

Real-World Performance of AI-Powered ICH Triage System

Khen Sela Peremen^{1*}, Naama Avni¹, Lilian Atlan¹, Dor Amran¹

¹150 Menachem Begin Road, We Tower, 19th floor, 6492105, Tel Aviv, Israel

***Corresponding Author:** Khen Sela Peremen, 150 Menachem Begin Road, We Tower, 19th floor, 6492105, Tel Aviv, Israel.

Received Date: 14 August 2023; **Accepted Date:** 25 January 2024; **Published date:** 14 March 2024

Citation: Khen Sela Peremen, Naama Avni, Lilian Atlan, Dor Amran. (2024). Real-World Performance of AI-Powered ICH Triage System. *Journal of Hematology and Disorders*. 3(1); DOI: 10.58489/2836-3582/10

Copyright: © 2024 Khen Sela Peremen, this is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Intracranial hemorrhage (ICH) is a life-threatening condition with high mortality and disability rates.

In March 2018, the FDA cleared Viz ICH [1], an AI tool that analyzes brain NCCT studies for the presence of an ICH. Upon detection of a suspected ICH, Viz.ai sends a notification to alert a multidisciplinary team to promote an efficient workflow, leading to quick diagnosis and treatment.

The objective of this study is to characterize the performance of Viz ICH in real-world settings. The FDA clearance is based on data generated in controlled settings; however, an effective AI tool should prove its potency and efficacy with heterogeneous data, originating from sites with varying equipment, protocols, and personnel skills.

3139 patients from 127 hospitals were analyzed using Viz ICH. Results were used to evaluate the detection of ICH, measured by accuracy, sensitivity, specificity, PPV, NPV, and time to notification.

On a real-life consecutive set of 3156 scans, Viz ICH showed a high accuracy rate of 98.57%, and sensitivity, specificity, PPV and NPV of 90.48%, 98.97%, 81.1% and 99.53%, respectively. Mean time to notification was 3.39 minutes for ICH suspected cases.

Viz ICH demonstrates robust performance in heterogeneous, real-world settings. Furthermore, the swift time to notification can assist clinicians in speeding up diagnoses. Viz ICH can therefore be considered a reliable AI tool for medical centers of various sizes and capabilities for the goal of improving patient outcomes.

Keywords: Intracranial hemorrhage, Stroke, Artificial intelligence, Non-contrast CT scan, Brain bleed

Abbreviations:

ICH – Intracranial hemorrhage; AI- Artificial intelligence; FDA - Food and Drug Administration; CT – Computerized tomography; NCCT - Non-contrast CT; PPV – Positive predictive value; NPV – Negative predictive value; TBI – Traumatic brain injury; IPH – Intraparenchymal cerebral hemorrhage; IVH – Intraventricular hemorrhage; SAH – Subarachnoid hemorrhage; SDH – Subdural hemorrhage; EDH – Epidural hemorrhage; PSC – Primary stroke centers; CSC – Comprehensive stroke centers; ROC - Receiver operating characteristics; AUC - Area Under the Curve; DICOM - Digital Imaging and Communications in Medicine

Introduction

ICH is a medical emergency, resulting in a mortality rate of 35% to 52% at 30-days. Survivors' functional outcome is often very poor, with fewer than 20% being independent at six months [2].

ICH is typically divided into spontaneous (non-traumatic) ICH and traumatic ICH. Spontaneous ICH accounts for 9%-27% of all strokes [3], but the global burden of brain hemorrhage is greater than that of

ischemic stroke in terms of death and disability, even though the incidence of ischemic stroke is twice as great [4].

ICH refers to the accumulation of blood outside blood vessels within various spaces in the brain. Common categorization is based on the hemorrhage location: Intraparenchymal cerebral hemorrhage (IPH), Intraventricular hemorrhage (IVH), Subarachnoid hemorrhage (SAH), Subdural hemorrhage (SDH),

Hematology and Disorders

and Epidural hemorrhage (EDH). Traumatic ICH is a common result of traumatic brain injury (TBI) [5], and is the most common cause of death in lethally injured trauma patients accounting for up to 50% of fatalities [6], resulting in a significant amount of long-term disability [7].

ICH is a dynamically evolving process [8], in which response time is critical. Disease outcome is primarily affected by hematoma growth, perihematomal edema, and evolving intracranial pressure, that may necessitate immediate neurosurgical intervention, which may be life-saving [9]. Early detection and management of ICH have been proven to be associated with improved clinical outcomes [10,11].

Importantly, also in the setting of an acute ischemic stroke, it is essential to rule-out ICH as quickly as possible to enable optimal treatment to restore blood flow before irreversible brain damage occurs [12].

Both acute ischemic stroke and ICH present similar clinical features, but whereas treatment with thrombolytics is an essential part of ischemic stroke management [13], it can be fatal when given to a patient with ICH. The detection of ICH in a time-efficient, and reliable manner is fundamental and crucial in both settings.

The use of CT exams, since the 1970s, has changed the way we manage and diagnose patients, especially in the acute setting. The use of CT improved the treatment of many acute medical conditions, including stroke [14]. According to Power et al. (2016) CT has many advantages, such as availability of scanners and fast results, allowing physicians to confirm diagnoses and exclude critical conditions.

Over the past decades, imaging and thereby radiologists' on-call workload in the ambulatory and ER setting has shown tremendous growth. The increase in imaging studies over the past years represents a big challenge for radiologists and medical centers around the world and necessitates adjustments in the workforce and workflow to accommodate these evolving changes while guaranteeing the quality and safety of patients and avoiding radiologist's burnout [15,16].

In the initial clearance in March 2018 [1], the United States Food and Drug Administration (FDA) approved Viz ICH as a notification-only, parallel workflow tool for use by hospital networks and trained clinicians to identify and communicate images of specific patients to a multidisciplinary team, independent of standard of care workflow. Viz ICH uses an AI based algorithm to analyze head NCCT

images for findings suggestive of ICH and notifies appropriate medical specialists of these findings in parallel to standard of care imaging interpretation, as shown in figure 1. When ICH is suspected, the system sends a notification to a multidisciplinary team, recommending review of those images. Images can be viewed through a mobile application (figure 2) or via the Viz web viewer (figure 3).

