Politechnika Śląska

GRAYLIGHT
MEDICAL IMAGING SOFTWARE

CAN AI MAKE US SEE BEYOND THE VISIBLE: TOWARD CE MARKED DEEP LEARNING SOFTWARE FOR MEDICAL IMAGE ANALYSIS

Jakub Nalepa
jnalepa@ieee.org

Pesapane *et al. European Radiology Experimental* (2018) 2:35
https://doi.org/10.1186/s41747-018-0061-6

European Radiology Experimental

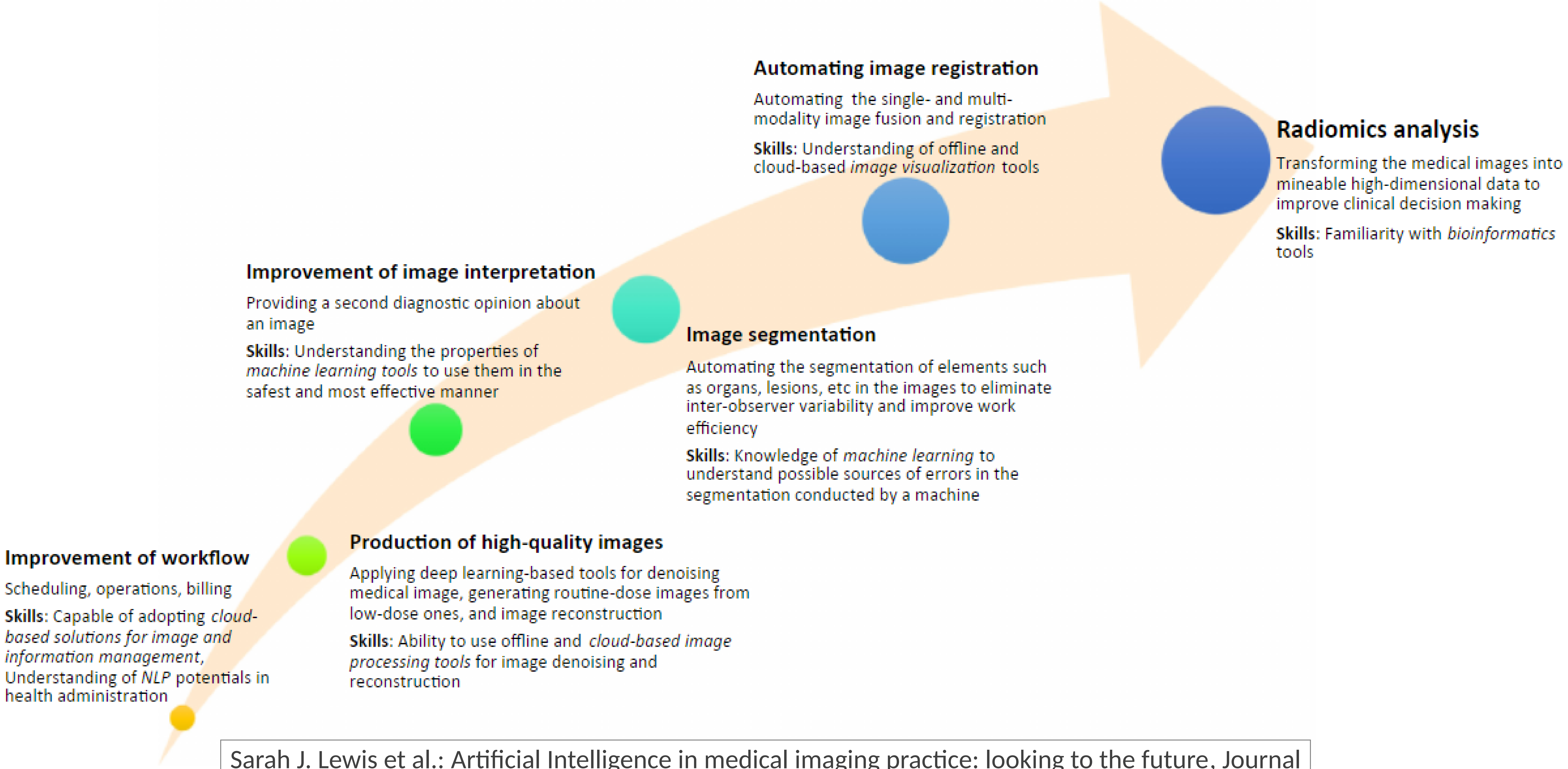
NARRATIVE REVIEW

Open Access

CrossMark

Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine

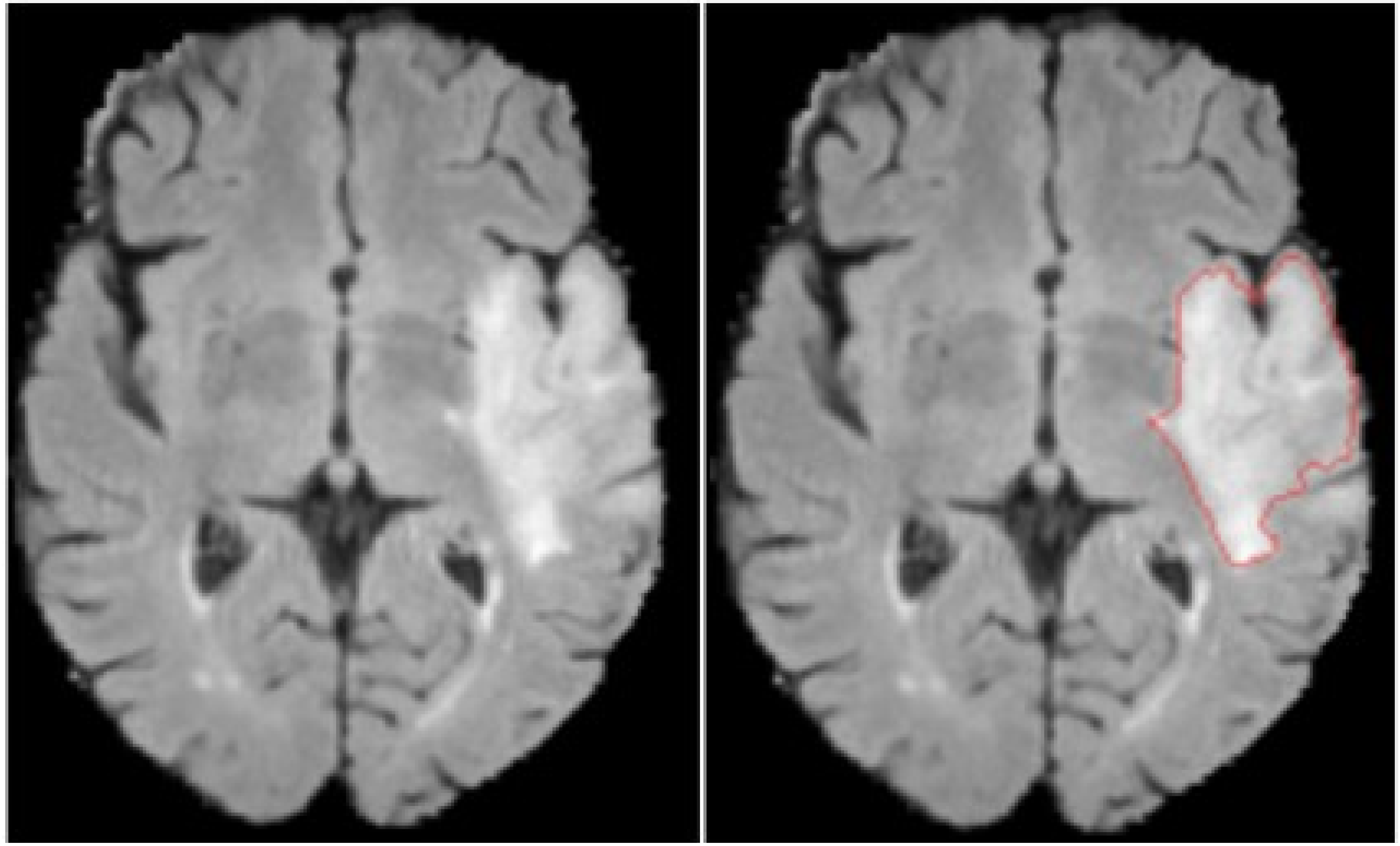
Filippo Pesapane[1†], Marina Codari[2*†] and Francesco Sardanelli[2,3]



Sarah J. Lewis et al.: Artificial Intelligence in medical imaging practice: looking to the future, Journal of Medical Radiation Sciences, First published: 10 November 2019 https://doi.org/10.1002/jmrs.369

L. Shen et al., Deep Learning to Improve Breast Cancer Detection on Screening Mammography, Scientific Reports volume 9, Article number: 12495 (2019)

CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (BRAIN MRI)

