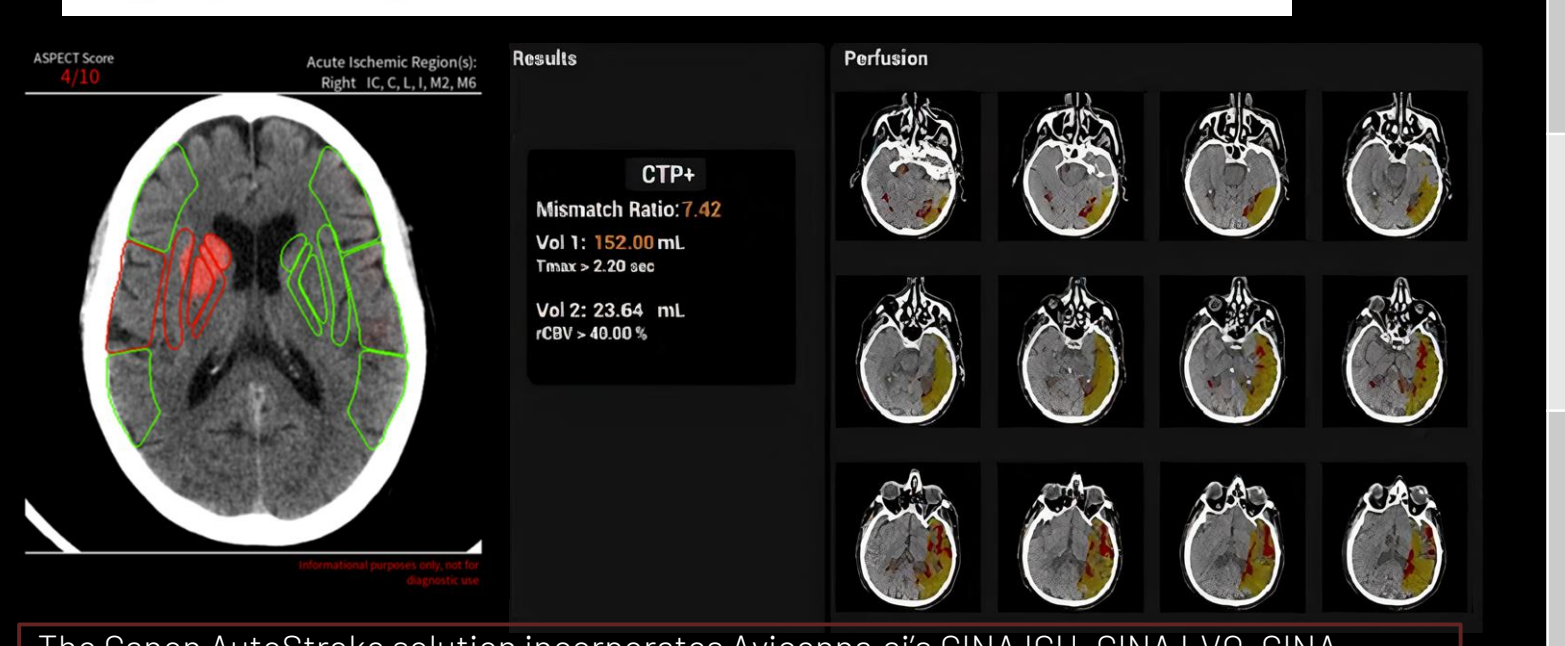


Brainomix e-CTP demonstrating (left to right): automated perfusion map with core infarct (red) and penumbra (green) with mismatch volume and ratio on top; arterial input function (AIF) and venous output function (VOF) curves, relative vertebral blood volume (rCBV) and Tmax. Images courtesy of Brainomix.