The primary focus of this study is to evaluate the performance of the Viz ICH device in terms of accuracy, sensitivity, specificity, PPV, NPV, and time-to-notification on a diverse dataset from different hospitals of varying sizes and capabilities to obtain a robust estimate of performance that represents real-world settings.

Methods

Data Collection

The data consists of 4 subsets, each one collected sequentially from multiple hospitals during the time periods shown in Table 1. Data was collected and reviewed on a quarterly basis as a post-market real world performance monitoring quality assurance mechanism of the Viz ICH medical device. Selected hospitals and collection dates were chosen at random to enable variety and examine device generalization. Collection time periods were selected to represent the full range of intended population scanned throughout a day in both week/end days.

Scans were excluded from the set-in cases of motion artifacts, metal artifacts and low quality that limit scan readability. Postoperative scans were also excluded due to inadequacy for the ICH screening task.

The collected cohort demonstrates diversity of patient gender, age, scanning technical parameters (such as CT manufacturer and model, scan resolution and protocols). The hospital type also varied, with both Primary Stroke Centers (PSC) and Comprehensive Stroke Centers (CSC) (see Table 2).

Truthing

The CT scans were interpreted by radiology trained clinical specialists and a neuroradiology fellow. Every scan was evaluated for the existence of ICH and the bleed type.

Time to Notification Measurement

To assess the time saved, the time-to-notification was computed. The time-to-notification per scan relates to the time elapsed between the CT acquisition start time and the time in which the notification was sent to the user.

Statistical Analysis

Statistical analysis was performed using descriptive data, including ranges, means, medians, standard deviations for continuous variables, and frequencies and percentages for categorical variables. System performance methods were examined using sensitivity, specificity, PPV and NPV.

Confidence intervals were calculated with the Clopper-Pearson interval based on Beta distribution. Receiver Operating Characteristics (ROC) analysis was used to determine Viz ICH classification performance over multiple suspected ICH size volumetric thresholds.

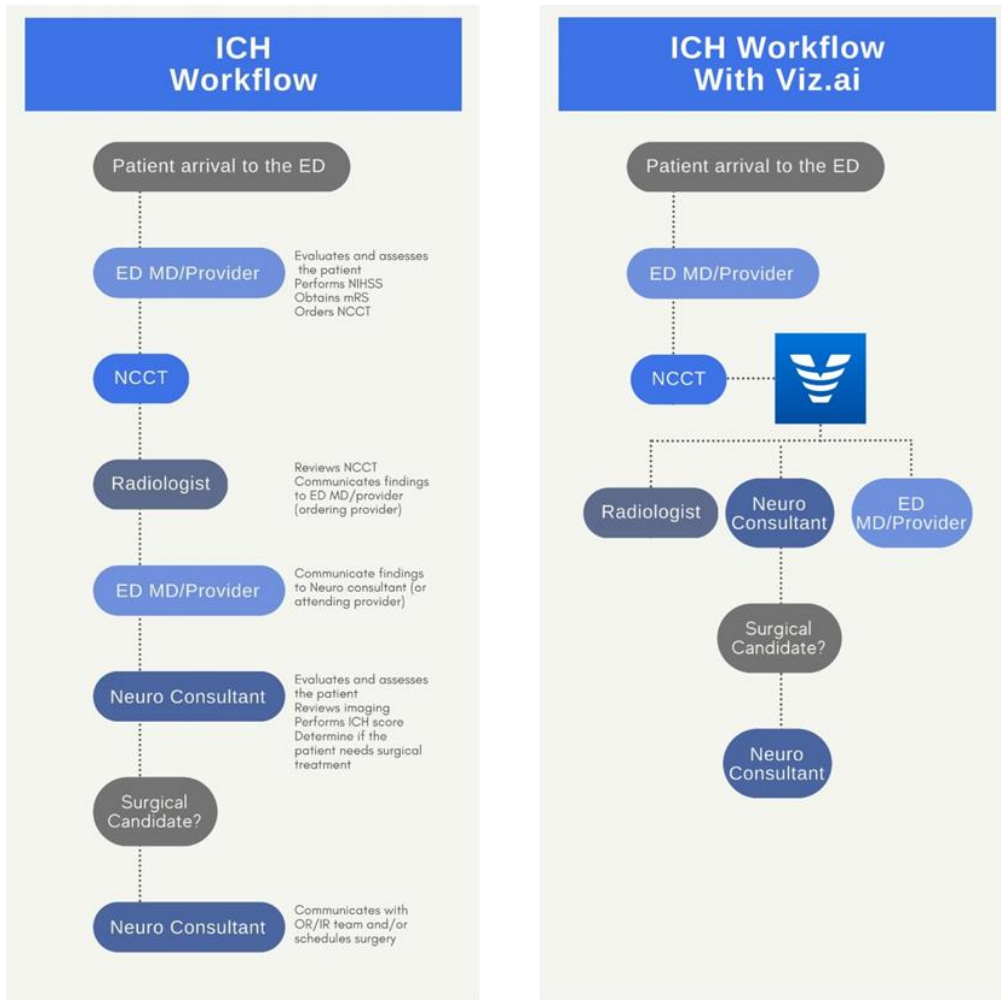


Fig 1: Management workflow for ICH, left: standard ICH workflow, right: ICH workflow with Viz ICH