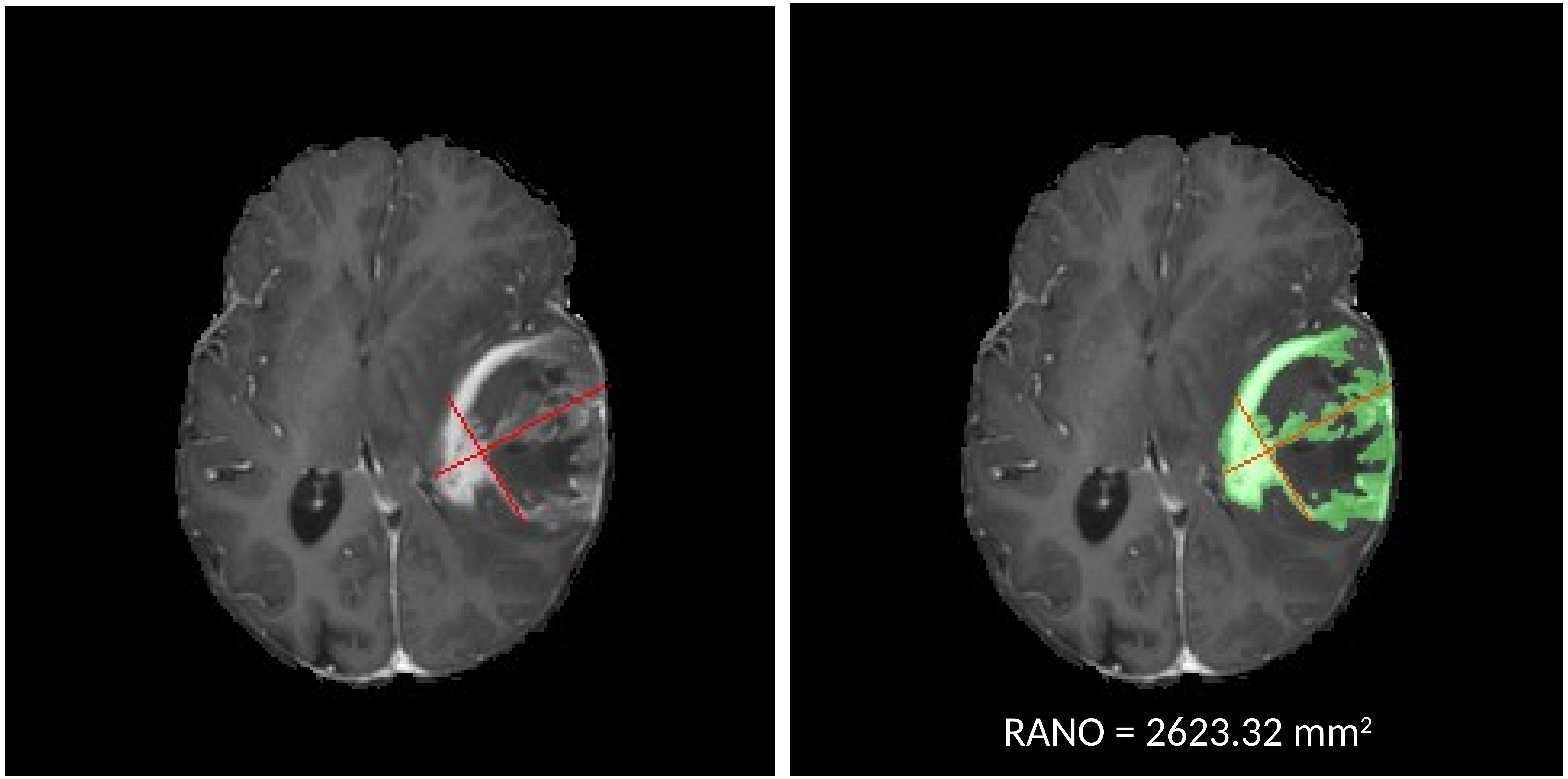


J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

RSNA Radiological Society of North America

ESR EUROPEAN SOCIETY OF RADIOLOGY

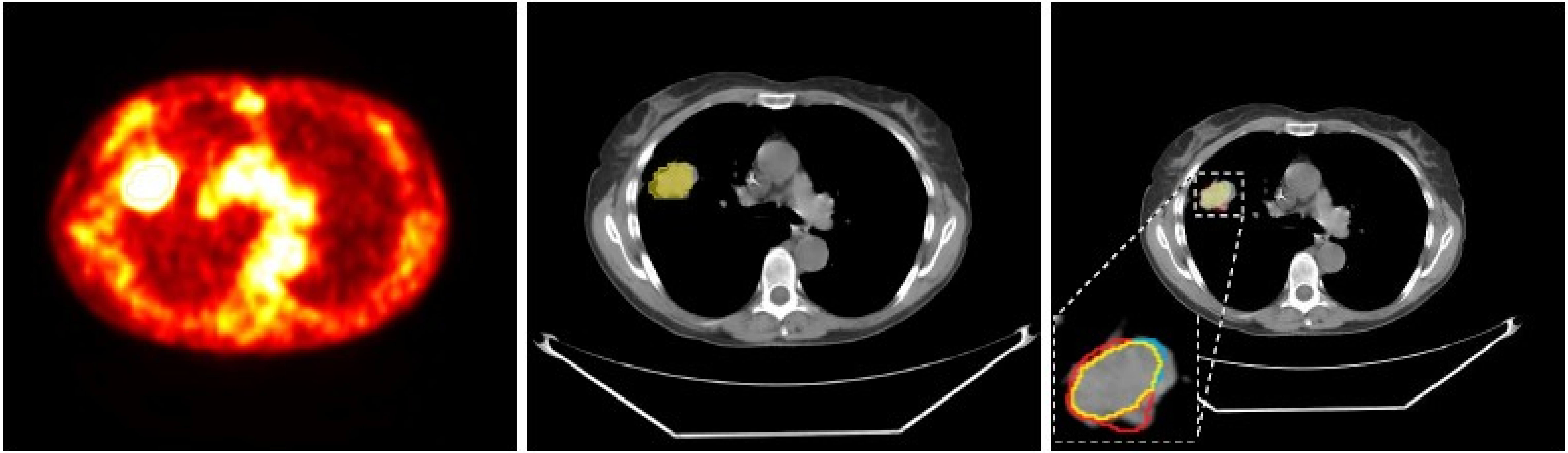
CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (BRAIN MRI)



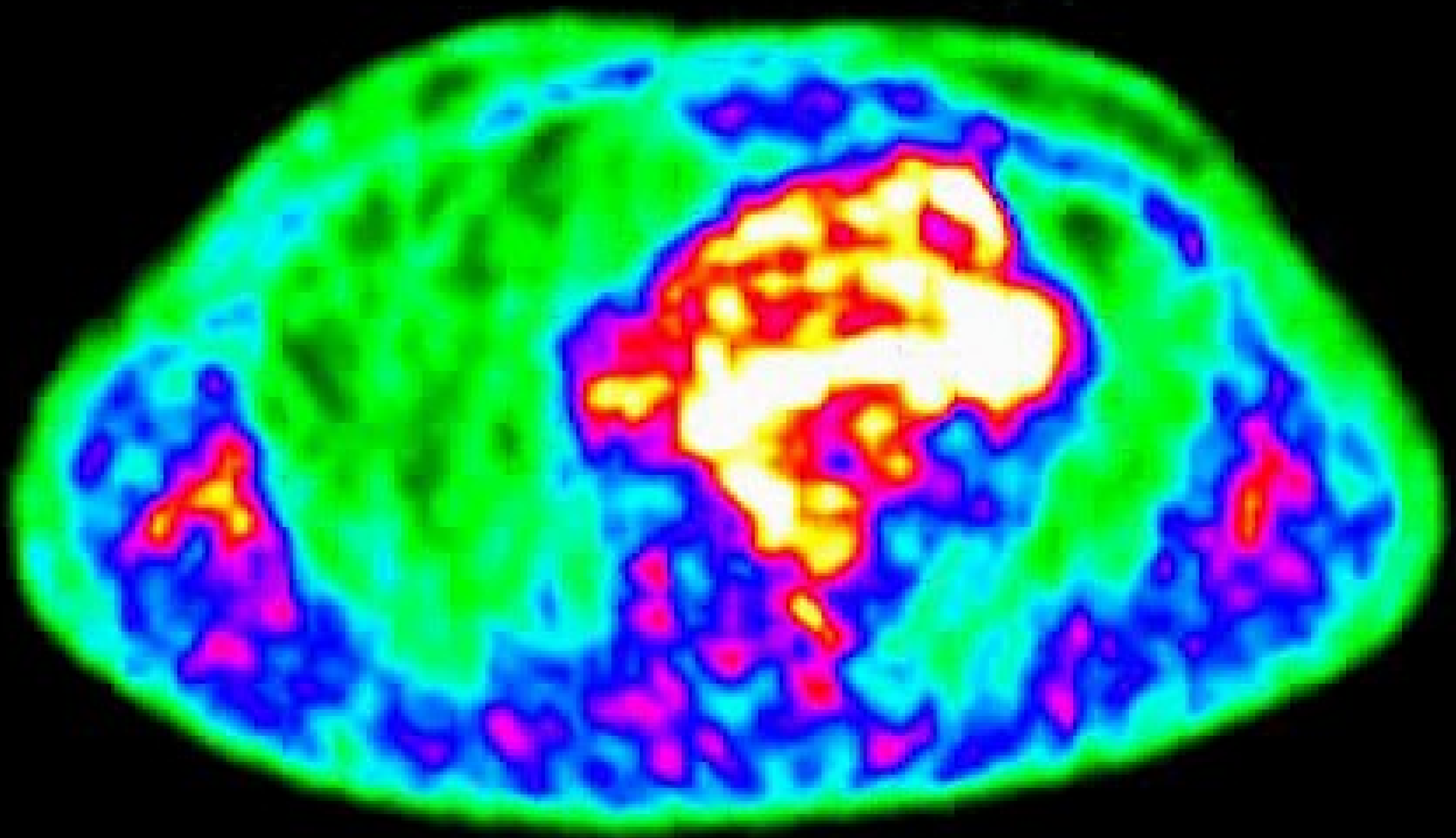
K. Chang et al.: Automatic assessment of glioma burden: a deep learning algorithm for fully automated volumetric and bidimensional measurement: Neuro-Oncology, Volume 21, Issue 11, November 2019, Pages 1412–1422, https://doi.org/10.1093/neuonc/noz106

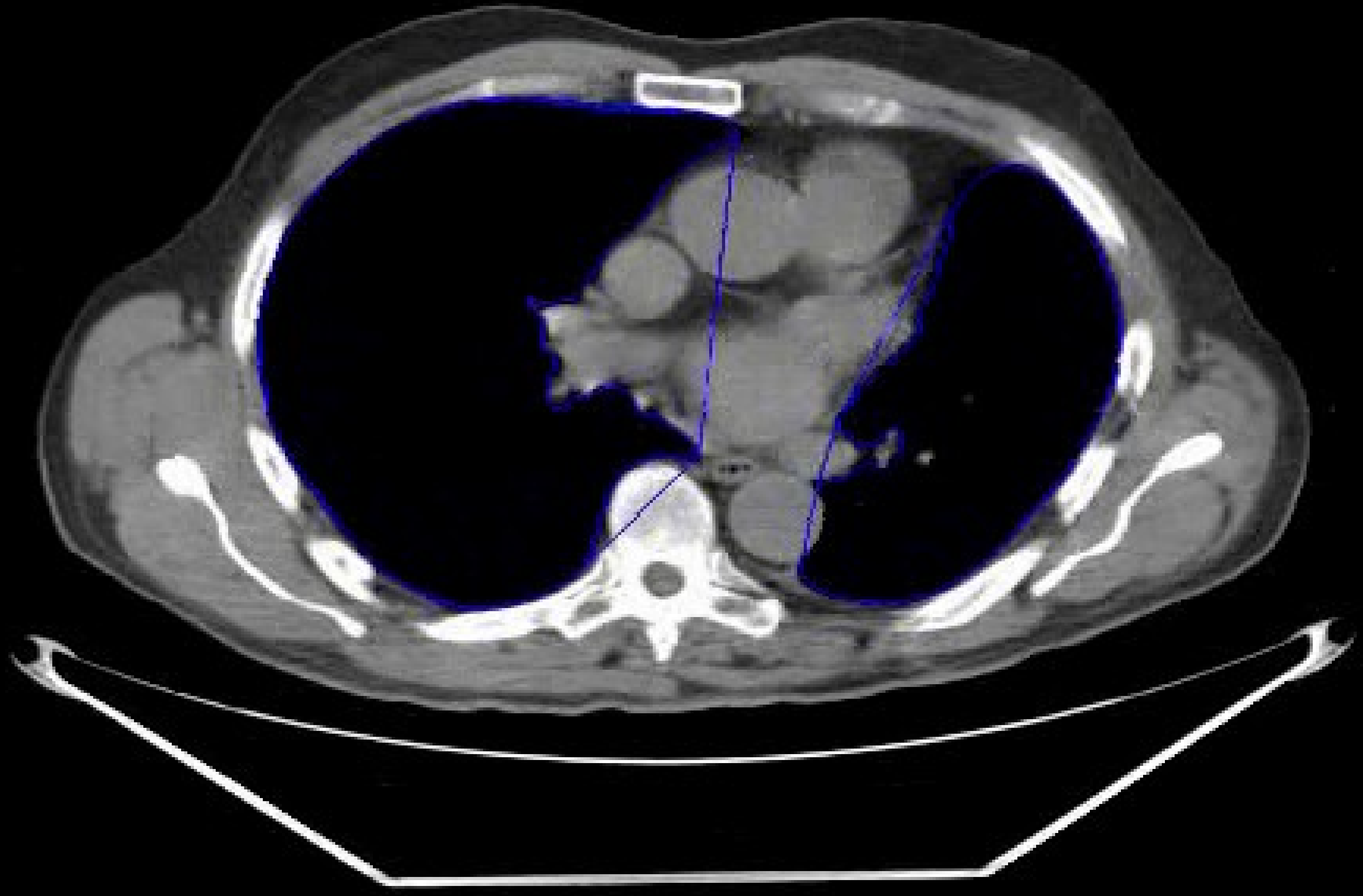
CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (LUNG PET/CT)

