AI-powered stroke triage system performance in the wild

Ofir Shalitin, Neta Sudry, Jonathan Mates, David Golan*

a. David Golan Ofir Shalitin, Jonathan Mates , Viz.ai

b. Neta Sudry, Tel Aviv university

*Author for correspondence:

David Golan, Viz.ai
E-mail: david@viz.ai

Abstract:

Background: In February of 2018 the Food and Drug Administration (FDA) cleared Viz ContaCT, known commercially as Viz LVO- the first clearance for artificial-intelligence based software that detects large vessel occlusion (LVO) stroke and directly alerts relevant specialists, for the purpose of triage. The potential benefit is to reduce time to notification and therefore treatment, with better outcomes for patients. Recent publications have demonstrated time savings, improved patient outcomes, and reduction in hospital length-of-stay following an implementation of Viz LVO. Clinical evidence to support FDA clearance is typically generated in controlled settings, so it is important to characterize the performance of such software in the real-world, in which the data is more heterogeneous and unpredictable, coming from a number of sites with varying equipment, protocols and personnel skill levels..

Methods: 2,544 patients from 139 hospitals analyzed using the commercially available Viz LVO were sequentially collected and annotated. The results were used to evaluate the sensitivity and specificity of the Viz LVO software, as well as the time-to-notification of an LVO alert to the stroke team.

Results: Viz LVO demonstrated high sensitivity (96.3% [92.6%-98.8%]) and specificity (93.8% [92.8%, 94.7%]) and a median time-to-notification of 5 minutes and 45 seconds across all of the sites involved.

Conclusions: Viz LVO demonstrates robust performance despite the heterogeneity of setting, equipment, and processes. This suggests that the benefit documented at specific sites or systems may generalize to other hospital systems.

Keywords: Acute ischemic stroke, ischemic attack, stroke rehabilitation

Introduction

Stroke is the number one cause of long term disability in the United States affecting nearly 800,000 people a year who are left with varying degrees of disability [1]. There are effective treatments for a subset of patients - such as endovascular thrombectomy for patients with a large vessel occlusion (LVO) - but the efficacy of these treatments is time-sensitive. The medical cost of delay has been estimated, with every additional delayed minute resulting in 1.9M dead neurons [2] and the loss of 4.2 days of healthy life [3]. Despite the time-sensitivity and the harmful impact of delays on patients, the median time to treatment in the STRATIS registry of 984 patients was 3 hours in a comprehensive stroke center (CSC) and 5 hours in a primary stroke center (PSC) [4].

In a new ruling in February of 2018 [5], the United States Food and Drug Administration (FDA) approved Viz LVO   and thus created a new category of medical devices: computer assisted triage (CADt). Viz LVO is an artificial-intelligence powered software application which automatically analyzes computed tomography angiography (CTA) of the brain of patients suspected of stroke, identifying large vessel occlusions (LVO) and flagging them directly to a specialist, who can make a treatment decision or guide the team to further investigate the differential diagnosis. Screenshots of

the device are provided in [Figure 1]. This device creates a parallel workflow that has the potential to significantly accelerate diagnosis and treatment, as outlined in [Figure 2]. The supporting evidence for the FDA submission included real-world data demonstrating a significant time savings, alerting the specialist on average 52 minutes faster than the standard of care.
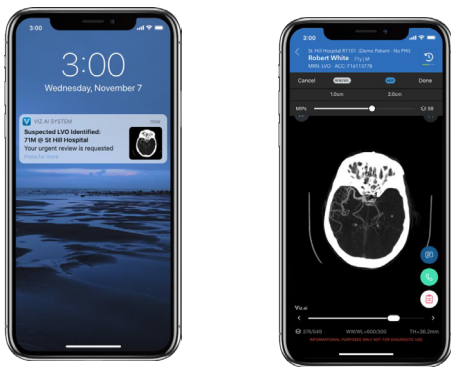

Figure 1: The Viz LVO device offers automatic stroke detection and notification (left), coupled with a mobile DICOM viewer (right).
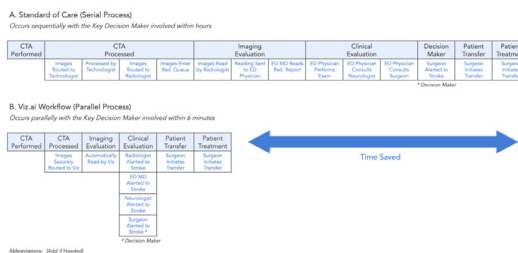

Figure 2: Accelerating notification of a specialist via an automated parallel workflow

Scans enter the radiologist's queue and are processed by the standard-of-care (black boxes) while also being automatically processed by the automated triage software.

In the time since its clearance, several publications demonstrated time-to-notification savings when using Viz ContaCT ranging from 33 to 66 minutes, as well as reduction in length of stay, and improvement of patient outcomes at discharge and at 5-day and 90-day follow up[6, 7].