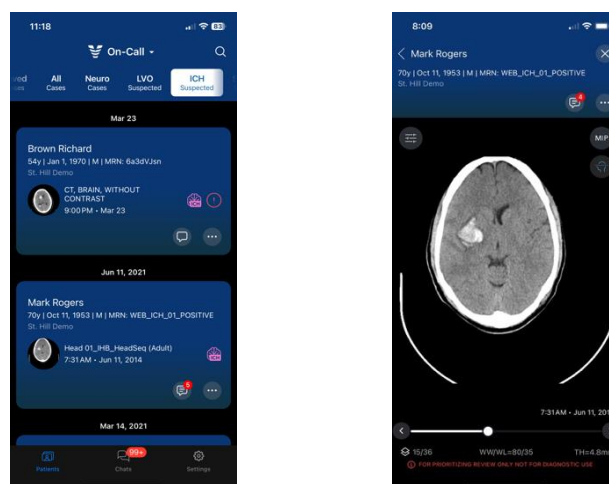


Fig 2: Viz ICH's mobile DICOM viewer, examples for ICH suspected scans

Hematology and Disorders

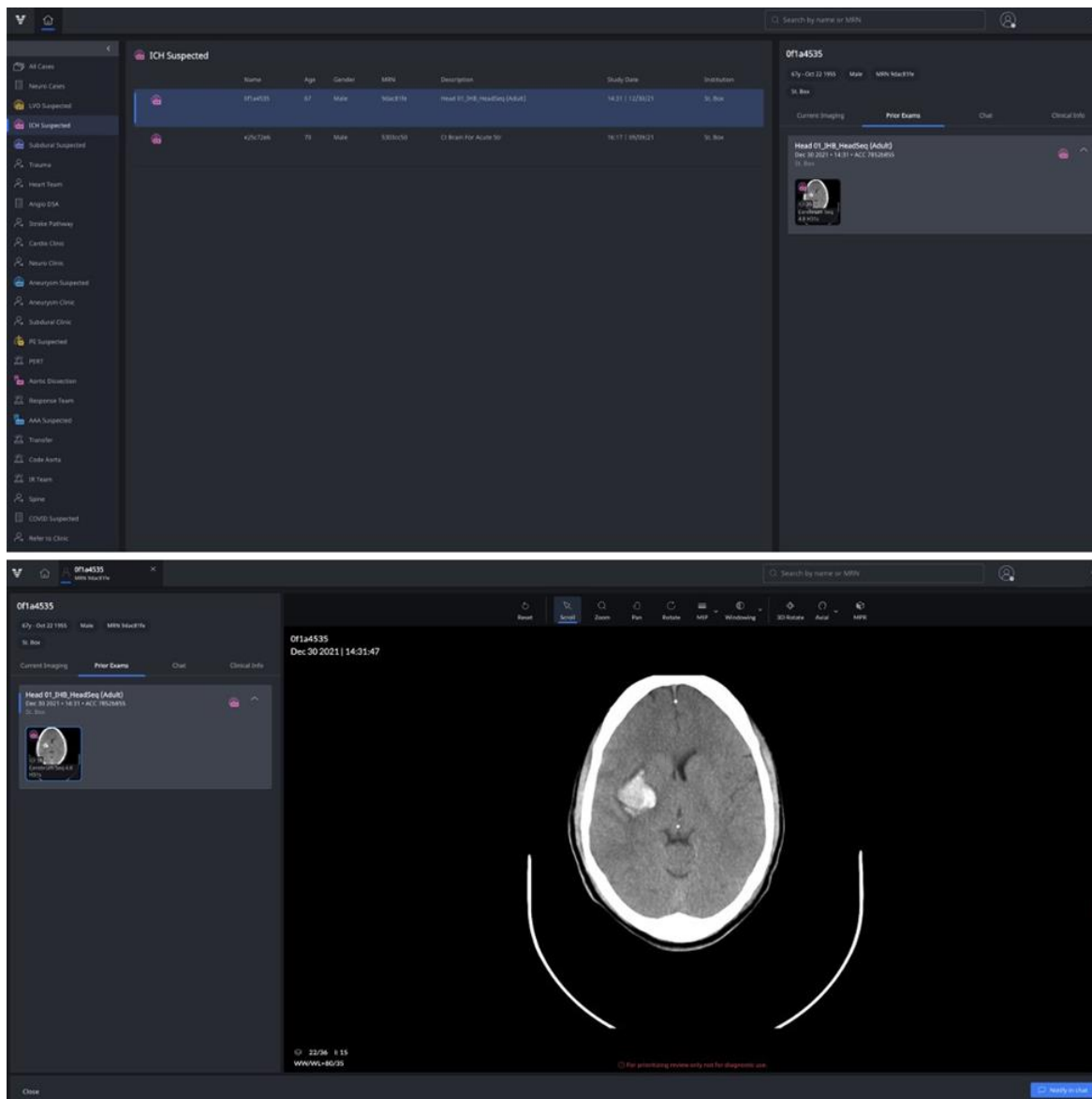


Fig 3. Viz Web ICH suspected subjects list (TOP), single subject scan view (bottom)

Table 1: Study consecutively collected subsets, time period, and hospital distribution. Notice that hospitals were used for data collection in multiple time periods as a part of timely, multi-site, consecutive performance evaluation on the Viz ICH device for quality control purposes.

| Subset # | Time Period | Number of hospitals | |
|----------|--|--|---|
| | | Number of unique hospitals (appear on specified subset only) | Number of repeat hospitals (Appear in more than one subset) |
| 1 | 10/23/2021 00:00 - 10/24/2021 23:59 | 0 | 8 |
| 2 | 11/07/2021 00:00 - 11/14/2021 23:59 | 26 | 48 |
| 3 | 02/01/2022 00:00 - 02/03/2022 23:59 | 3 | 18 |
| 4 | 05/01/2022 - 00:00-08:00 06/01/2022 - 08:00-12:00 07/01/2022 - 12:00-15:00 | 46 | 50 |

Table 2. Study population distribution

| Category | Subcategory | Number of Cases (out of 3156) | Percentage of Cases |
|----------------------|--|----------------------------------|---------------------|
| Age range [Years] | < 50 | 823 | 26.08% |
| | 50-80 | 1672 | 52.98% |
| | > 80 | 661 | 20.94% |
| Patient gender | F | 1707 | 54.09% |
| | M | 1449 | 45.91% |
| hospital Hierarchy | Comprehensive Stroke Center (CSC - Hub) | 1585 | 50.22% |
| | Primary Stroke Center (PSC - Spoke) | 1571 | 49.78% |
| Slice Thickness [mm] | 2.5 | 868 | 27.50% |
| | 3.0 | 572 | 18.12% |
| | 4.0 | 183 | 5.80% |
| | 5.0 | 1533 | 48.57% |
| | 18-39 | 1529 | 48.45% |
| Number of Slices | 40-59 | 779 | 24.68% |
| | 60-125 | 848 | 26.87% |
| | Canon Medical Systems | 22 | 0.70% |
| manufacturers | GE MEDICAL SYSTEMS | 1309 | 41.48% |
| | Hitachi, Ltd. | 6 | 0.19% |
| | Philips | 154 | 4.88% |
| | SIEMENS | 1172 | 37.14% |
| | TOSHIBA | 493 | 15.62% |