PET **CT** **Segmented lesion**

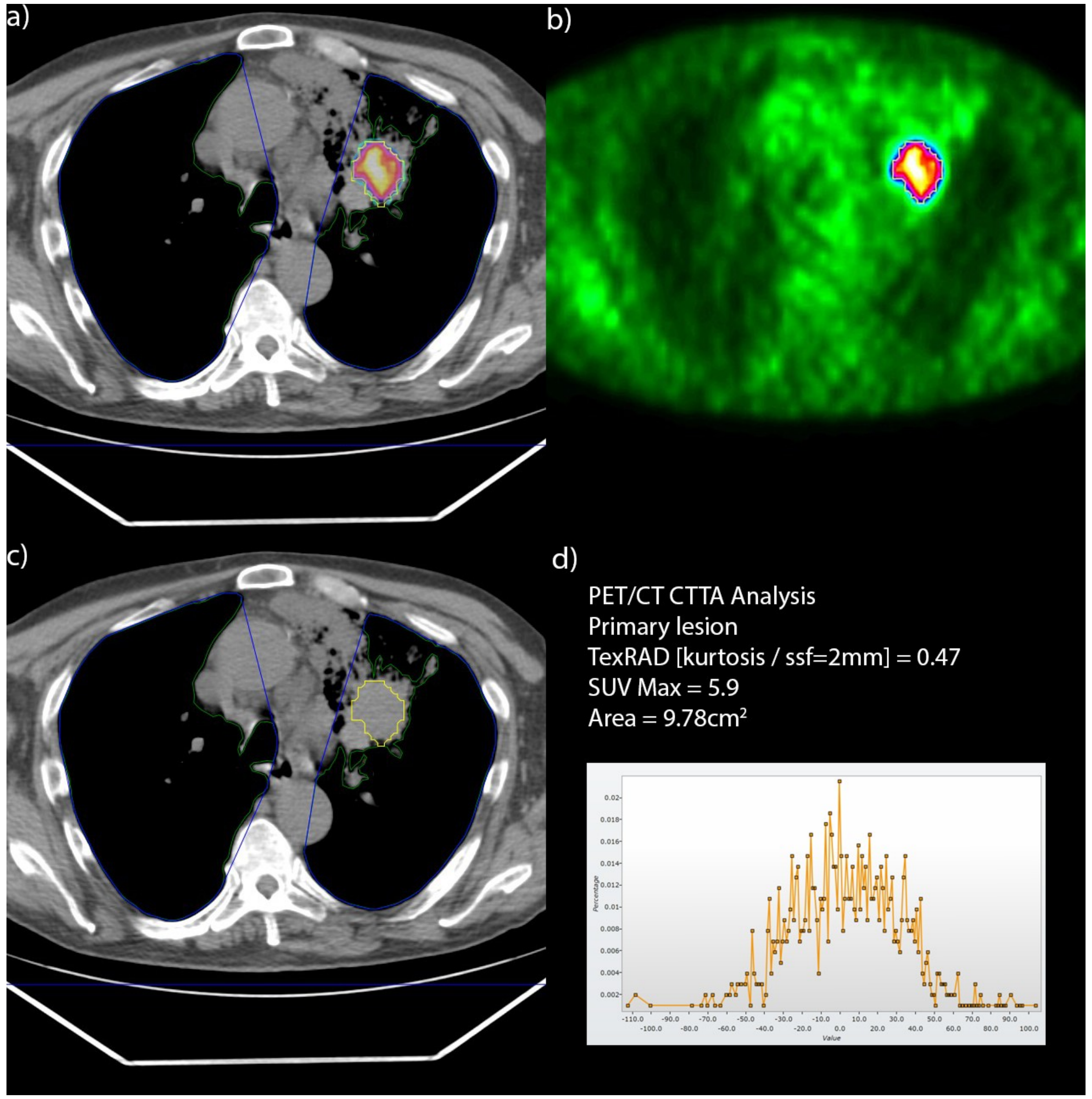


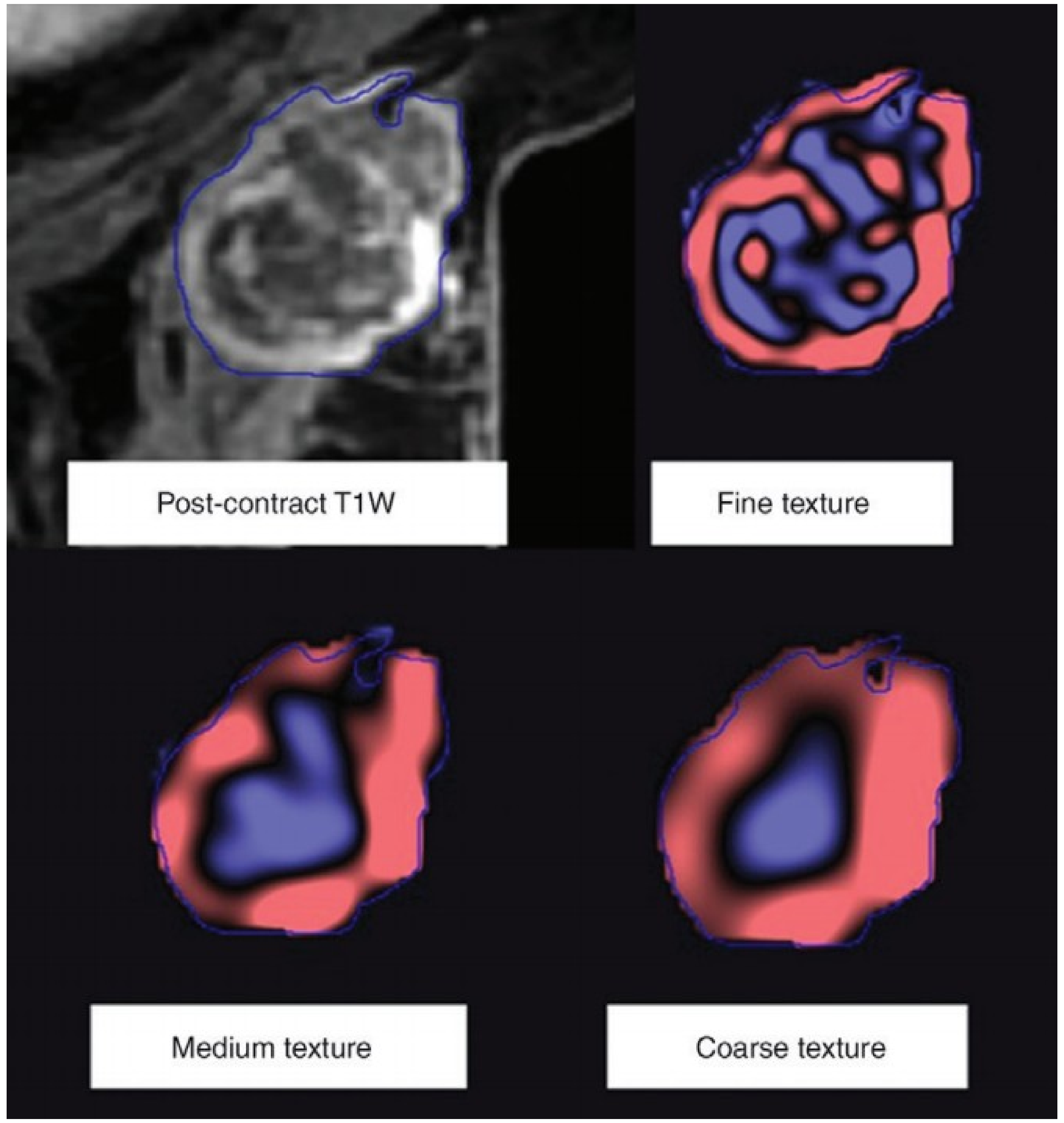
K. Pawelczyk, M. Kawulok, J. Nalepa, M. P. Hayball, S. J. McQuaid, V. Prakash, B. Ganeshan: Towards Detecting High-Uptake Lesions from Lung CT Scans Using Deep Learning. ICIAP (2) 2017: 310-320





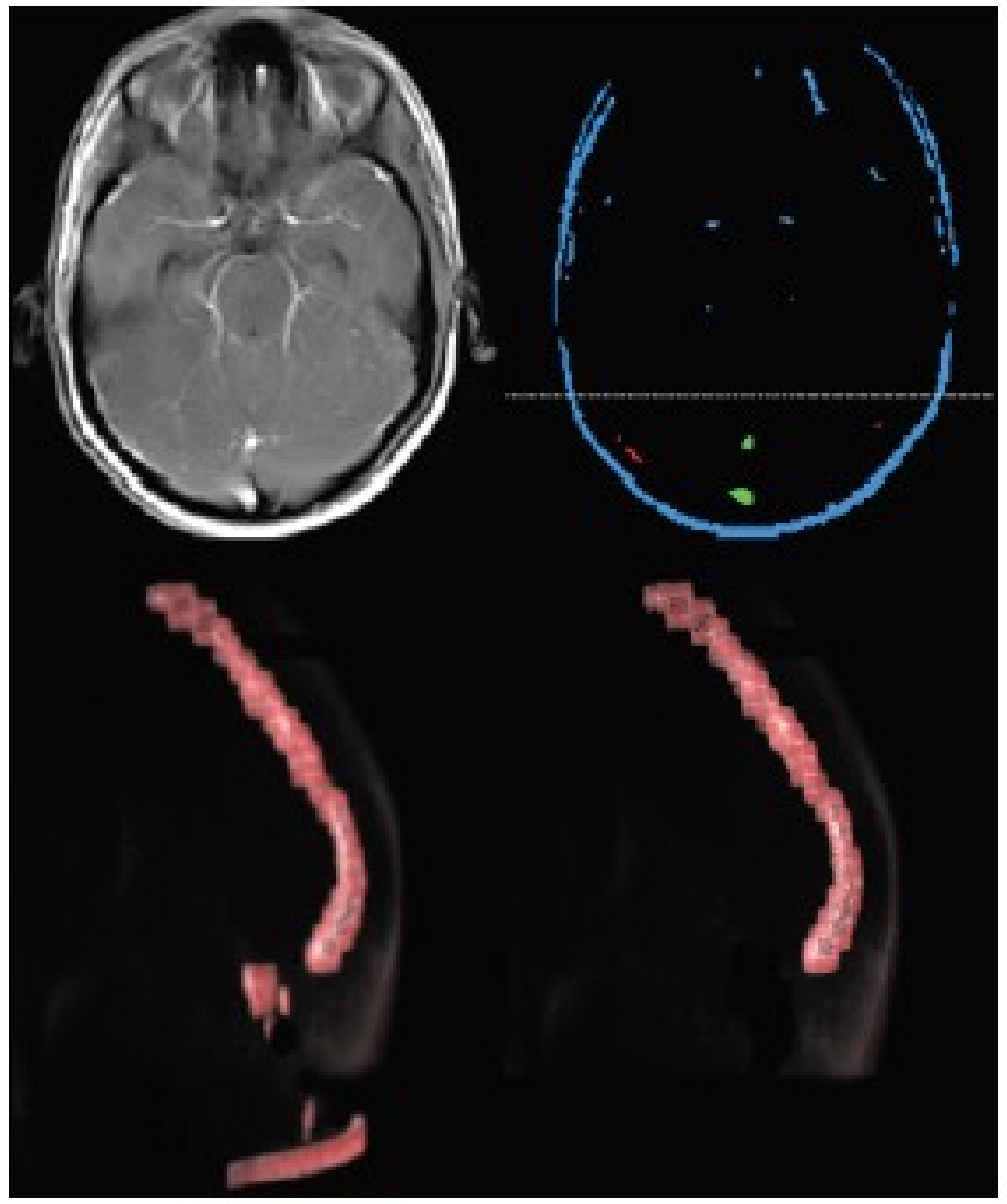
CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (LUNG PET/CT)