Paper	Purpose	Study Type	Inclusion Criteria	Ground Truth	n	Performance Data	Author Conclusions
Grunwald et al. 2019	AI vs neuro-radiologist CTA collateral scoring	Retrospective single center	Patients eligible for MT with LVO+ CTA	3 Neuro-rads (NRs)	98	e-CTA improved interclass correlation between NRs from 0.58 to 0.77. For collateral flow, sens. 99%, spec. 94%	Brainomix e-CTA provided real-time automated collateral scoring with potential to improve consistency of interpretation amongst NRs.
Austein et al. 2020	AI vs AI non-inferiority trial; e-CTP vs Rapid CTP	Retrospective single center	Anterior LVO patients	Rapid CTP	206	Mean difference and correlation of 5 mL and r=0.85 between AIs for core. 19 mL and r=0.77 for penumbra. Correlation with FIV similar between AI (r=0.57 and 0.56 for e-CTP and Rapid.	e-CTP and Rapid CTP generated similar core and penumbra volumes. Prediction of FIV is limited in both, which appears to be a generic limitation of CTP modality.
Pisani et al. 2021	Compare AI CTP accuracy to FIV: Rapid, Brainomix and Viz	Retrospective single center	LVO involving ICA-T or proximal M1-MCA, s/p reperfusion	FIV on follow up imaging	242	Correlation with FIV moderate in 3 CTP algorithms: Rapid r = 0.64, Viz r = 0.60, e-CTP r=0.61. Strong correlation with each other (r=0.83-0.89). FIV over-estimation (>10cc): Rapid 15%, Viz 24%, e-CTP 5%.	e-CTP, Rapid CTP, and Viz CTP strongly correlate with each other; however CTP predictability of FIV remains moderate and requires further evaluation.

Brainomix has yet to gain FDA approval, however is currently being utilized in practice in the UK. Consistently reported amongst CTP platforms is the limitation of final infarct volume prediction. While this may be a generic limitation of CT perfusion, it must be noted that the evolution of penumbra to core over time as well as clinical intervention (e.g. tPA) will skew comparison of core infarct volumes and make comparison difficult, even if follow-up imaging is performed within a reasonable time frame.