Abbreviations:

CSC - Comprehensive Stroke Center

PSC - Primary Stroke Center

MM - Millimeter

Results

Over the chosen timeframe, 3420 non-contrast head CT scans were analyzed by the Viz ICH algorithm, out of which 264 were excluded (see Truthing method section). Out of the 3156 scans, 147 were positive for ICH, with ICH prevalence of 4.6%. Scans were collected from 127 hospitals, 33 of them were comprehensive stroke centers (Hubs) and 94 primary stroke centers (spokes). Table 2 illustrates the distribution of study population according to patients' age range and gender, hospital hierarchy and scan technical parameters, such as slice thickness, number of slices and scanner manufacturer.

Study Scans Categorical Sub-Grouping**Performance Analysis**

Table 3 describes Viz ICH's performance. Viz ICH showed a high accuracy rate of 98.57%, [CI: 98%,99%]. The algorithm identified 133/147 of ICH positive scans, and 2978/3009 of ICH negative scans, resulting in sensitivity, specificity, PPV and NPV of 90.48% [CI: 85%,95%], 98.97% [CI:99%,99%], 81.1% [CI:74%, 87%] and 99.53% [CI: 99%,100%],

respectively. Performance was stratified by age, gender, and hospital hierarchy. As shown in Table 3, baseline factors (age, gender, and hospital hierarchy) did not show significant effect on sensitivity and specificity of ICH detection, based on a T-test held over these subgroups' metrics differences (P-Value<0.05). With respect to the type of hemorrhage, Viz ICH showed high sensitivity for the detection of IPH (97.62%) and SDH (90.0%) and a mixture of ICH subtypes (96.08%), whereas all other subtypes prevalence (EDH, IVH) was too low to establish a solid performance claim. Additional scan factors-based performance stratification can be seen in Appendix A.

In addition, Viz ICH mean time to notification measured was 203.14 sec (3.39 minutes) for all suspected ICH cases. The Receiver Operating Characteristic (ROC) curve for Viz ICH based on ICH volumetric detection threshold, is viewed in Figure 4. The curve demonstrates AI model detection reliability starting a volumetric threshold of the suspected ICH size detected. The Area Under the ROC Curve (AUC) was 0.97.

Hematology and Disorders

Error Analysis

Table 4 classifies False Positive (FP) cases according to the reasons for the classification confusion. Figure 5 demonstrates common FP cases. Artifacts category, containing above mentioned artifact types, only contained scans where the

artifacts did not limit the ability to identify ICH. Space occupying lesions that present as hyperdense lesions, may have a similar appearance as ICH, especially IPH. Calcifications in scans included calcified areas (pathological or benign) that were misidentified by the algorithm as ICH. Low scan quality refers to scans with a low signal to noise ratio.

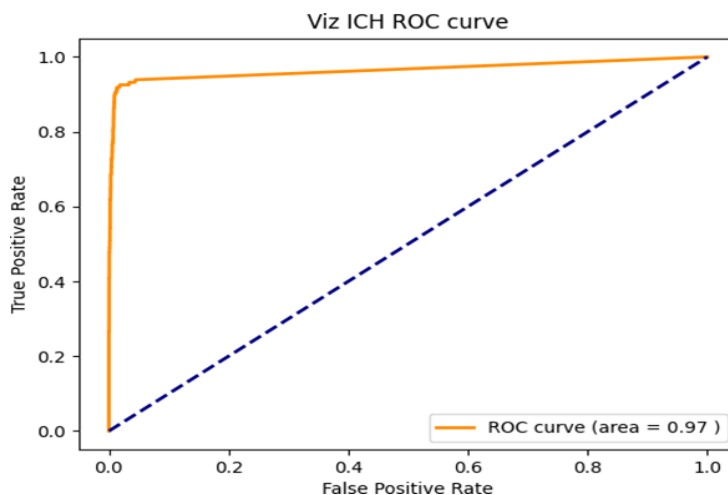


Fig 4: Viz ICH classification ROC plot, based on ICH volumetric classification threshold (ranging from 0 to ~206 mL, the largest suspected ICH volume in the examined set) determining final device classification. The ROC curve shows an AUC of 97%.

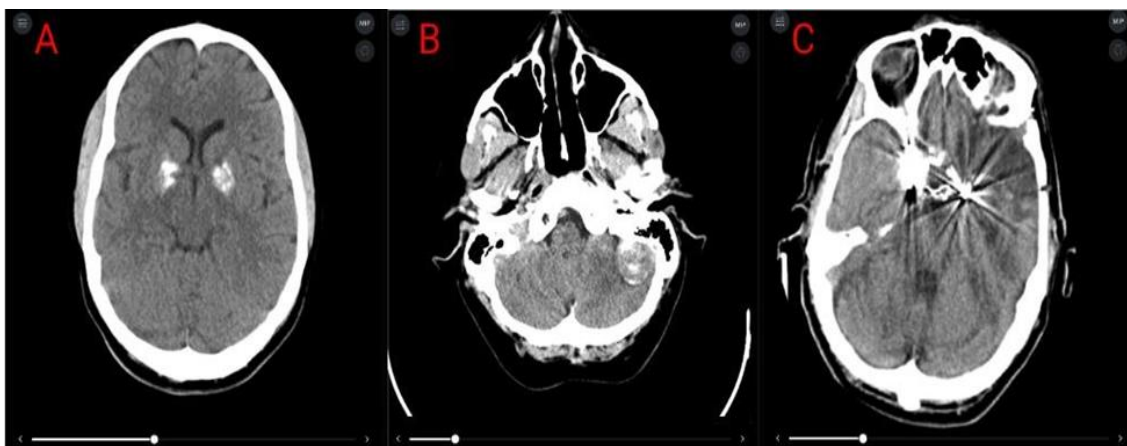


Fig 5: FP examples. Left to right: (A) Calcifications, (B) Space occupying lesion, (C) Metal artifacts.