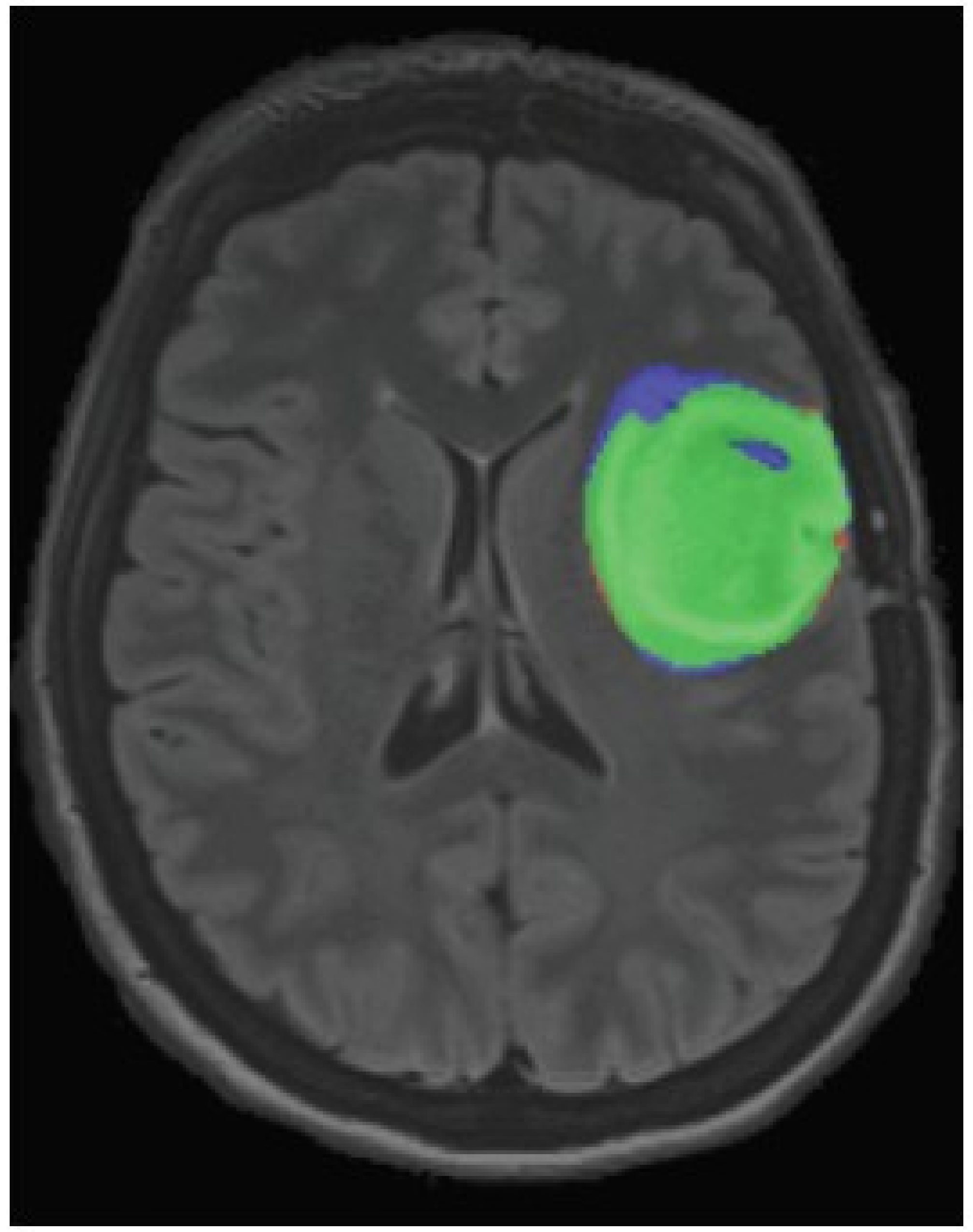


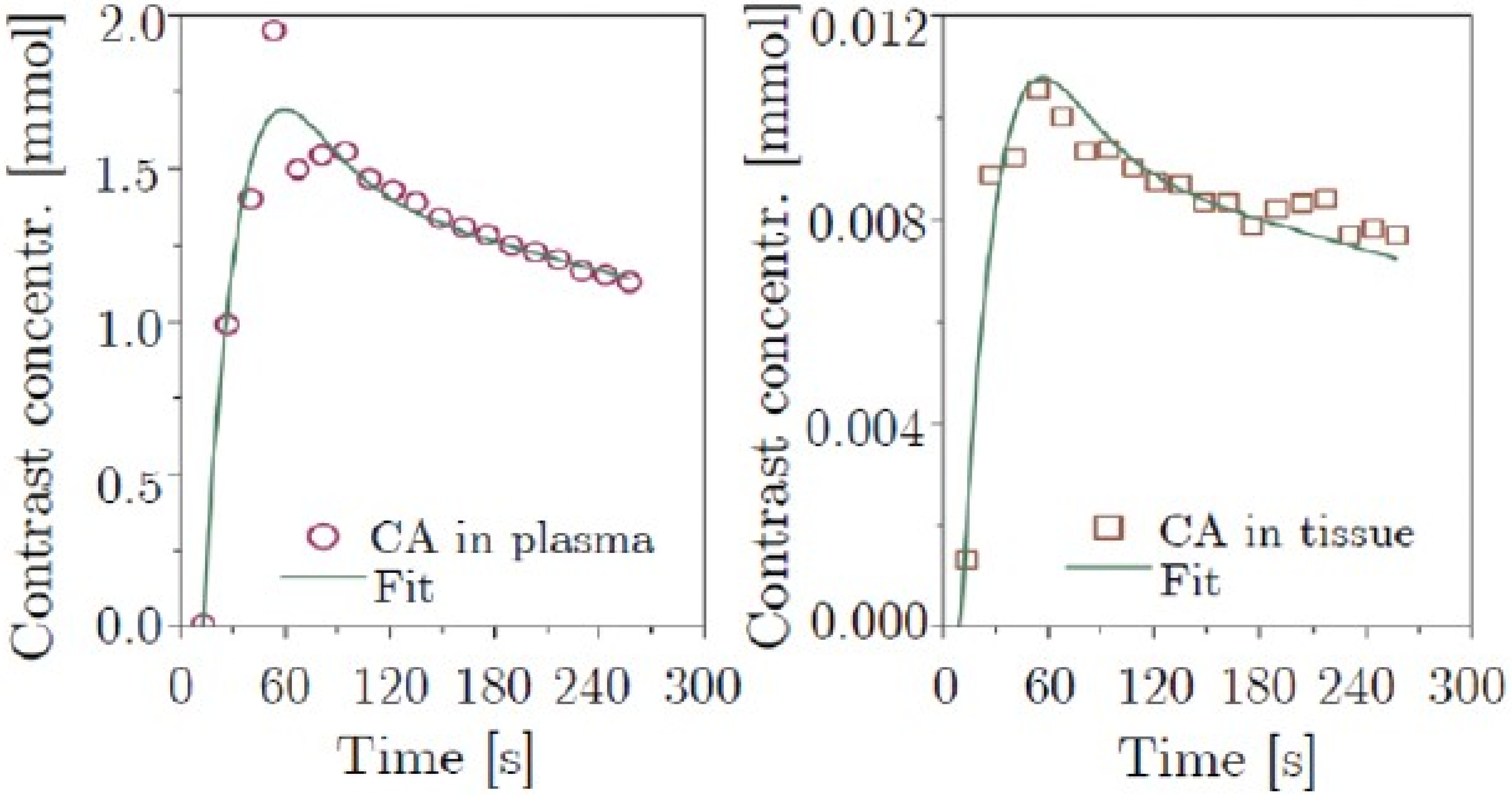


J. Nalepa et al: PET/CT in Lung Cancer: An Automated Imaging Tool for Decision Support, RSNA 2016. Left image: a) CT, b) PET, c) lung and lesion segmentation, and d) visualized textural features. Right image: textural features in breast imaging (Chamming et al., RSNA 2017).

CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (BRAIN DCE-MRI)





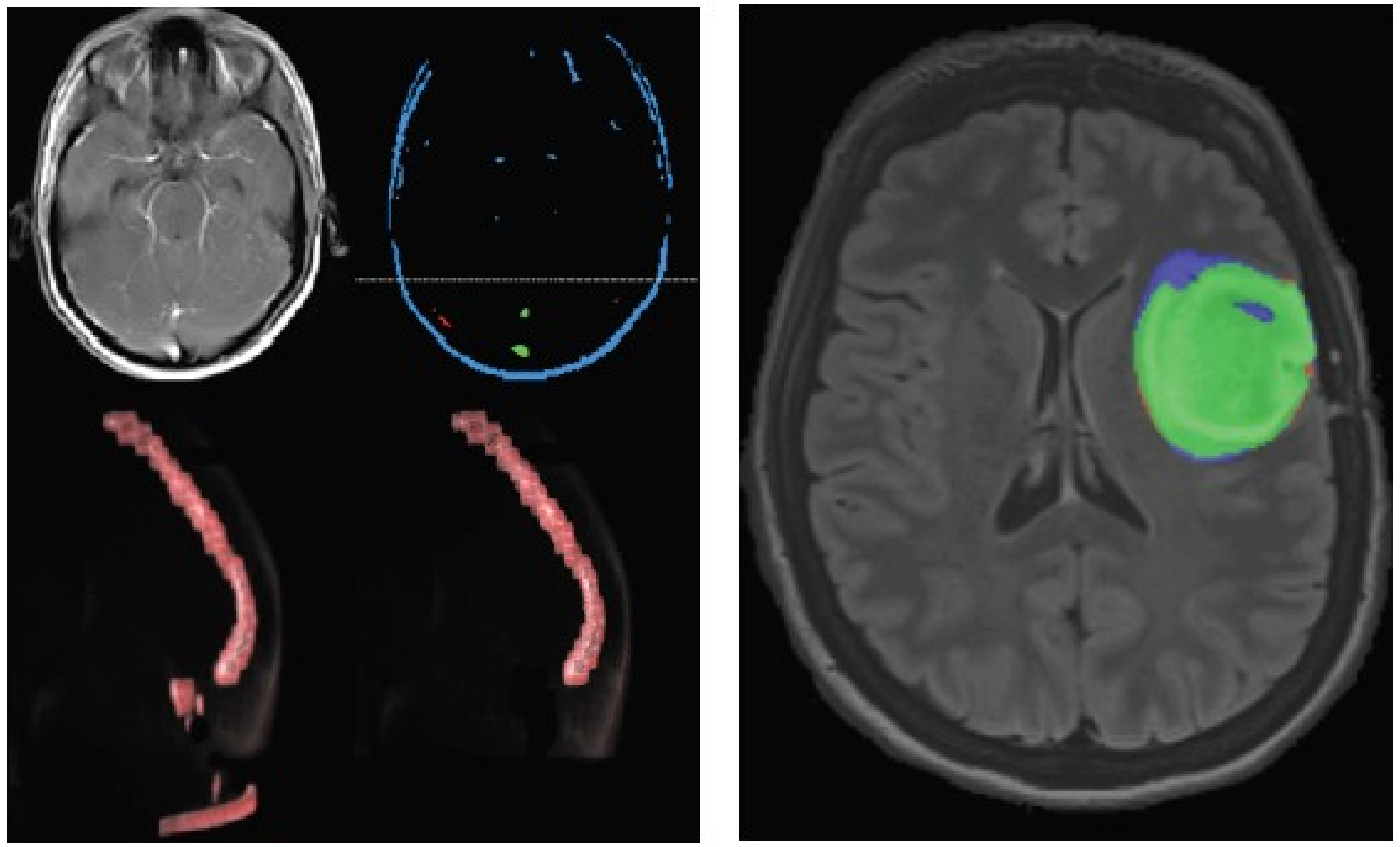


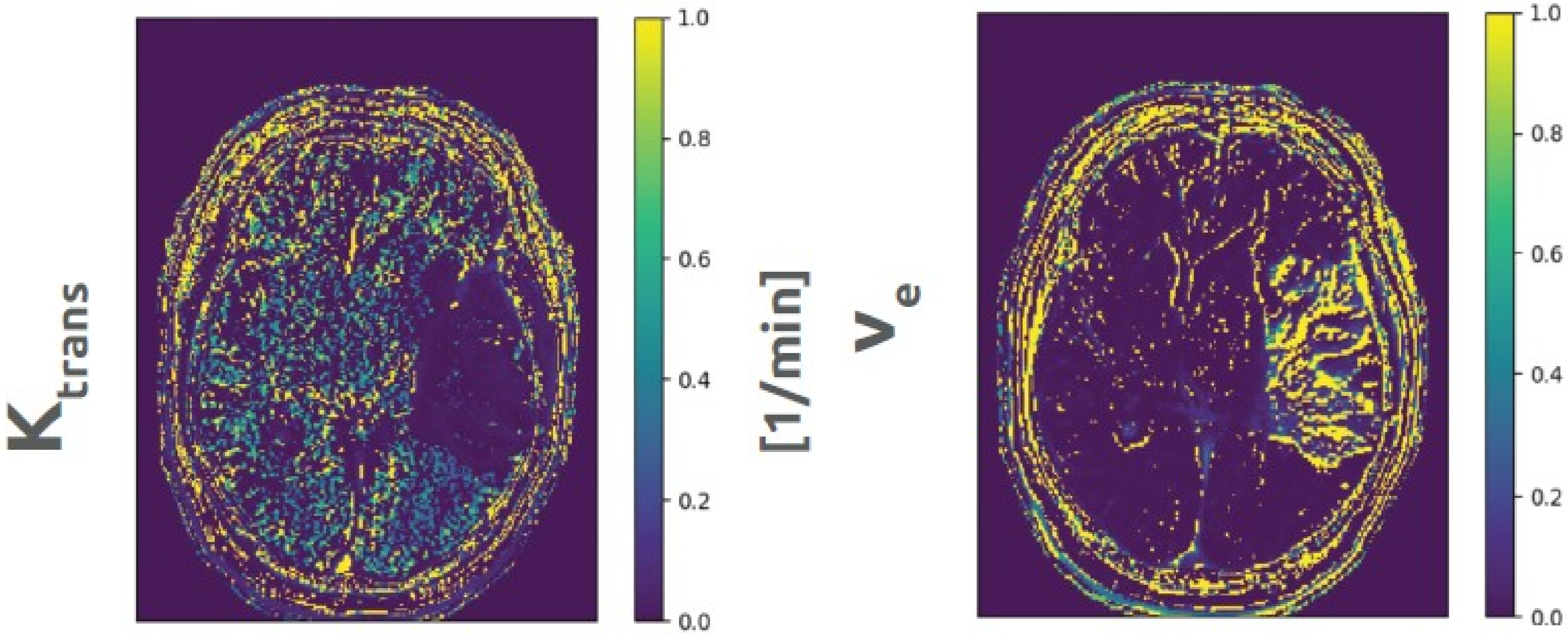
J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

