

Lesion-Level Validation of a Deep-Learning Model
for Intracranial Aneurysm Detection in TOF-MRA

deephealth
IMAGIR
Centres d'imagerie médicale

B. Doche de Laquintane¹, A. Drier¹, P. Rubini², J. Brioulet², A. El Ameen³, M. Raufaste¹, V.Munoz-Ramirez²
(1) IMAGIR centre Lecocq, Bordeaux, France (2) Pixyl, Grenoble, France (3) Lyon, France

Introduction

- Early detection of intracranial aneurysms enables risk stratification and treatment planning, reducing the risk of rupture in high-risk aneurysms.
- TOF-MRA is non-ionizing and non-invasive which makes it ideal for screening.
- Detection sensitivity is prone to high inter-rater sensitivity, depending on expertise.

How does this AI model for aneurysm detection compare with general radiologists ?
Could they help improve detection consistency ?

Data & Methods

251
TOF-MRA exams

8
Participant centers

148
Positive exams

199
Aneurysms

4 US and 4 FR centers
82 US and 169 FR exams

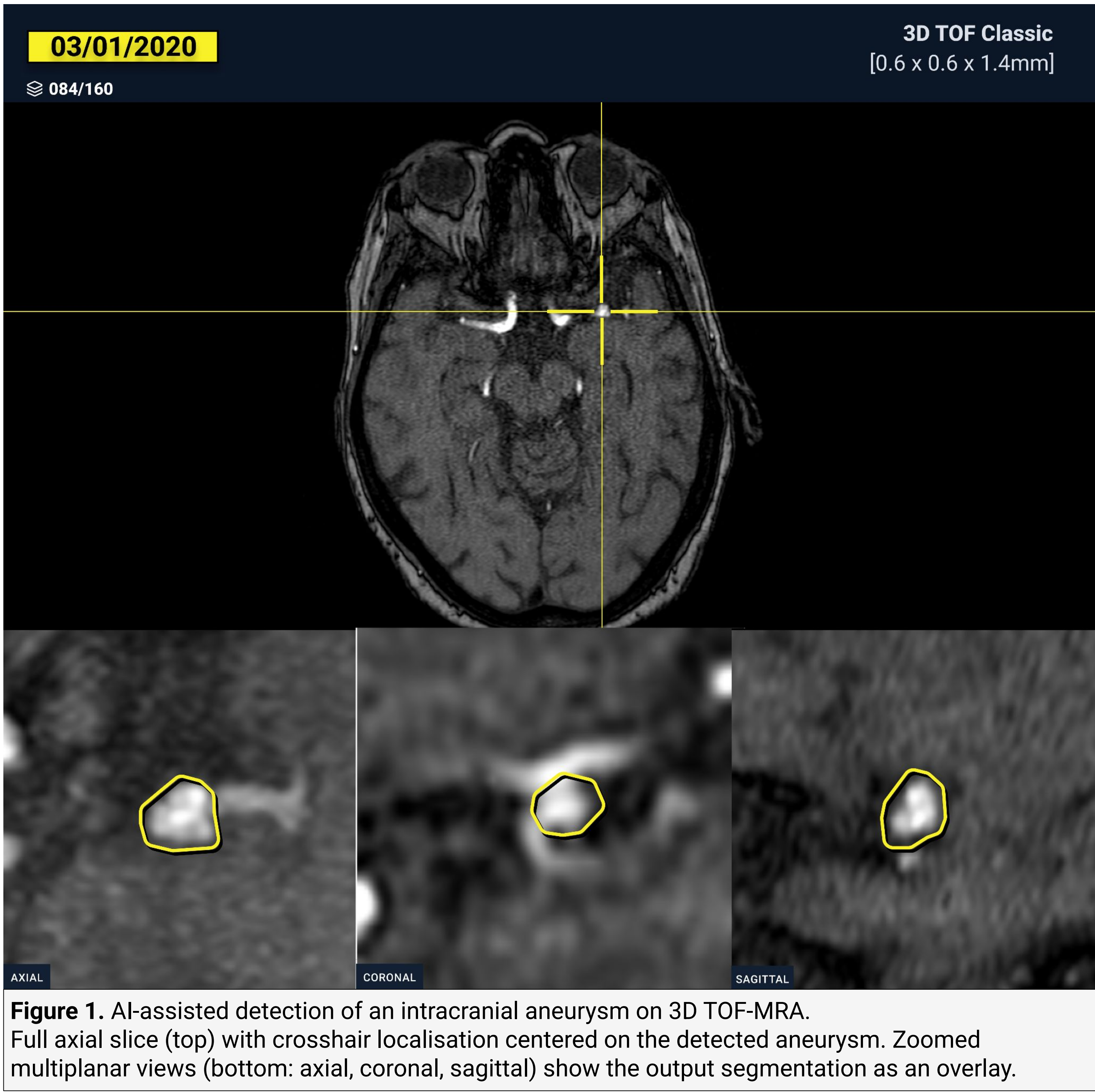
Ground truth:
2 expert neuroradiologists
(18 and 23 years of experience)