Table 3: Viz ICH's diagnostic performance

| Category | Subcategory | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---------------------|-------------|-------------|-------------|-------------|------------|-------------|
| General Performance | - | 98.57% | 90.48% | 98.97% | 81.1% | 99.53% |
| | - | [98%, 99%] | [85%, 95%] | [99%, 99%] | [74%, 87%] | [99%, 100%] |
| | - | (3111/3156) | (133/147) | (2978/3009) | (133/164) | (2978/2992) |
| | - | 97.62% | 97.62% | | | |
| | IPH | [87%, 100%] | [87%, 100%] | **1 | ** | ** |
| | (41/42) | (41/42) | | | | |
| | 75.0% | 75.0% | | | | |
| | IVH | [19%, 99%] | [19%, 99%] | ** | ** | ** |
| | (3/4) | (3/4) | | | | |

| | | | | | | |
|----------------------------|------------------------------------|-------------------------------------|----------------------------------|-------------------------------------|-------------------|--------------------------------------|
| | | 90.0% | 90.0% | | | |
| Bleed Type | SDH | [73%, 98%] (27/30) | [73%, 98%] (27/30) | ** | ** | ** |
| | | 66.67% | 66.67% | | | |
| | SAH | [41%, 87%] (12/18) | [41%, 87%] (12/18) | ** | ** | ** |
| | | 66.67% | 66.67% | | | |
| | EDH | [9%, 99%] (2/3) | [9%, 99%] (2/3) | ** | ** | ** |
| | Multiple Bleeds | 96.08% [87%, 100%] (49/51) | 96.08% [87%, 100%] (49/51) | ** | ** | ** |
| | | | | | 80.0% | |
| Age range | | 98.91% | 88.89% | 99.25% | [61%, | 99.62% |
| | <50 | [98%, 99%] (814/823) | [71%, 98%] (24/27) | [98%, 100%] (790/796) | 92%] (24/30) | [99%, 100%] (790/793) |
| | | | | | 81.1% | |
| General Performance | | 98.57% | 90.48% | 98.97% | [74%, | 99.53% |
| | - | [98%, 99%] (3111/3156) | [85%, 95%] (133/147) | [99%, 99%] (2978/3009) | 87%] (133/164) | [99%, 100%] (2978/2992) |
| | | | | | 80.61% | |
| | | 98.27% | 88.76% | 98.8% | [71%, | 99.36% |
| | 50-80 | [98%, 99%] (1643/1672) | [80%, 94%] (79/89) | [98%, 99%] (1564/1583) | 88%] (79/98) | [99%, 100%] (1564/1574) |
| | | | | | 83.33% | |
| | | 98.94% | 96.77% | 99.05% | [67%, | 99.84% |
| | >80 | [98%, 100%] (654/661) | [83%, 100%] (30/31) | [98%, 100%] (624/630) | 94%] (30/36) | [99%, 100%] (624/625) |
| | | | | | 89.02% | |
| | | 98.62% | 86.9% | 99.34% | [80%, | 99.2% |
| Patient gender | Male | [98%, 99%] (1429/1449) | [77%, 93%] (73/84) | [99%, 100%] (1356/1365) | 95%] (73/82) | [99%, 100%] (1356/1367) |
| | | | | | 73.17% | |
| | | 98.54% | 95.24% | 98.66% | [62%, | 99.82% |
| | Female | [98%, 99%] (1682/1707) | [87%, 99%] (60/63) | [98%, 99%] (1622/1644) | 82%] (60/82) | [99%, 100%] (1622/1625) |
| | | | | | | |
| | Comprehensive Stroke Center | 98.42% [98%, 99%] (1560/1585) | 90.53% [83%, 96%] (86/95) | 98.93% [98%, 99%] (1474/1490) | 84.31% [76%, | 99.39% [99%, 100%] (1474/1483) |
| | | | | | 91%] | |

Hematology and Disorders

| hospital | (Hub) | (86/102) | | | | |
|-----------|---------|-------------|------------|-------------|---------|-------------|
| Hierarchy | Primary | 98.73% | 90.38% | 99.01% | 75.81% | 99.67% |
| | Stroke | [98%, 99%] | [79%, 97%] | [98%, 99%] | [63%, | [99%, 100%] |
| | Center | (1551/1571) | (47/52) | (1504/1519) | 86%] | (1504/1509) |
| | (Spoke) | | | | (47/62) | |

Abbreviations:

PPV – Positive predictive value

NPV – Negative predictive value

IVH – Intraventricular hemorrhage

SAH – Subarachnoid hemorrhage

SDH – Subdural hemorrhage

EDH – Epidural hemorrhage

PSC – Primary stroke centers

CSC – Comprehensive stroke centers

Table 4: Classification of false positive cases according to FP reason

| | Number of Cases | Percentage of Cases |
|---|-----------------|---------------------|
| Artifacts (Skull artifacts, scan artifacts and metal artifacts) | 9 | 29.03% |
| Space Occupying Lesion (SOL) | 7 | 22.58% |
| Calcifications | 3 | 9.68% |
| Low Scan Quality | 3 | 9.68% |
| Other | 9 | 29.03% |
| Total | 31 | |

Table 5: Viz ICH's diagnostic performance - additional technical factors sub-grouping.