Canon AUTOStroke solution

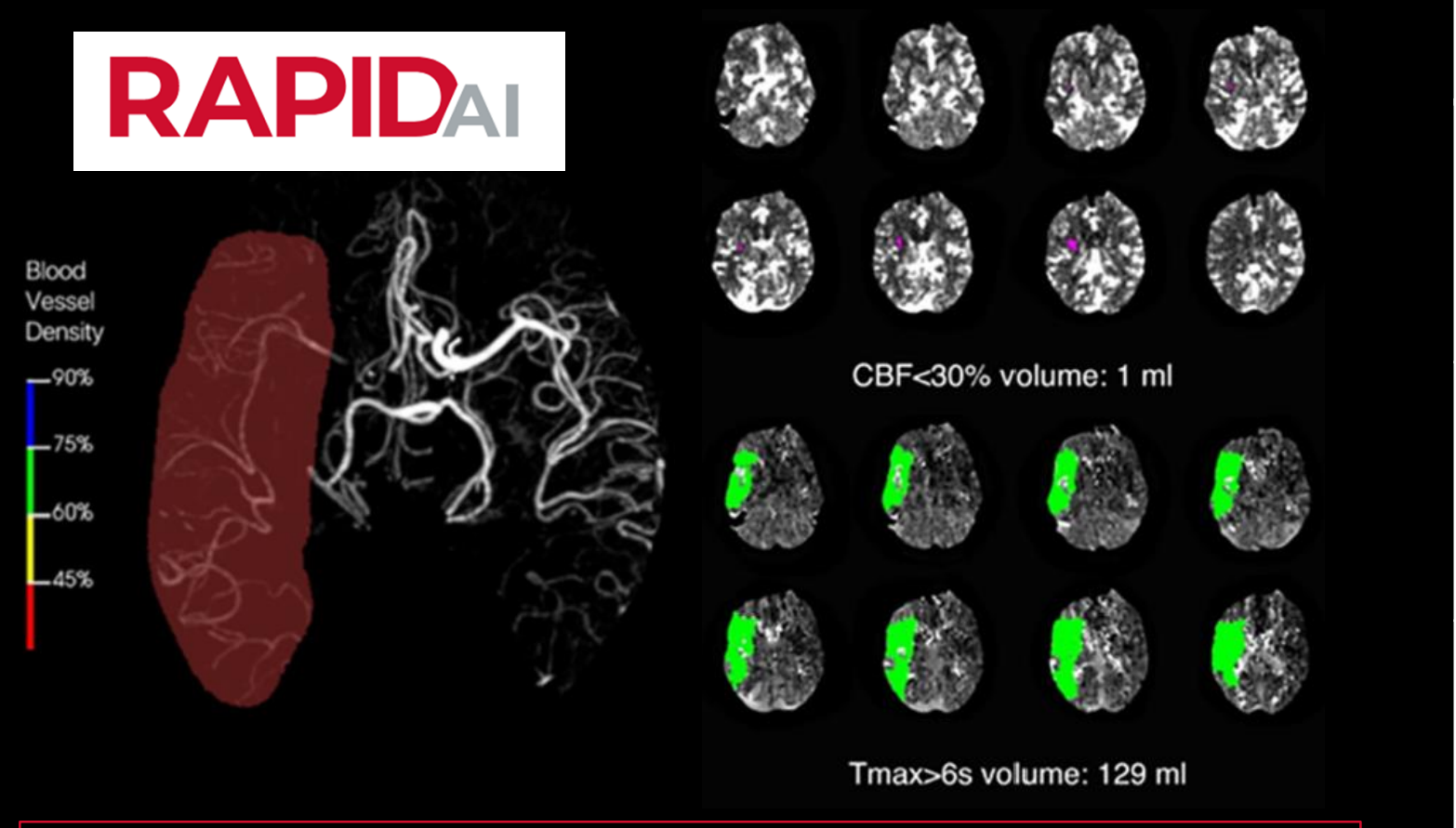


The Canon AutoStroke solution incorporates Avicenna.ai's CINA ICH, CINA LVO, CINA ASPECTS (left image showing ASPECTS regions, score and ischemic regions), and Vitrea's CT Perfusion (right, demonstrating core volume rCBV and penumbra volume Tmax). Images courtesy of Avicenna.ai.

Paper	Purpose	Study Type	Inclusion Criteria	Ground Truth	n	Performance Data	Author Conclusions
Soun et al. 2020	Assess CINA ICH for detecting and quantifying ICH	Multicenter retrospective analysis	NCCTs from 41 hospitals and 4 vendors	2 NRs	255	CINA sens. 91%, spec. 98%, AUC 0.94. For small (<5cm^3), PPV 73%. For medium (5-25cm^3) and large (>25cm^3) PPV 100%.	Limitations of missing small volume ICH, particularly with noise, motion, or streak artifacts.
McLouth et al. 2021	Assess AI generalizability for ICH detection	Multicenter retrospective analysis	ICH, LVO cases, multiple sites and vendors	2 NRs	814 NCCT 378 CTA	ICH sens. 91%, spec. 98%. For small (<5cm^3), PPV 72%. For medium (5-25cm^3) and large (>25cm^3) PPV 100%. CINA LVO sens. 94%, spec. 97%.	AI can assist with emergent findings in a variety of settings. Limitations in detection of small ICH and distal M2 occlusion.
Rava et al. 2020	Assess accuracy of Vitrea CTP in determining penumbra and final infarct volumes.	Single center retrospective analysis	Ischemic stroke patients s/p MT and LVO patients without intervention	DWI and FLAIR for FIV	105	Vitrea more accurate than Rapid CTP for FIV for patients without intervention; r=0.74 and -1.5mL mean infarct difference versus r=0.67 and -50mL. For post-intervention cohort, Vitrea and Rapid similar.	Vitrea noninferior to Rapid with similar infarct differences and correlations with FIVs for post-intervention cases. For non-intervention, Vitrea performed superior.

A commonly seen flaw with AI ICH algorithms is the propensity of high false positives for smaller hemorrhages, where the AI may mistakenly predict ICH in regions of relatively higher density, for instance, at a dural venous sinus or along the falx. Artifacts such as those listed in the chart additionally seem to be a substantial limitation of these algorithms. Despite this, AI may assist in screening outpatients for ICH where there may be a relative delay in care compared to more urgent or emergent studies from the ED.

