Not all AI is equal:
A comparative study of three imaging algorithms

aidoc

RÖNTGEN-VECKAN

SÖS SÖDERSJUKHUSET

NOT ALL AI IS EQUAL: A COMPARATIVE STUDY OF THREE IMAGING ALGORITHMS

This pre-publication report, presented with consent from the authors below, is provided by Aidoc to highlight key findings from a retrospective comparative study of ICH algorithms, as presented at Röntgenveckan in September 2023.

Authors: *Dr. Kristoffer Järlevi, Dr. Michael Wilczek, Dr. Chikako Suzuki and Mr. Gustav Alvfeldt MMSc*
Affiliation: *Radiology Department, Stockholm South Emergency Hospital (Södersjukhuset, SÖS)*

PURPOSE

To assess the potential of three commercially available AI algorithms in flagging and triaging cases of intracranial hemorrhage—a critical condition requiring swift intervention for positive patient outcomes. This evaluation is of utmost importance given the current landscape of workforce shortages and increased imaging demands, compounded by the common perception that all AI algorithms are equal - and interchangeable - and are therefore often selected based on non-performance-related factors.

Moreover, despite the growing clinical evidence demonstrating AI's benefits, most published studies tend to feature individual algorithms, making meaningful comparisons a challenge. Hence, this research not only aids SÖS's radiology department in making well-informed algorithm selections but also holds substantial value for the wider healthcare industry as they consider their own adoption of AI for medical imaging.

MATERIALS AND METHODS

Three AI algorithms were evaluated for their ability to flag intracranial hemorrhage (ICH) cases in non-contrast computed tomography (CT) scans. Aidoc was one of the algorithms, while the identities of the other two providers remain confidential and are denoted as Vendor A and Vendor B.

4027 adult head CT studies without contrast were taken from the Sectra RIS/PACS (including images and corresponding reports). 125 scans were manually excluded prior to the study, leaving 3902 cases of unique patients for analysis.

To ensure a fair comparison, each vendor's algorithm complied with the same methods for the analysis of retrospective data, which meant that only DICOM-related data from the PACS was analysed, without additional support from any other data sources. Notably, Aidoc's ICH algorithm ran on its award-winning AI operating system (aiOS™), and two other algorithms from vendors "A" and "B" were deployed via an AI marketplace.

Figure 1 illustrates how scans were analysed by each algorithm individually and how any case flagged as positive by the AI algorithms, or positive or uncertain according to the original radiology reports were subjected to manual review to establish the ground truth.

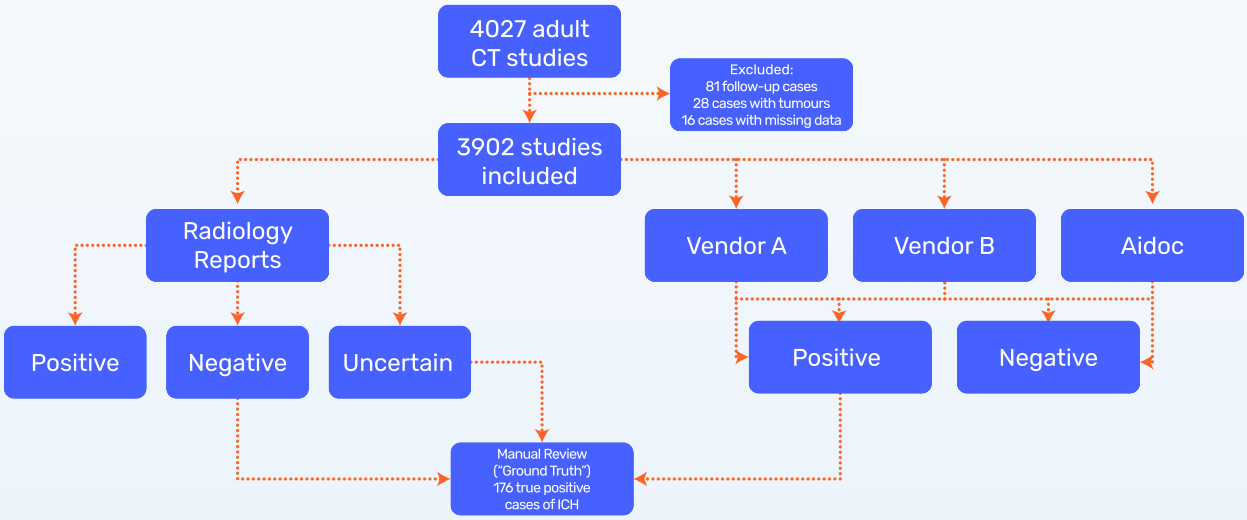

Figure 1: Method to determine the ground truth.

RESULTS

A total of 3,902 scans were analysed, of which there were 176 (5%) true positive cases of ICH. The results in **Table 1** present the sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV) of the three algorithms and the primary and double radiology reviews. Among the three algorithms, Aidoc's ICH algorithm consistently outperformed Vendor A and B, exhibiting a 29% higher sensitivity and 1.75% greater specificity compared to the average outcomes of the other two algorithms. Notably, Aidoc's algorithm was also the only algorithm with a performance comparable to the unaided, primary reviewing radiologist.

	Sensitivity	Specificity	PPV	NPV
Primary review	86.4%	99.4%	87.9%	99.4%
Double review	95.5%	99.5%	90.8%	99.8%
Vendor A	60.2%	97.1%	48.4%	98.2%
Vendor B	62.8%	97.4%	52.4%	98.3%
Aidoc	90.3%	99.0%	80.3%	99.5%

Table 1: Sensitivity, specificity, PPV and NPV results from Aidoc's algorithm, anonymised vendors B and C and primary and double radiologist reviews.

Unsurprisingly, **Table 2** shows that as the highest performer, only the Aidoc ICH algorithm enhanced the sensitivity of the primary review, compared to the radiologist working alone.

	Sensitivity	Specificity	PPV	NPV
Primary review	86.4%	99.4%	87.9%	99.4%
Double review	95.5%	99.5%	90.8%	99.8%
Primary Review + Aidoc	96.0%	99.4%	88.9%	99.8%
Double Review + Aidoc	98.9%	99.5%	91.1%	99.9%

Table 2: Sensitivity, specificity, PPV and NPV results from Aidoc's algorithm when shown in conjunction with primary and double radiology review.

Conclusion

The results clearly demonstrate significant disparities among the algorithms, highlighting the unequal capabilities of different AI solutions. Among the three prominent commercially available algorithms, Aidoc's algorithm run via its aiOS™ platform, stood out with notably higher sensitivity and specificity in comparison to the other two. Furthermore, it demonstrated the potential to substantially enhance sensitivity when used alongside a radiologist, as opposed to a radiologist working alone.

A key distinguishing feature of Aidoc's platform is its proprietary AI operating system, known as aiOS™, setting it apart from other AI vendors. This unified operating system offers various advantages, including the utilisation of image-based orchestration to select the relevant studies and most suitable image series, ensuring increased resilience against real-world data variability, as well as a robust monitoring infrastructure ensuring consistent performance over time. The combination of the algorithm with aiOS™ played a significant role in Aidoc's performance. Algorithms deployed via an AI marketplace typically lack key capabilities in terms of optimal studies and series selection.

Following this retrospective research, Södersjukhuset has seamlessly integrated the Aidoc aiOS™ into their clinical production workflow, running algorithms for ICH, iPE and PE, with the intention of validating previous findings through a prospective study. This initiative is geared towards elevating patient care and optimising the efficiency of care pathways.

For more information on Aidoc, please visit **www.aidoc.com**