| Category | Sub- Category | Accuracy | Sensitivity | Specificity | PPV | NPV | |
|------------------|---------------|-------------|-------------|-------------|-----------|---------|-----------|
| Number of Slices | 18-39 | 98.82% | 92.54% | 99.11% | 82.67% | 82.67% | |
| | | [98%, 99%] | [83%, | [98%, | [72%, | [72%, | |
| | | (1511/1529) | 98%] | 100%] | 90%] | 90%] | |
| | 40-59 | | | (62/67) | 2) | (62/75) | (62/75) |
| | | | | 86.11% | 98.12% | 68.89% | 99.32% |
| | | | 97.56% | | | | |
| | 60-125 | | | [71%, | [97%, | [53%, | [98%, |
| | | | [96%, 99%] | 95%] | 99%] | 82%] | 100%] |
| | | | (760/779) | | | | |
| | | | | (31/36) | (729/743) | (31/45) | (729/734) |
| | | 90.91% | 99.5% | 90.91% | 99.5% | | |
| | 99.06% | | | | | | |
| | | [78%, | [99%, | [78%, | [99%, | | |
| | [98%, | 97%] | 100%] | 97%] | 100%] | | |
| | (840/848) | | | | | | |
| | | (40/44) | (800/804) | (40/44) | (800/804) | | |
| | | 89.13% | 99.39% | 89.13% | 99.39% | | |
| | | 98.85% | | | | | |

| | | | | | | |
|------------------------|--------------------|-------------------------------------|----------------------------------|-----------------------------|-------------------------|-----------------------------|
| | | [98%, 99%] (858/868) | [76%, 96%] | [99%, 100%] | [76%, 96%] | [99%, 100%] |
| | 2.5 | | (41/46) 100.0% | (817/822) 98.21% | (41/46) 58.33% | (817/822) 100.0% |
| Slice Thickness | | 98.25% [97%, 99%] (562/572) | [77%, 100%] | [97%, 99%] | [37%, 78%] | [99%, 100%] |
| | 3 | | (14/14) 83.33% | (548/558) 98.79% | (14/24) 88.24% | (548/548) 98.19% |
| | | | [59%, [94%, 99%] (178/183) | [96%, 100%] | [64%, 99%] | [95%, 100%] |
| | 4 | | (15/18) 98.7% | (163/165) 99.04% | (15/17) 81.82% | (163/166) 99.59% |
| | 5 | 98.7% [98%, 99%] (1513/1533) | [82%, 97%] | [98%, 99%] (1450/146 | [71%, 90%] | [99%, 100%] (1450/145 |
| | | | (63/69) | 4) | (63/77) | 6) |
| | | | | 98.67% | | 99.64% |
| | | 98.38% [97%, 99%] (1153/1172) | 91.11% [79%, 98%] | [98%, 99%] (1112/112 | 73.21% [60%, 84%] | [99%, 100%] (1112/111 |
| | SIEMENS | | (41/45) | 7) | (41/56) | 6) |
| | | | | 99.2% | | 99.52% |
| | | 98.78% [98%, 99%] (1293/1309) | 90.62% [81%, 96%] | [99%, 100%] (1235/124 | 85.29% [75%, 93%] | [99%, 100%] (1235/124 |
| Manufacturers | GE MEDICAL SYSTEMS | | (58/64) 85.71% | 5) | (58/68) | 1) |
| | | 99.35% | | 100.0% | 100.0% | 99.32% |
| | | [96%, 100%] (153/154) | [42%, 100%] | [98%, 100%] | [54%,100 %] | [96%, 100%] |
| | Philips | | (6/7) 90.32% | (147/147) 98.92% | (6/6) 84.85% | (147/148) 99.35% |
| | | 98.38% [97%, 99%] (485/493) | [74%, 98%] | [97%, 100%] | [68%, 95%] | [98%, 100%] |
| | TOSHIBA | | (28/31) | (457/462) | (28/33) | (457/460) |

Hematology and Disorders

Discussion

Viz ICH is an AI based tool, intended for analyzing NCCT images of the brain are often acquired in the acute setting and demand rapid interpretation for optimal patient care. When the algorithm identifies a scan as ICH suspected, the system notifies a multidisciplinary team, recommending immediate review of those images. Images can be promptly previewed through a web interface or through a mobile application, allowing clinicians to assess the case in a remote manner.

Over the last years, radiologists' workload has increased dramatically [17]. Long-hour shifts, increasing number of scans, and manpower shortage may not only lead to radiologist' burnout but also risk critical diagnostic errors [18,19]. Since medical imaging is a major contributor to the overall diagnostic process, it is also a major potential source of diagnostic error [20]. Clinically significant errors may vary between 2-20% [21]. Different methods such as double reporting was shown to be effective [22] but are time consuming and demand more workforce [23]. AI based algorithms, such as Viz ICH, could be used to improve the workload on radiologists, shorten reporting time and serve as additional tools for radiologists.

ICH is a life-threatening condition, with high mortality rates, and complex morbidity [24]. ICH is generally an acutely evolving process, for which quick diagnosis and treatment are key to better outcomes [24,25]. Most importantly missed ICH in NCCT scans may have a fatal impact on patients, leading to delay in diagnosis, complications, and death [26]. AI tools, designed to detect and notify about suspected ICH, could help decrease error rate, minimize false negatives diagnosis [27] and shorten notification time.

In this study, we aimed to examine the performance of Viz ICH in real-life settings. 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, 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 patient motion.

- **Overfitting due to lack of diversity in data:** Overfitting [28] is a general problem in modeling whereby a model may have good predictive capabilities on a training dataset, but lower performance on real-world data. While there are many reasons for this phenomenon, the most relevant for AI applications 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. When a device is developed using a dataset with limited diversity, there is risk that performance on data of greater diversity will be lower.
- **Accuracy:** Since Viz ICH was approved, a great amount of data was processed from many sites. This real-world data could be leveraged to evaluate the AI performance more accurately.
- **Generalization:** 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. The value of such evidence is to demonstrate the realistic possibility.
- **Data Drift:** Supervised AI algorithms employ past examples-based learning to predict requested outputs. In a clinical dynamic environment setting, this static approach preserves the original data distributions learned. A likely outcome of this setup is a decrease in performance over time due to a change in data distribution due to either technical or clinical changes that manifest in imaging visual modifications. Practically this means deployed algorithms must be monitored to ensure their safety over time.

In 2022, Matsoukas, Stavros, et al [29] published an article examining Viz ICH real world performance, in all hospitals of the Mount Sinai Health System. In the

here presented study, we aimed to examine a larger cohort of sites, varying in equipment, population, protocol, and personnel skills, thus exploring AI model generalization over a diversified dataset.