RAPIDAI

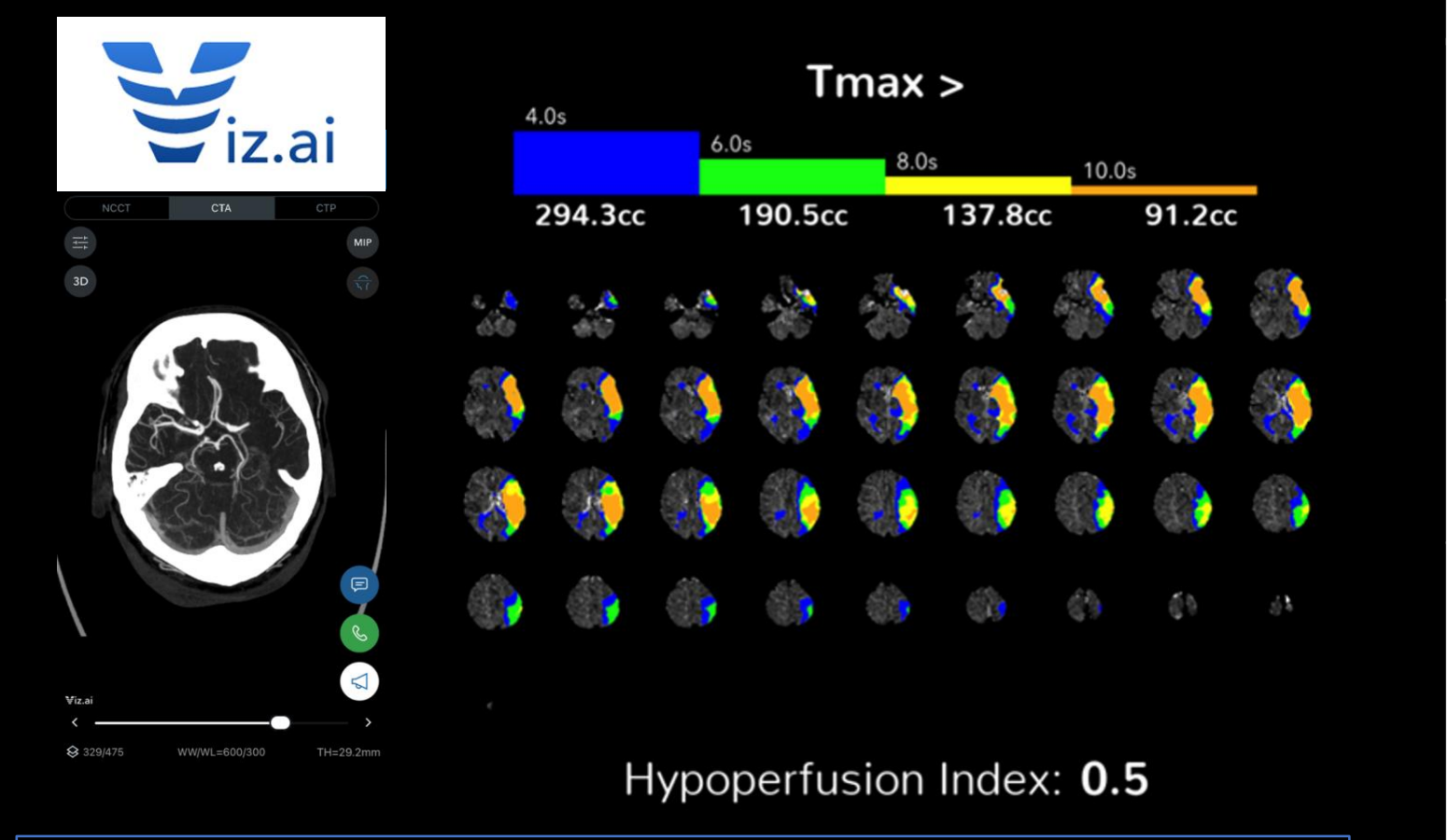


Rapid LVO (left) demonstrating CTA MIP image of patient with right M1-MCA occlusion and RAPID CTA detection of vessel paucity with a >45% reduction in blood vessel density with heat map. Rapid CTP (right) demonstrating small core infarct (top series, purple annotation) calculated by CBF<30% with 1 mL volume; and relatively large penumbra (bottom series, green annotation) calculated by Tmax >6s with 129 mL volume. Images adapted from Murray et al, 2019.