AI-model:
Supervised deep-learning

Comparison:
3 general radiologists
(5 years of experience each)

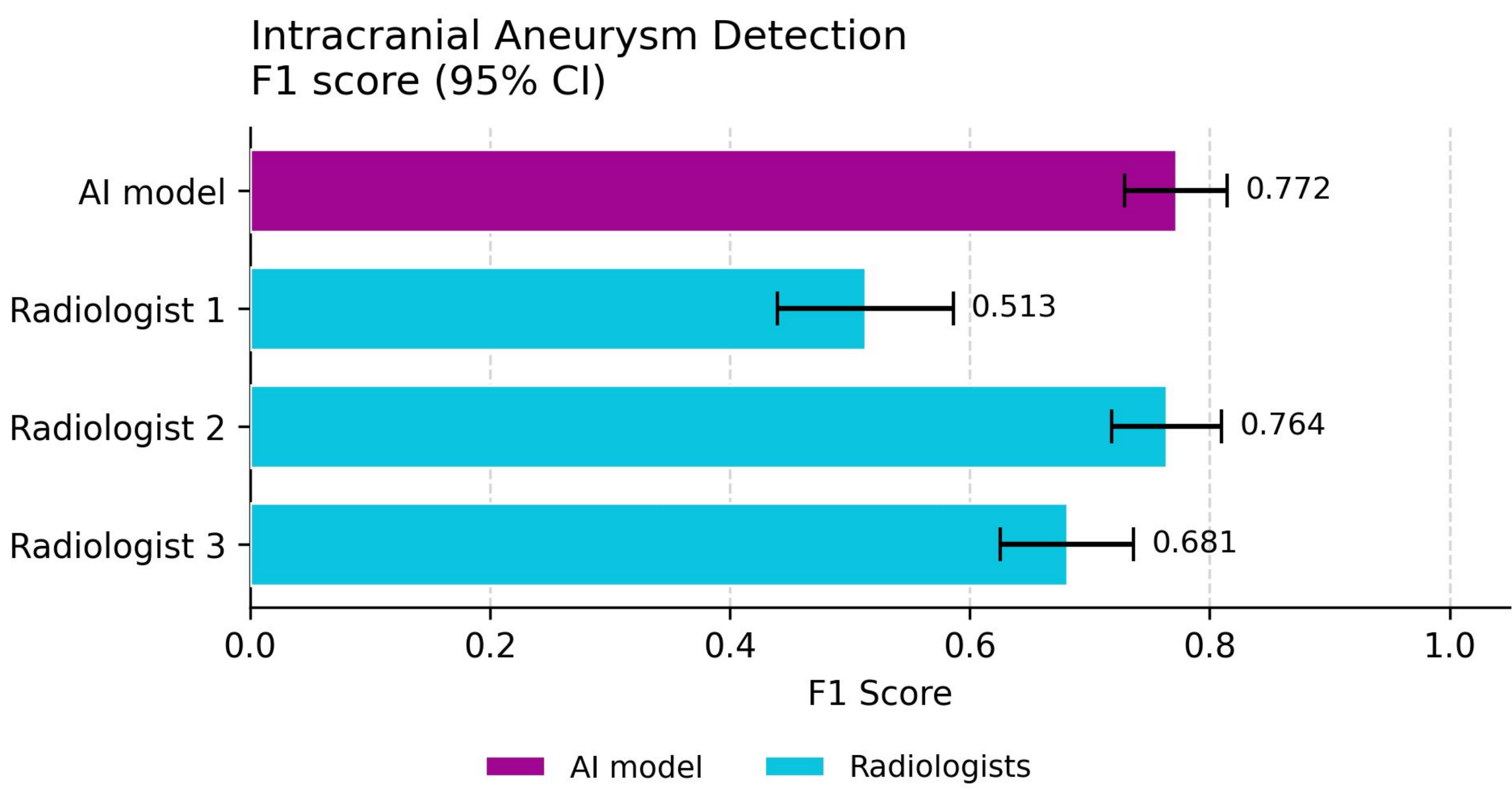
Metrics:
Lesion-level sensitivity,
precision and F1-score

Results



The AI model reached a lesion-level sensitivity of 73.4% [95% CI: 67-79], close to the best radiologist (74.9% [68-81]) and higher than the others (35.7% [29-43] and 55.8% [49-63]). Precision was intermediate (81.6% [75-87]): lower than that of the least sensitive readers (91% [82-96] and 87,4% [80-93]) but higher than that of the most sensitive reader (78% [71-84]).

Overall, the AI achieved the most balanced performance, with the highest F1-score (77.2% [73-81]), indicating the strongest sensitivity–precision compromise.



Conclusion

The AI model demonstrates a more favourable balance between sensitivity and precision than general radiologists, suggesting its potential to improve sensitivity, enhance consistency, and reduce inter-reader variability in TOF-MRA aneurysm detection.

Limitations & Perspectives

This retrospective study is limited by its relatively small cohort, and the comparison against general radiologists only. Future work could assess its impact on reader performance when used as a second-reader with a wider variety of expertise levels in neuroradiology.