In this study, we examined scans from 127 hospitals, 33 comprehensive stroke centers (Hubs) and 94 primary stroke centers (spokes). The hospitals are using different CT manufactures and protocols as seen in table 2. The extensive and varied nature of our study mitigate the risk of bias and the artificial inflation of performance metrics, commonly observed in studies featuring minimally diverse and homogeneous cohorts. Such comprehensive coverage ensures our findings are robust, reflecting true clinical utility rather than the biased results of more limited research settings, such as single site studies.

The algorithm demonstrated robustness to differences in sensitivity and specificity over all subgroups such as gender, age group and hospital hierarchy. When comparing FP scans causes between CSC and PSCs, the latter group had shown larger amounts of SOLs misidentified as ICH by Viz ICH, a factor unrelated to the hospital hierarchy. In addition, a slight increase in artifacts containing scans was spotted in the PSCs subgroup, which in combination with the lower sample size of this group compared with CSC, may be the cause for the difference detected in PPV. A different prevalence of stroke suspected patients, arriving at PSCs vs. CSC can also provide further justification to the difference in PPV.

Viz ICH mean time to notification is 3.39 minutes. Such notification time allows a multidisciplinary team to review suspected cases with higher priority, which can lead to faster diagnosis, assist treatment decisions, and result in better outcomes. When examining radiologists' performance, their mean time to report is 132 minutes. When an AI algorithm assists the radiologists, their mean time to report is shortened to 73 minutes [30].

When examining Viz ICH's sensitivity to bleed types, variations in sensitivity are seen. Viz ICH is more sensitive to IPH (97.62% CI 87%,100%) and SDH (90.0% CI: 73%,98%), than to IVH (75% CI 19%,99%), EDH (66.67% CI 9%,99%) and SAH (64.71% CI 38%,86%). It is worth mentioning that IVH EDH, and SAH have low prevalence in this study (five IVH positive cases, 3 EDH positive cases, 18 SAH positive cases), resulting in a wide confidence interval, questioning the statistical significance of these specific findings. Hence, further investigation with a cohort enhanced with positive cases for EDH,

IVH and SAH is needed to determine Viz ICH's sensitivity for these ICH subtypes. It is also worth mentioning, that the prevalence of SAH in this study was low with only 18 positive cases, equivalent to 0.006%. In 2019, the US recorded 0.85 million SAH cases [31], with an estimated population of 328.24 million for that year [32]. Based on these parameters, we can estimate the SAH prevalence to be 0.003%. SAH incidence rates vary between 6.2 and 11.15 per 100,000 people [33,34]. A low disease prevalence is typically associated with lower sensitivity in detection. The low prevalence of SAH, could account for the observed lower sensitivity of SAH compared to the overall sensitivity for all bleed types in this study.

ICH prevalence in this study was 4.6%, which is significantly higher than ICH prevalence in the general population documented in epidemiological studies [35,36]. Other than the fact that the subpopulation of people who go through head CT in the hospital setting is more prone to have ICH than the general population, this gap could be explained by the fact that scans reviewed by Viz ICH algorithm are only scans that were included upon certain criteria (e.g., head injury, stroke protocol, etc.), which are more susceptible to be positive for ICH, compared to the general population.

Limitations

The study data was collected from the data available for the Viz ICH algorithm. Important information regarding patients, such as medical history, clinical presentation, laboratory exams and clinical outcomes, was not available for this study.

Literature review of ICH subtype prevalence lacked coherent measurements of comparison [33,34,35], which leads to a difficulty determining each subtype prevalence in the general population. Thus, we couldn't assess the external validity of this article in relation to real world prevalence of ICH subtypes.

Conclusion

Viz ICH demonstrates robust performance in identifying ICH on NCCT in most parameters, despite the heterogeneity of setting, equipment, and processes. Therefore, it can be considered a reliable and efficient AI tool for medical centers of all sizes and capabilities. It shows excellent accuracy, sensitivity, specificity and NPV values, as well as swift time to notification, which can have a significant impact on patients' lives.

Conflict of interest

All the authors are employees of Viz.ai, and receive a salary from Viz.ai, the company responsible for the

Hematology and Disorders

development and commercialization of the AI algorithm discussed in this article.