Paper	Purpose	Study Type	Inclusion Criteria	Ground Truth	n	Performance Data	Author Conclusions
Amukotuwa et al 2019	AI vs neuro-radiologist; accuracy evaluation of Rapid LVO.	Single center retrospective review	ICA/M1 or proximal M2 stroke cases	2 NRs	106	Sensitivity 95%, Specificity 79%, PPV 58% and NPV 98% for both ICA/M1 and proximal M2 LVO detection.	RAPID CTA had high sensitivity and NPV for LVO detection. It can be used to screen in emergent setting.
Dehkharghani et al 2021	AI vs neuro-radiologist; Rapid LVO.	Multicenter retrospective analysis	CTA data from CRISP + DASH trials; ICA or M1 LVO	2 NRs	217	109/217 positive for LVO. AI sensitivity 96% and specificity 98% for detection of LVO, AUC 99%.	Results confirm feasibility of fast automated high-performance detection of ICA and M1 LVO
Aghaebrahim et al. 2018	Assess door to puncture time after Rapid CTP integration	Multicenter retrospective review	Ischemic stroke pts. transferred from Rapid and non-Rapid using facilities	Non-Rapid door-to-puncture times	132	Door-to-puncture time was significantly shorter in Rapid-using facilities; median 12 mins (IQR 8-16) versus 48.5 min (32.8-71.8).	Triaging from PSC after CTP Rapid is feasible and significantly reduced door-to-puncture time without delay in transfer.

It is important to note that when comparing algorithm performances, one must be aware of what specifically the algorithm is designed to detect – for instance, Rapid LVO offers indirect LVO detection by searching for relative reductions in collateral vessel density, whereas other CNN-based LVO algorithms like Viz LVO or Brainomix e-CTA offer direct LVO detection. This makes sensitivity and specificity comparisons difficult, as one must compare apples to apples, not apples to oranges.

Viz.ai



Viz.ai LVO (left) and CT perfusion maps (right) with axial series from inferior-most to superior-most slice (top left to bottom right) with variations in hypoperfusion calculated by adjustments to Tmax and their associated annotated variations in volume. Hypoperfusion index may allow for prediction of brain regions with adequate collateral flow. Images adapted from Viz.ai training mobile app.

Paper	Purpose	Study Type	Inclusion Criteria	Ground Truth	n	Performance Data	Author Conclusions
Chatterjee et al. 2019	AI LVO detection accuracy from CTA	Single center retrospective review	Anterior circulation LVO (ICA, M1)	Radiology report	650	Sensitivity 82%, specificity 94%, PPV 77%, NPV 95%	Viz LVO demonstrated accurate identification of proximal intracranial LVOs. Need to optimize detection in poor quality CTAs.
Barreira et al 2018	CNN vs neurologist to detect LVOs from CTA	Multicenter retrospective review	Acute ischemic stroke pts.	Stroke neurologists	875	Sensitivity of 90% and specificity 83%, PPV 82%, NPV 91%; overall accuracy 86%	Convolutional neural network AI detects more than 82% of anterior circulation LVOs
Dornbos et al 2020	AI vs radiologist; accuracy of LVO AI	Single center retrospective review	All AI LVO alerts from May - Dec 2019	NRs	68	Sensitivity = 66.2%, Specificity = 91%, PPV = 45%, NPV = 96%. Majority of AI missed LVOs were distal M2 or posterior circulation occlusions (60.8%).	Viz.ai LVO is a robust screening tool for identifying LVO based on its high NPV. It can improve workflow metrics.

A consistently reported finding amongst AI algorithms with the ability to detect LVO is the relatively low positive predictive values (PPV). AI software is often unable to take certain contextual factors into account, such as artifacts (motion, bolus timing, streak artifact, etc.) or intracranial calcifications or vessel tortuosity/branching variations, which may lead to high false positive rates. This can potentially reduce radiologist efficiency by distraction by repeated case flagging or "over-searching" on a truly negative CT that has been flagged positive by AI.

<u>SUMMARY</u>: Each platform includes ICH, LVO and CTP analysis. Aidoc and Viz.AI do not include automated ASPECTS scoring. RapidAI indirectly detects LVO by assessing collateral vasculature. The definition of LVO by AI software appears to only include ICA-terminus up to distal M1-MCA, and even with this limited capacity, consistently reported LVO performance data were that of low PPVs. Automated mapping on CT perfusion mainly utilized CBF <30% and Tmax>6s, with Vitrea utilizing CBV instead. Not all CT perfusion softwares offer Tmax parameter adjustments (for instance, extending Tmax threshold to 10 seconds, which allows for hypoperfusion index calculations and the assessment of brain parenchyma with adequate collateralization and subsequent delayed perfusion.