RSNA® Radiological Society of North America

ESR EUROPEAN SOCIETY OF RADIOLOGY

CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (BRAIN DCE-MRI)



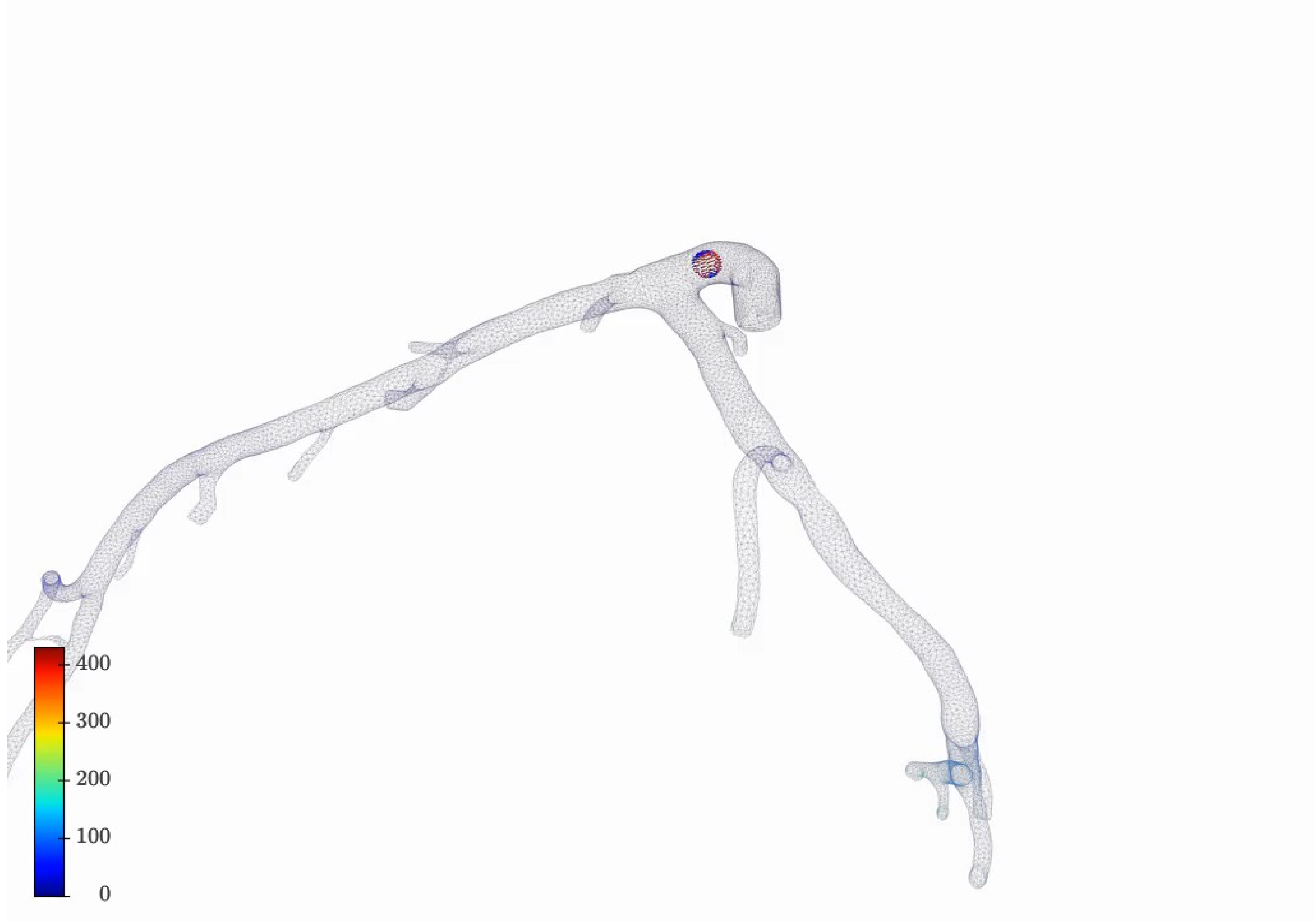


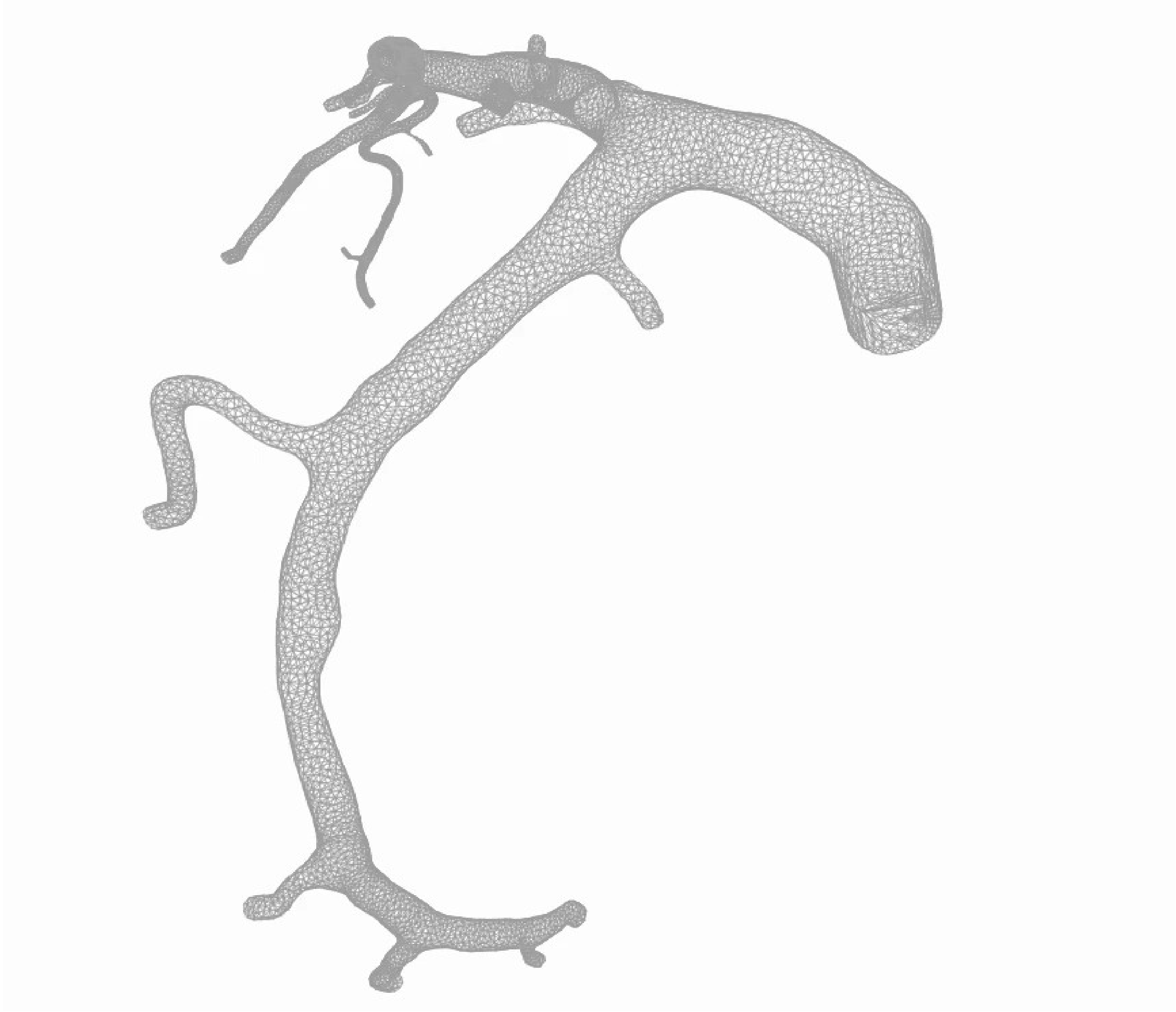
J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

RSNA® Radiological Society of North America

ESR EUROPEAN SOCIETY OF RADIOLOGY

CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (CORONARY CT ANGIOGRAPHY)