References

1. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K210209>
2. Elliott, J., & Smith, M. (2010). The acute management of intracerebral hemorrhage: a clinical review. *Anesthesia & Analgesia*, 110(5), 1419-1427.
3. Feigin, V. L., Lawes, C. M., Bennett, D. A., Barker-Collo, S. L., & Parag, V. (2009). Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *The Lancet Neurology*, 8(4), 355-369.
4. Krishnamurthi, R. V., Feigin, V. L., Forouzanfar, M. H., Mensah, G. A., Connor, M., Bennett, D. A., ... & Murray, C. (2013). Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *The Lancet Global Health*, 1(5), e259-e281.
5. Subaiya, S., Roberts, I., Komolafe, E., & Perel, P. (2012). Predicting intracranial hemorrhage after traumatic brain injury in low and middle-income countries: a prognostic model based on a large, multi-center, international cohort. *BMC emergency medicine*, 12(1), 1-7.
6. Kauvar, D. S., & Wade, C. E. (2005). The epidemiology and modern management of traumatic hemorrhage: US and international perspectives. *Critical Care*, 9(5), 1-9.
7. Corrigan, J. D., Selassie, A. W., & Orman, J. A. L. (2010). The epidemiology of traumatic brain injury. *The Journal of head trauma rehabilitation*, 25(2), 72-80.
8. Elliott, J., & Smith, M. (2010). The acute management of intracerebral hemorrhage: a clinical review. *Anesthesia & Analgesia*, 110(5), 1419-1427.
9. Flower, O., & Smith, M. (2011). The acute management of intracerebral hemorrhage. *Current Opinion in Critical Care*, 17(2), 106-114.
10. Kellner, C. P., Schupper, A. J., & Mocco, J. (2021). Surgical evacuation of intracerebral hemorrhage: the potential importance of timing. *Stroke*, 52(10), 3391-3398.
11. Bir, S. C., Maiti, T. K., Konar, S., & Nanda, A. (2016). Overall outcomes following early interventions for intracranial arteriovenous malformations with hematomas. *Journal of Clinical Neuroscience*, 23, 95-100.
12. Herpich, F., & Rincon, F. (2020). Management of acute ischemic stroke. *Critical care medicine*, 48(11), 1654.
13. Phipps, M. S., & Cronin, C. A. (2020). Management of acute ischemic stroke. *Bmj*, 368.
14. Power, S. P., Moloney, F., Twomey, M., James, K., O'Connor, O. J., & Maher, M. M. (2016). Computed tomography and patient risk: Facts, perceptions and uncertainties. *World journal of radiology*, 8(12), 902.
15. Lantsman, C. D., Barash, Y., Klang, E., Guranda, L., Konen, E., & Tau, N. (2022). Trend in radiologist workload compared to number of admissions in the emergency department. *European Journal of Radiology*, 149, 110195.
16. Kocher, K. E., Meurer, W. J., Fazel, R., Scott, P. A., Krumholz, H. M., & Nallamothu, B. K. (2011). National trends in use of computed tomography in the emergency department. *Annals of emergency medicine*, 58(5), 452-462.
17. Bruls, R. J. M., & Kwee, R. M. (2020). Workload for radiologists during on-call hours: dramatic increase in the past 15 years. *Insights into imaging*, 11, 1-7.
18. Ruutiainen, A. T., Durand, D. J., Scanlon, M. H., & Itri, J. N. (2013). Increased error rates in preliminary reports issued by radiology residents working more than 10 consecutive hours overnight. *Academic radiology*, 20(3), 305-311.
19. Hanna, T. N., Shekhani, H., Lamoureux, C., Mar, H., Nicola, R., Sliker, C., & Johnson, J. O. (2017). Emergency radiology practice patterns: shifts, schedules, and job satisfaction. *Journal of the American College of Radiology*, 14(3), 345-352.
20. Bruno, M. A., Walker, E. A., & Abujudeh, H. H. (2015). Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *Radiographics*, 35(6), 1668-1676.
21. Goddard, P., Leslie, A., Jones, A., Wakeley, C., & Kabala, J. (2001). Error in radiology. *The British journal of radiology*, 74(886), 949-951.
22. Pow, R. E., Mello-Thoms, C., & Brennan, P. (2016). Evaluation of the effect of double reporting on test accuracy in screening and diagnostic imaging studies: a review of the evidence. *Journal of medical imaging and*

- radiation oncology, 60(3), 306-314.
23. Lauritzen, P. M., Hurlen, P., Sandbæk, G., & Gulbrandsen, P. (2015). Double reading rates and quality assurance practices in Norwegian hospital radiology departments: two parallel national surveys. *Acta radiologica*, 56(1), 78-86.
 24. Elliott, J., & Smith, M. (2010). The acute management of intracerebral hemorrhage: a clinical review. *Anesthesia & Analgesia*, 110(5), 1419-1427.
 25. Kellner, C. P., Schupper, A. J., & Mocco, J. (2021). Surgical evacuation of intracerebral hemorrhage: the potential importance of timing. *Stroke*, 52(10), 3391-3398.
 26. Jenny, C., Hymel, K. P., Ritzen, A., Reinert, S. E., & Hay, T. C. (1999). Analysis of missed cases of abusive head trauma. *Jama*, 281(7), 621-626.
 27. Rao, B., Zohrabian, V., Cedeno, P., Saha, A., Pahade, J., & Davis, M. A. (2021). Utility of artificial intelligence tool as a prospective radiology peer reviewer—Detection of unreported intracranial hemorrhage. *Academic radiology*, 28(1), 85-93.
 28. Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1), 1-12.
 29. Matsoukas, S., Chennareddy, S., Kalagara, R., Scaggiante, J., Smith, C. J., Bazil, M. J., ... & Kellner, C. P. (2022). Pilot deployment of viz–intracranial hemorrhage for intracranial hemorrhage detection: real-world performance in a stroke code cohort. *Stroke*, 53(9), e418-e419.
 30. Wismüller, A., & Stockmaster, L. (2020, February). A prospective randomized clinical trial for measuring radiology study reporting time on Artificial Intelligence-based detection of intracranial hemorrhage in emergent care head CT. In *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging* (Vol. 11317, pp. 144-150). SPIE.
 31. Renedo, D., Acosta, J. N., Leasure, A. C., Sharma, R., Krumholz, H. M., de Havenon, A., ... & Sheth, K. N. Burden of Ischemic and Hemorrhagic Stroke Across the US From 1990 to 2019. *JAMA neurology*.
 32. <https://www.census.gov/newsroom/press-releases/2019/popest-nation.html>
 33. Korja, M., Lehto, H., Juvela, S., & Kaprio, J. (2016). Incidence of subarachnoid hemorrhage is decreasing together with decreasing smoking rates. *Neurology*, 87(11), 1118-1123.
 34. Linn, F. H. H., Rinkel, G. J. E., Algra, A., & Van Gijn, J. (1996). Incidence of subarachnoid hemorrhage: role of region, year, and rate of computed tomography: a meta-analysis. *Stroke*, 27(4), 625-629.
 35. Rincon, F., & Mayer, S. A. (2013). The epidemiology of intracerebral hemorrhage in the United States from 1979 to 2008. *Neurocritical care*, 19, 95-102.
 36. An, S. J., Kim, T. J., & Yoon, B. W. (2017). Epidemiology, risk factors, and clinical features of intracerebral hemorrhage: an update. *Journal of stroke*, 19(1), 3.
 37. Wang, S., Zou, X. L., Wu, L. X., Zhou, H. F., Xiao, L., Yao, T., ... & Zhang, L. (2022). Epidemiology of intracerebral hemorrhage: A systematic review and meta-analysis. *Frontiers in Neurology*, 13, 915813.
 38. Li, Q., Li, R., Zhao, L. B., Yang, X. M., Yang, W. S., Deng, L., ... & Xie, P. (2020). Intraventricular hemorrhage growth: definition, prevalence and association with hematoma expansion and prognosis. *Neurocritical Care*, 33, 732-739.
 39. Tenny S, Thorell W. Intracranial Hemorrhage. [Updated 2023 Feb 13]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK470242/>