While FDA clearances of computer-aided triage devices typically include supporting information regarding the sensitivity and specificity of the cleared device as well as average time-savings, these data-points need to be validated in a real world setting for several reasons::

- **Quality bias:** sites participating in pivotal studies supporting

FDA submissions are typically at the cutting edge of the field. As such, their programs are typically well staffed, well trained and well funded, and thus the average quality of a scan at a site participating in a pivotal study may not represent the average quality of a scan in an average hospital in the US. Thus these data may be biased to have higher quality imaging due to more modern or advanced equipment. Greater experience of the CT technologists and staff may lead to fewer technical issues or patient-driven issues, such as contrast bolus timing and patient motion, respectively.

- **Overfitting due to lack of diversity in data:** Overfitting[8] is a general problem in modelling whereby a model may have good predictive capacities on a training dataset, but lower performance on real-world data. While there are many reasons for this phenomenon, the most relevant in the case of applications of AI in medical imaging is when the distribution of the training set data is different from the distribution of the real-world test data. For example, a device developed using data from only a subset of scanner makes, models and configurations might obtain good performance on these data, but yield lower performance when applied on more diverse data. Since most CADt devices use deep-learning models as a key building block of their detection algorithm, they are particularly susceptible to such issues. When a device is developed using a dataset with limited diversity, there is risk that performance on data of greater diversity will be lower.

The primary focus of this study is to evaluate the performance of the Viz LVO device in terms of sensitivity, specificity and time-to-notification on a diverse dataset from different hospitals of varying sizes and capabilities, to obtain a robust estimate of performance that aims to avoid the aforementioned pitfalls. To the best of our knowledge, this is the first study aimed at assessing the performance of an AI triage system for stroke with such a diverse data set from multiple hospital systems.

Methods

Over 400 CTA scans from different hospital systems are analyzed by the Viz ContaCT device on a daily basis. As part of the on-boarding process of new institutions, all sequential scans within a defined date range are reviewed and annotated, with ground truth established by a team of radiology trained annotators.

The data used in this study included 2,544 patients scans from 139 sites, belonging to 37 different hospital systems across the US, with diversity of patient sex, patient age, and technical parameters

such as CT manufacturer. Hospital type also varied, with both Primary Stroke Centers (PSC) and Comprehensive Stroke Centers (CSC) (see Table 1).

Table 1: Summary statistics of the data used in this study.

Patient demographics		
	Age	Median = 66 years (SD=17.42)
	Sex	Women = 52.48%
		Men = 46.62%
		Unspecified = 0.9%
	LVO	6.41% (163/2544)
Site properties		
	# patients	Average = 67.38 (SD=68.61)
		Median = 47
	PSC vs. CSC	PSC = 73 sites
		CSC = 34 sites
		Other or unspecified = 32 sites
Perfusion imaging		
	Out of total scans	31.05% of scans (790/2544)
	Sites with CTP available	45.32% (63/139)
CT manufacturer		
	GE	47.17%
	Siemens	26.49%
	Philips	4.72%
	Toshiba	21.34%
	other/NA	0.28%

Statistical analysis

Performance of the algorithm was measured by sensitivity and specificity, with medical annotation team interpretations serving as the ground truth. In addition, workflow performance was measured by time-to-notification, defined as CTA acquisition start time to alert send time.

Results

Data were collected from a total of 139 sites managed under 37 hospital systems. Of the 139 sites, 34 areCSCs, while 73 are PSCs. Only 63 of the 139 sites perform CT perfusion scans, indicating a diversity of levels of experience: from established comprehensive stroke centers to smaller stand alone emergency departments. The sample covers all major CT manufacturers, with GE scanners being the most common. Descriptive statistics are provided in [Table 1].

A total of 163 ground-truth LVOs were identified. The overall sensitivity and specificity were 96.32% [0.9268, 0.9884] and 93.83% [0.9283, 0.9475], respectively. The median time to notification of the software was 5 minutes and 45 seconds, discounting 7 historical cases which were sent to the device manually for testing. 4.14% of scans were rejected by the device

due technical reasons such as metal artifacts in close proximity to the Circle of Willis, and were therefore excluded from the analysis. Summary of performance statistics is provided in [Table 2]. Examples of device outputs are provided in [Figure 3].

Table 2: Performance statistics of the Viz ContaCT device.