AI Platform	ICH, LVO & CTP	Automated ASPECT Scoring	Direct LVO or Indirect Collateral	Penumbra Determination	Mobile App & Auto-Notification
aidoc	✓	✗	Direct LVO	CBF & Tmax	✓
BRAINOMIX™	✓	✓	Direct LVO	CBF & Tmax	✓
Canon	✓	✓	Direct LVO	CBV & Tmax	✓
RAPIDAI	✓	✓	Indirect Collateral	CBF & Tmax	✓
Viz.ai	✓	✗	Direct LVO	CBF & Tmax	✓

DISCUSSION

It is widely known that artificial intelligence is demonstrating increasing significance in medical imaging, especially in regard to the evaluation and interpretation of cases of acute hemorrhage and ischemic stroke. What may not be as apparent, however, are the limitations of applying these automated algorithms into clinical practice. We presented a series of literature reviews and analyses breaking down major AI platforms into their features and provided general limitations of AI as a whole. Potential barriers of AI implementation in clinical practice can be divided into four main categories: Performance overestimation and misinterpretation, ethical and liability implications, technical difficulties and monetary expense.

Performance Overestimation and Misinterpretation

- Limitation of LVO detection to ICA terminus and M1-MCA
- Vendors do not always report low PPVs, which can potentially reduce efficiency by distraction from repeated case flagging or leading to "over searching" on a negative CT that has been flagged positive by AI
- Vendor datasets are created in a homogenous, controlled environment which is not always generalizable to real life and may lead to AI being susceptible to technical artifacts or protocol variability across institutions.
- Full publication of internal validation study is not required to be made public to obtain FDA 510(k) clearance.

Ethical and Liability Implications

- More than one disease or abnormality is often present in the same patient – AI may lead to the "satisfaction of search" phenomenon in radiologists.
- Patient maleficence may arise as a result of AI error that would have been read appropriately by a human interpreter – for instance, automated CT perfusion map overestimating penumbra/underestimating core infarct leading to unwarranted endovascular intervention.
- Cloud-based or off-site servers may be subjected to cyber attacks and patient information may be compromised.

Technical Difficulties

- Issues of workflow integration may arise, especially in smaller institutions without a robust IT department.
- May require more powerful GPUs which are more expensive and less readily available than regular CPU workstations.

Significant Monetary Expense

- Substantial upfront cost without a guaranteed return on investment.
- NTAP (New Technology Add-on Payment) often publicized by AI companies) requires annual renewal, and only applies to Medicare patients in cases where hospital losses are incurred.

> *"…It turns out [when] you take that same AI system to an older hospital down the street, with an older machine, and the technician uses a slightly different imaging protocol, that data drifts to cause the performance of the AI system to degrade significantly. In contrast, any human radiologist can walk down the street to the older hospital and do just fine."*
>
> Andrew Ng
> Founder and CEO of Landing AI, deeplearning.ai
> Adjunct professor at Stanford University

CONCLUSION

We hope this exhibit will educate viewers on the currently available stroke-based AI platforms and the importance of developing a greater understanding of their algorithms while not only recognizing their potential, but also remaining cognizant of their limitations. While automated analysis of stroke imaging may be useful as an adjunct, both with regards to image interpretation and with patient triage, continued work is nevertheless required in this field to better improve the practicality of AI in acute stroke management in clinical practice.

REFERENCES & ACKNOWLEDGEMENTS

We thank Aidoc, Brainomix, Avicenna.ai, RapidAI and Viz.ai for their support with the generation of this project.

[1] Alexander A, Jiang A, Ferreira C, and Zurkiya D. An intelligent future for medical imaging: A market outlook on Artificial intelligence for medical imaging. JACR 2019.

[2] Topol E. High performance medicine: the convergence of human and artificial intelligence. Nature Medicine 2019.

Please scan the QR code on the left for additional references and complete literature review table.