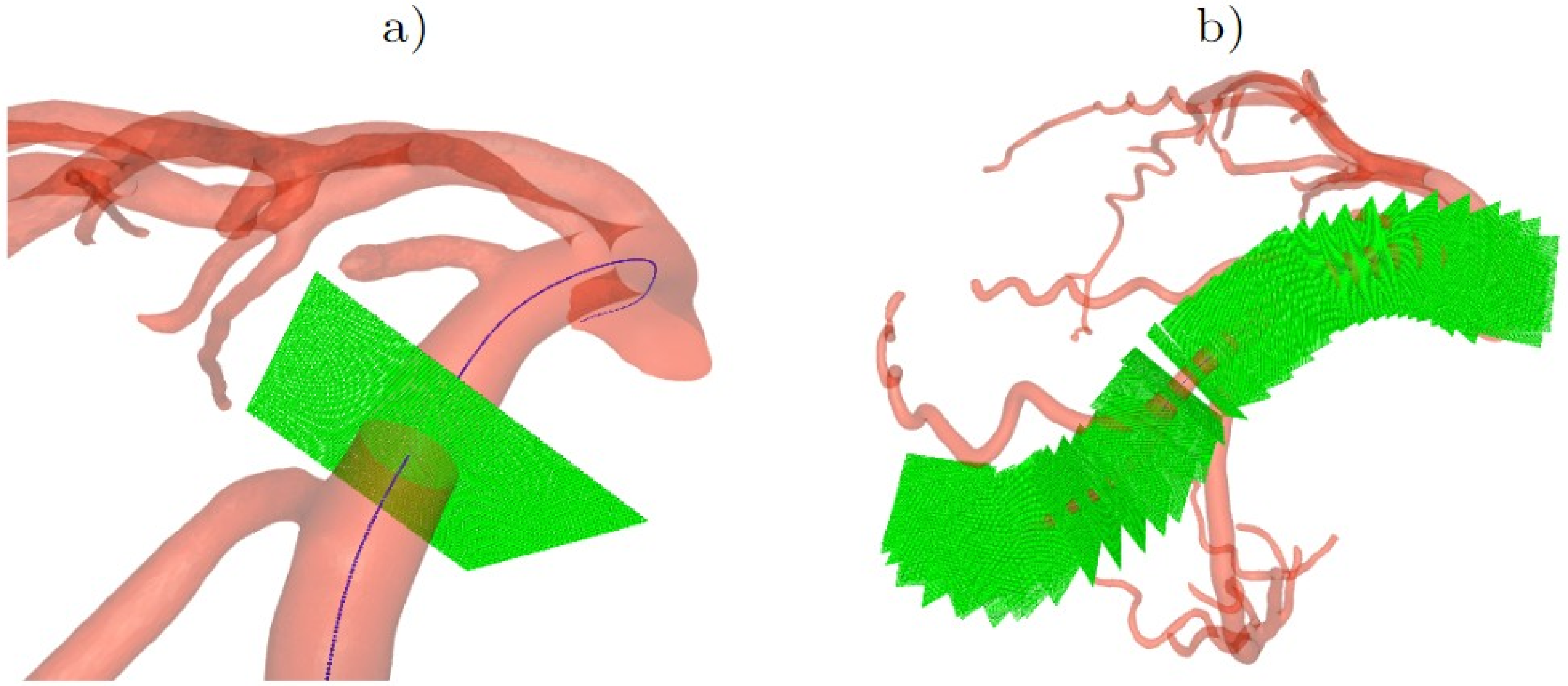


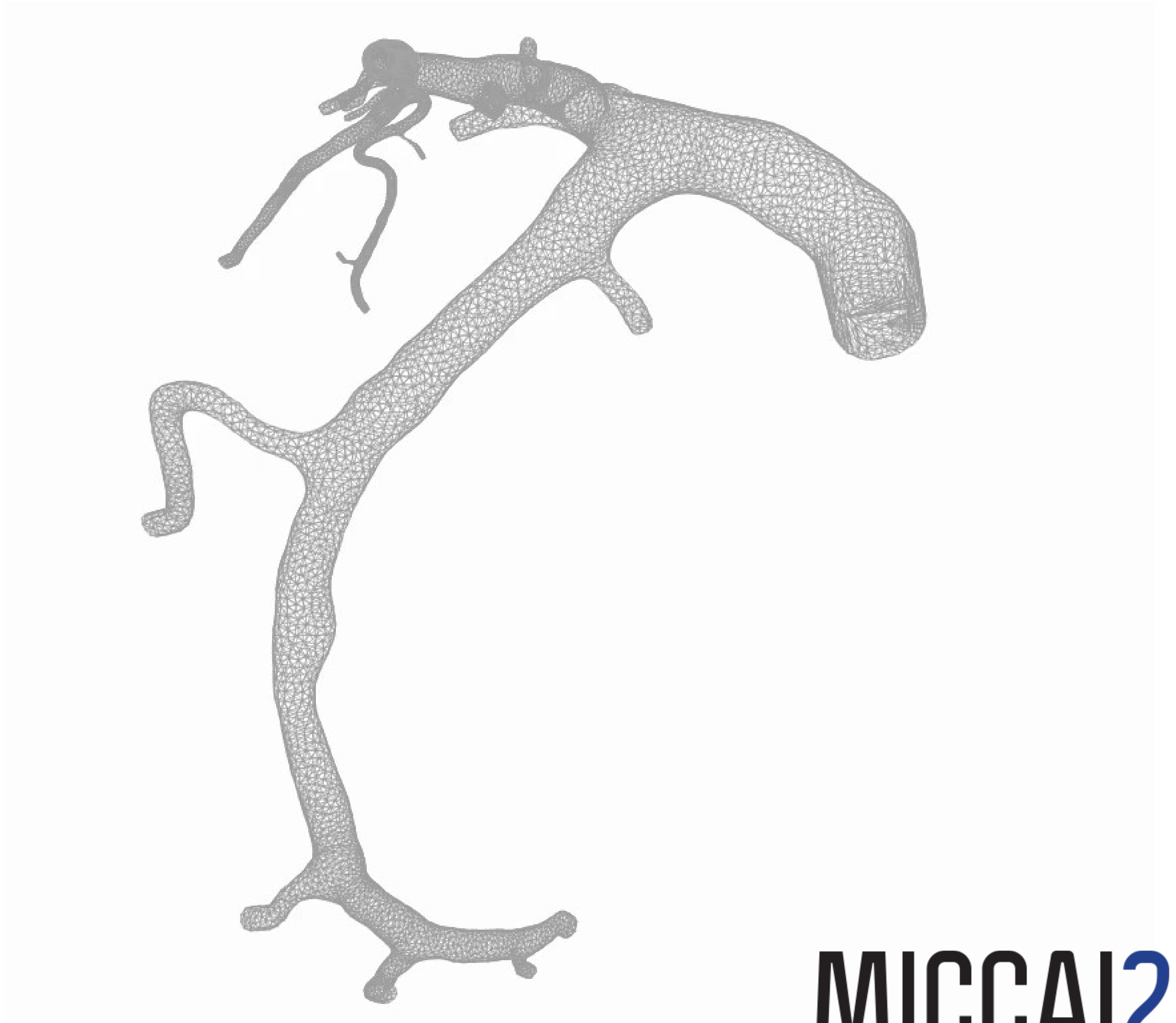


F. Malawski et int J. Nalepa.: Deep Learning Meets Computational Fluid Dynamics to Assess CAD in CCTA. AMAI@MICCAI 2022 (in press).

CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (CORONARY CT ANGIOGRAPHY)

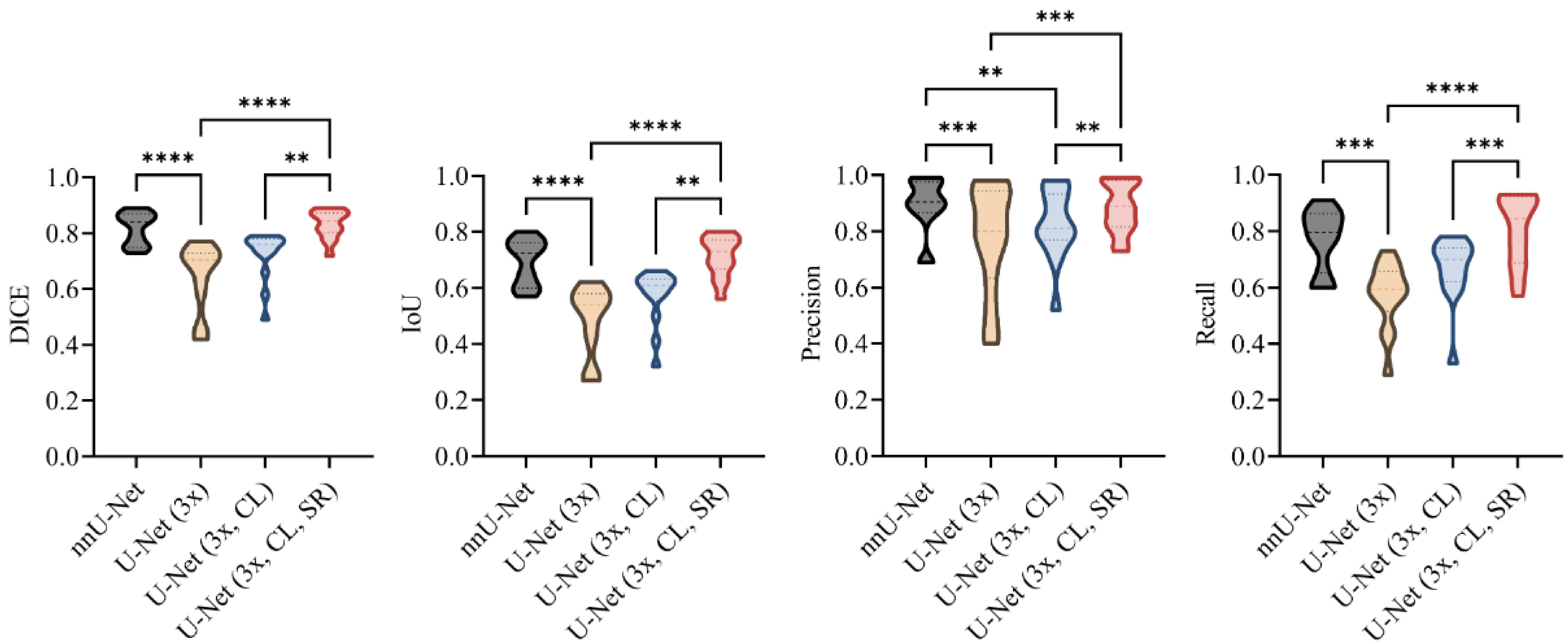
Straigthened representation of a vessel, a) a single 2D plane, together with a subset of b) all 2D planes.





F. Malawski et int J. Nalepa.: Deep Learning Meets Computational Fluid Dynamics to Assess CAD in CCTA. AMAI@MICCAI 2022 (in press).

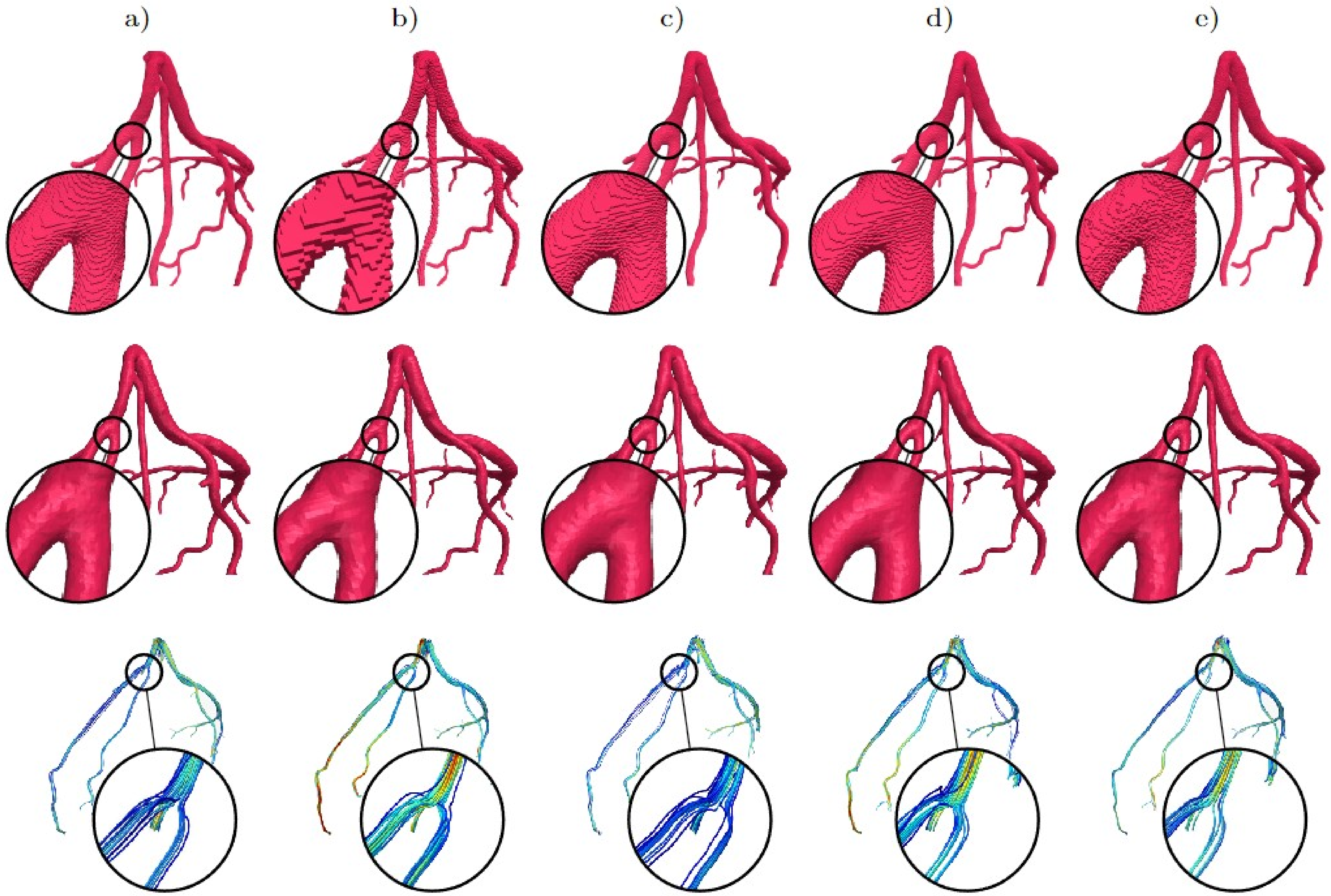
CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (CORONARY CT ANGIOGRAPHY)



MICCAI2022 Singapore

F. Malawski et int J. Nalepa.: Deep Learning Meets Computational Fluid Dynamics to Assess CAD in CCTA. AMAI@MICCAI 2022 (in press).

CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (CORONARY CT ANGIOGRAPHY)



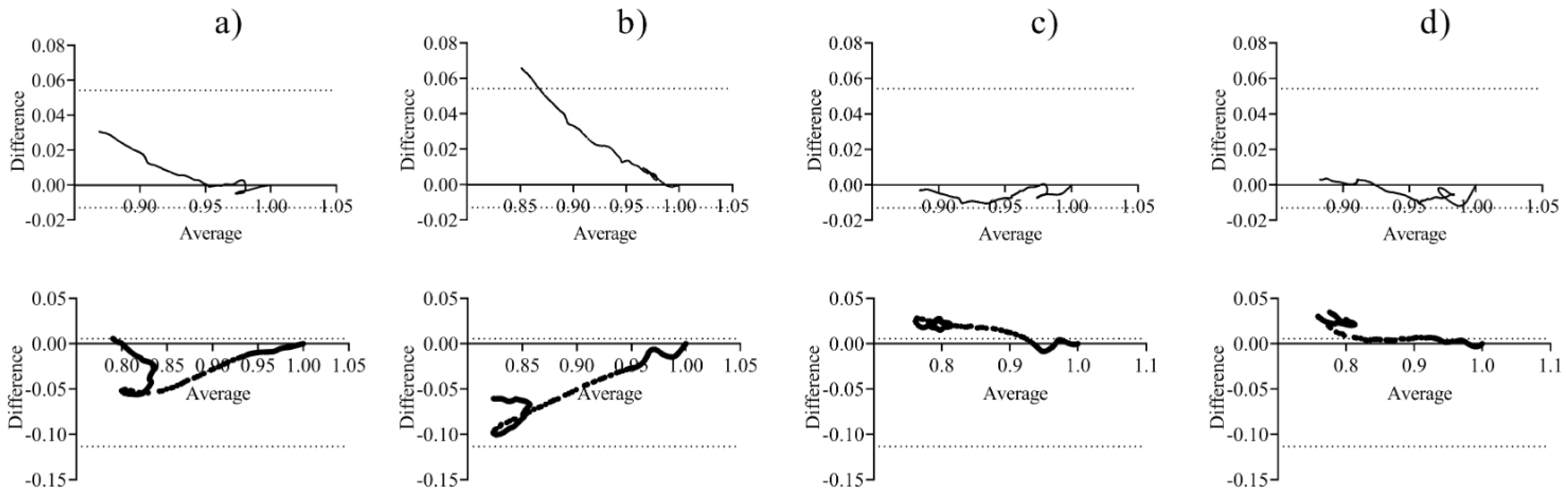
Example
a) ground truth,
b) nnU-Net (DICE: 0.86),
c) U-Net (3×) (DICE: 0.62),
d) U-Net (3×, CL) (DICE: 0.76),
e) U-Net (3×, CL, SR) (DICE: 0.89)
segmentations.

MICCAI2022 Singapore

F. Malawski et int J. Nalepa.: Deep Learning Meets Computational Fluid Dynamics to Assess CAD in CCTA. AMAI@MICCAI 2022 (in press).

CAN AI MAKE US SEE BEYOND THE VISIBLE: EXAMPLES (CORONARY CT ANGIOGRAPHY)

The disagreement for Patient A (top) and Patient B (bottom) between the blood flow parameters obtained for the ground truth and automated segmentation by
a) nnU-Net,
b) U-Net (3×),
c) U-Net (3×, CL), and
d) U-Net (3×, CL, SR).



F. Malawski et int J. Nalepa.: Deep Learning Meets Computational Fluid Dynamics to Assess CAD in CCTA. AMAI@MICCAI 2022 (in press).

nature

Explore our content ⌄ Journal information ⌄

nature › articles › article

Article | Published: 01 January 2020

International evaluation of an AI system for breast cancer screening

Scott Mayer McKinney, Marcin Sieniek, [...] Shravya Shetty

Nature **577**, 89–94(2020) | Cite this article

Matters Arising to this article was published on 14 October 2020

An Addendum to this article was published on 14 October 2020

nature

Explore our content ⌄ Journal information ⌄

nature › articles › article

Article | Published: 01 January 2020

International evaluation of an AI system for breast cancer screening

Scott Mayer McKinney, Marcin Sieniek, [...] Shravya Shetty

Nature **577**, 89–94(2020) | Cite this article

Matters Arising to this article was published on 14 October 2020

An Addendum to this article was published on 14 October 2020

Matters arising

Transparency and reproducibility in artificial intelligence

https://doi.org/10.1038/s41586-020-2766-y

Received: 1 February 2020

Accepted: 10 August 2020

Published online: 14 October 2020

Check for updates

Benjamin Haibe-Kains[1,2,3,4,5] ✉, George Alexandru Adam[3,5], Ahmed Hosny[6,7], Farnoosh Khodakarami[1,2], Massive Analysis Quality Control (MAQC) Society Board of Directors*, Levi Waldron[8], Bo Wang[2,3,5,9,10], Chris McIntosh[2,5,9], Anna Goldenberg[3,5,11,12], Anshul Kundaje[13,14], Casey S. Greene[15,16], Tamara Broderick[17], Michael M. Hoffman[1,2,3,5], Jeffrey T. Leek[18], Keegan Korthauer[19,20], Wolfgang Huber[21], Alvis Brazma[22], Joelle Pineau[23,24], Robert Tibshirani[25,26], Trevor Hastie[25,26], John P. A. Ioannidis[25,26,27,28,29], John Quackenbush[30,31,32] & Hugo J. W. L. Aerts[6,7,33,34]

ARISING FROM S. M. McKinney et al. *Nature* https://doi.org/10.1038/s41586-019-1799-6 (2020)

Leakage and the Reproducibility Crisis in ML-based Science

PRINCETON UNIVERSITY

We argue that there is a reproducibility crisis in ML-based science. We compile evidence of this crisis across fields, identify data leakage as a pervasive cause of reproducibility failures, conduct our own reproducibility investigations using in-depth code-review, and propose a solution.

Draft paper

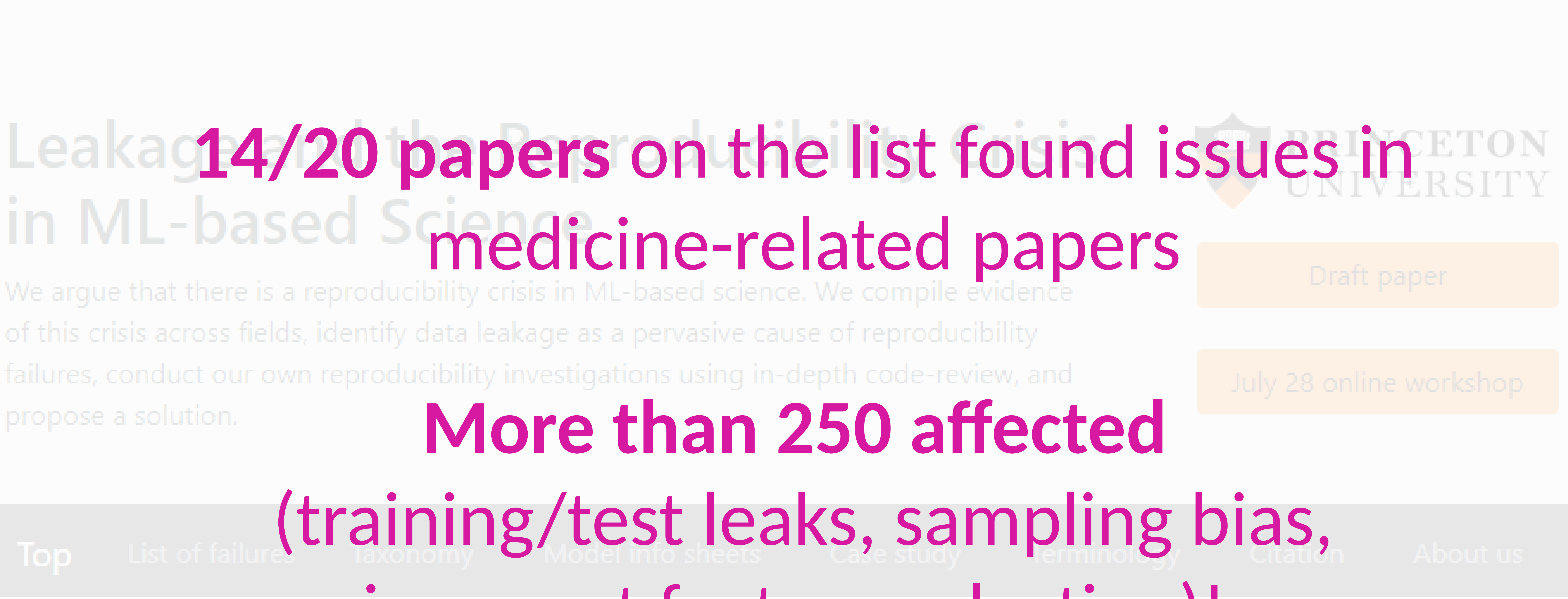
July 28 online workshop

Top | List of failures | Taxonomy | Model info sheets | Case study | Terminology | Citation | About us

14/20 papers on the list found issues in medicine-related papers

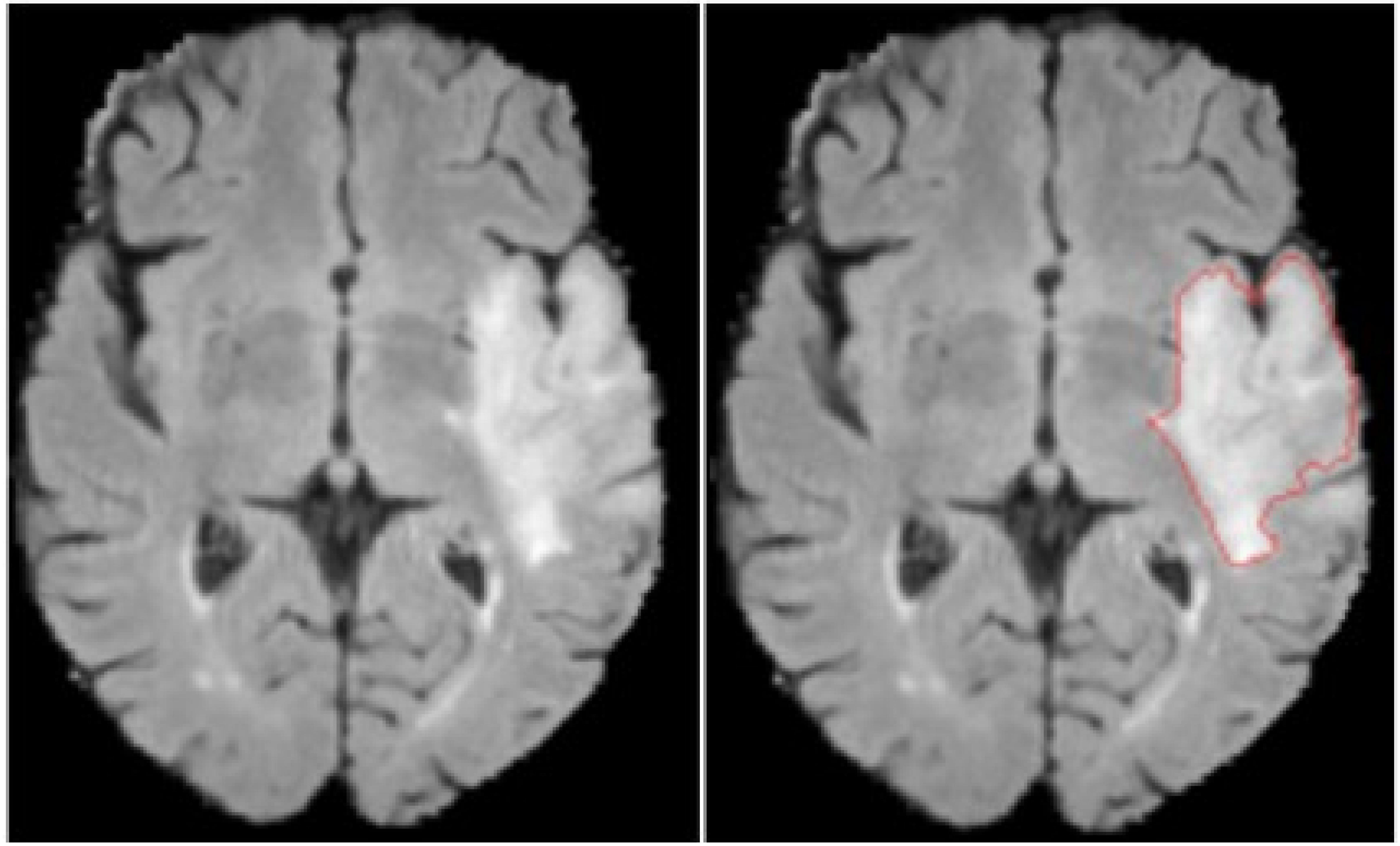
More than 250 affected

(training/test leaks, sampling bias, incorrect feature selection)!