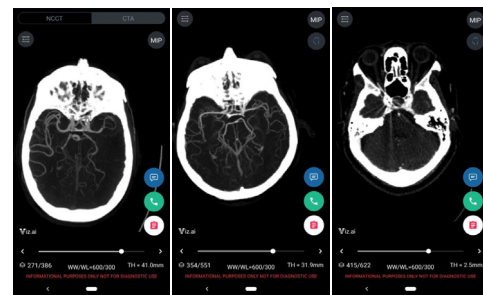| Device performance		
	Sensitivity	96.32% [0.9268, 0.9884]
	Specificity	93.83% [0.9283, 0.9475]
	Time-to-notification	Median = 5 minutes and 45 seconds

Figure 3: left: MIP of a scan presenting with a left M1 occlusion that was flagged by Viz LVO. Middle: MIP of a normal scan that was not flagged by Viz ContaCT. Right: Scan with a metal artifact that was rejected by the device.

Discussion

Automatic systems for detection and triage of stroke have a potential to significantly accelerate stroke triage, diagnosis and treatment, reduce length of stay and improve patient outcomes. Our work extends recent work in the field focused on measuring the effect of installing Viz Contact at specific systems by performing a wider analysis demonstrating the performance of the device is robust across different settings. This allows the hypothesis that the downstream benefits, demonstrated by the works of Hassan et al. and Morey et al. are likely to extend to other healthcare systems.

While random control trials are the gold standard in demonstrating the effectiveness of new therapies or methodologies, they often reduce bias and confounding evidence by limiting participation, with strict inclusion and exclusion criteria. The cost of this internal validity may be a reduction in generalizability of the results to the less controlled everyday practice [9]. The use of real world evidence, such as that presented here, may help payors, clinicians, and administrators decide if there is sufficient evidence that the proposed solutions would work in their specific situations. Moreover, demonstration of effectiveness of workflow enhancements - such as earlier alerting of specialists - does not necessarily guarantee transferability of that enhancement to a different system without understanding the specific circumstances of that system. If there is

limited capacity to perform thrombectomy, for example, or if there are other idiosyncratic rate-limiting factors which would preclude reducing the speed to treatment for stroke patients, those factors would have to be taken into account. The value of such evidence is to demonstrate the realistic possibility. The question that should be asked by the participants and owners of the stroke workflow in any system is whether, given the sensitivity/specificity of a detection system such as that presented here and the ability to alert specialists within several minutes of CTA acquisition would enable workflow enhancements that would improve time-to-treatment for stroke patients.

Limitations of study

This investigation focuses on the performance of the device: sensitivity, specificity and time-to-notification.. High sensitivity, specificity and timely alerts enable prompt review by stroke specialists.

Conclusion

In conclusion, Viz ContaCT presents high sensitivity and specificity even on data collected from 139 hospitals, of 2,544 patients, more than half of which were scanned at non specialist centers that may be more likely to have older scanners and less expertise in scanning acute stroke patients. We thus speculate that AI-powered detection and triage of stroke, coupled with a mobile platform for fast and easy communication and collaboration may be an essential tool for stroke teams.

Conflict of Interest

We hereby confirm that there is no conflict of interest associated with publication.

Acknowledgement

Recent publications have demonstrated time savings, improved patient outcomes, and reduction in hospital length-of-stay following an implementation of Viz LVO.

Author's Contributions

David Golan and Ofir Shalitin designed the experiment, Ofir Shalitin and Neta Sudry collected and analyzed the data. Jonathan Mates provided clinical oversight and guidance. David Golan, Ofir Shalitin and Jonathan Mates wrote the manuscript.
All authors read and approved the final manuscript.

References

1. Saver JL. Time is brain--quantified. Stroke. 2006;37(1):263-266.

2. Meretoja A, Keshtkaran M, Tatlisumak T, Donnan GA, Churilov L. Endovascular therapy for ischemic stroke: Save a minute-save a week. Neurology. 2017;88(22):2123-2127.

3. Froehler MT, Saver JL, Zaidat OO, et al. interhospital transfer before thrombectomy 24 is associated with delayed treatment and worse outcome in the stratis registry (systematic evaluation of patients treated with neurothrombectomy devices for acute ischemic stroke). Circulation 2017;136:2311–21.

4. Evaluation of automatic class iii designation for contact.

5. Current commercial name: Viz LVO. Clearance number: DEN170073.

6. Impact of Viz LVO on Time-to-Treatment and Clinical Outcomes in Large Vessel Occlusion Stroke Patients Presenting to Primary Stroke Centers Jacob R Morey, Emily Fiano, Kurt A Yaeger, Xiangnan Zhang, Johanna T Fifi medRxiv 2020.

7. Hassan AE, Ringheanu VM, Rabah RR, Preston L, Tekle WG, Qureshi AI. (2020). Early experience utilizing artificial intelligence shows significant reduction in transfer times and length of stay in a hub and spoke model. Interventional Neuroradiology.

8. Hawkins, Douglas M. "The problem of overfitting." Journal of chemical information and computer sciences. 2004; 44(1): 1-12.

9. Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and Impact of Real-World Clinical Data for the Practicing Clinician. Adv Ther. 2018;35(11):1763–74.