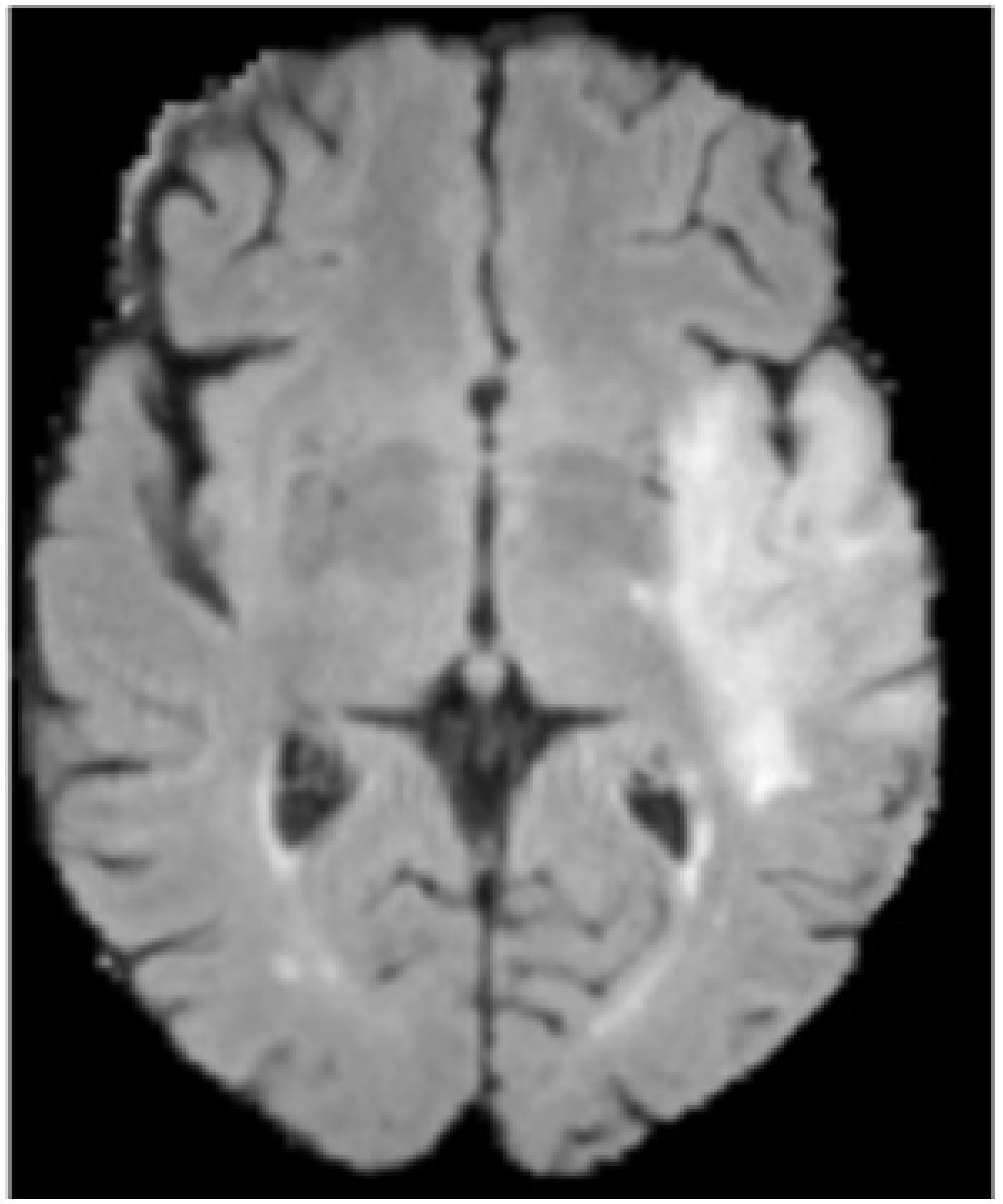
TOWARDS CE-MARKED DEEP LEARNING SOFTWARE FOR BRAIN TUMOR ANALYSIS

Deep learning in medical products

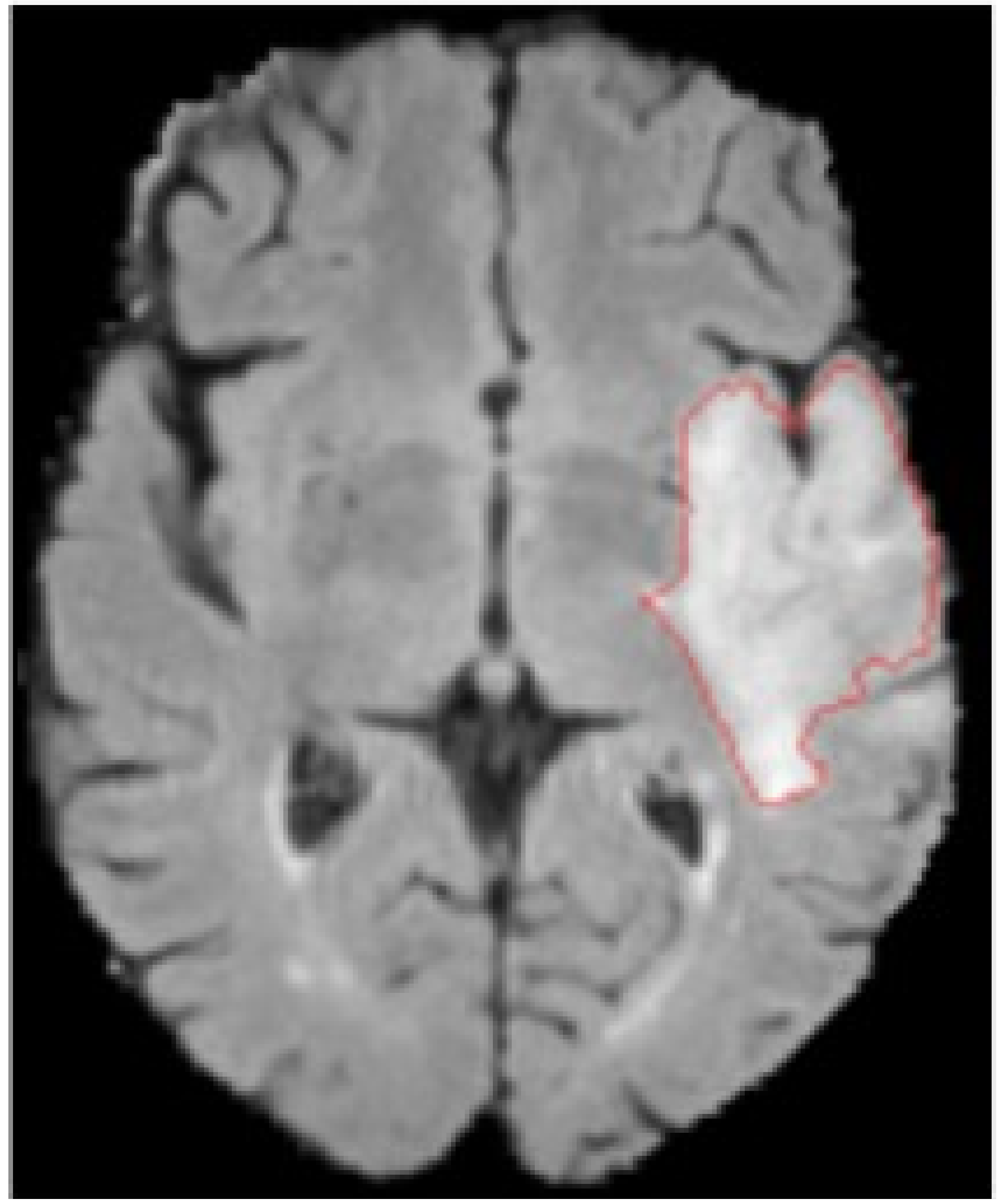


J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

Deep learning in medical products



Sens.Ai

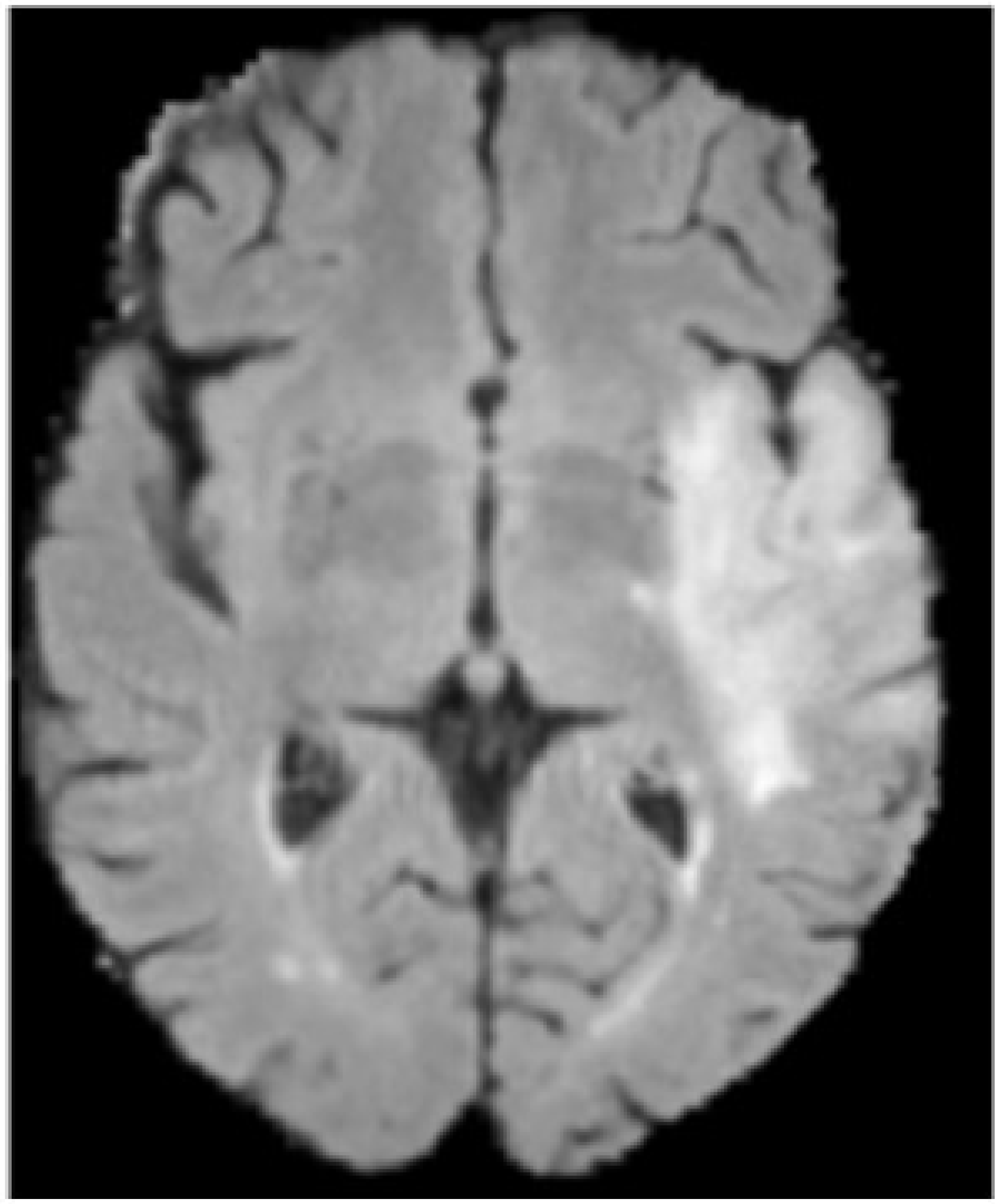


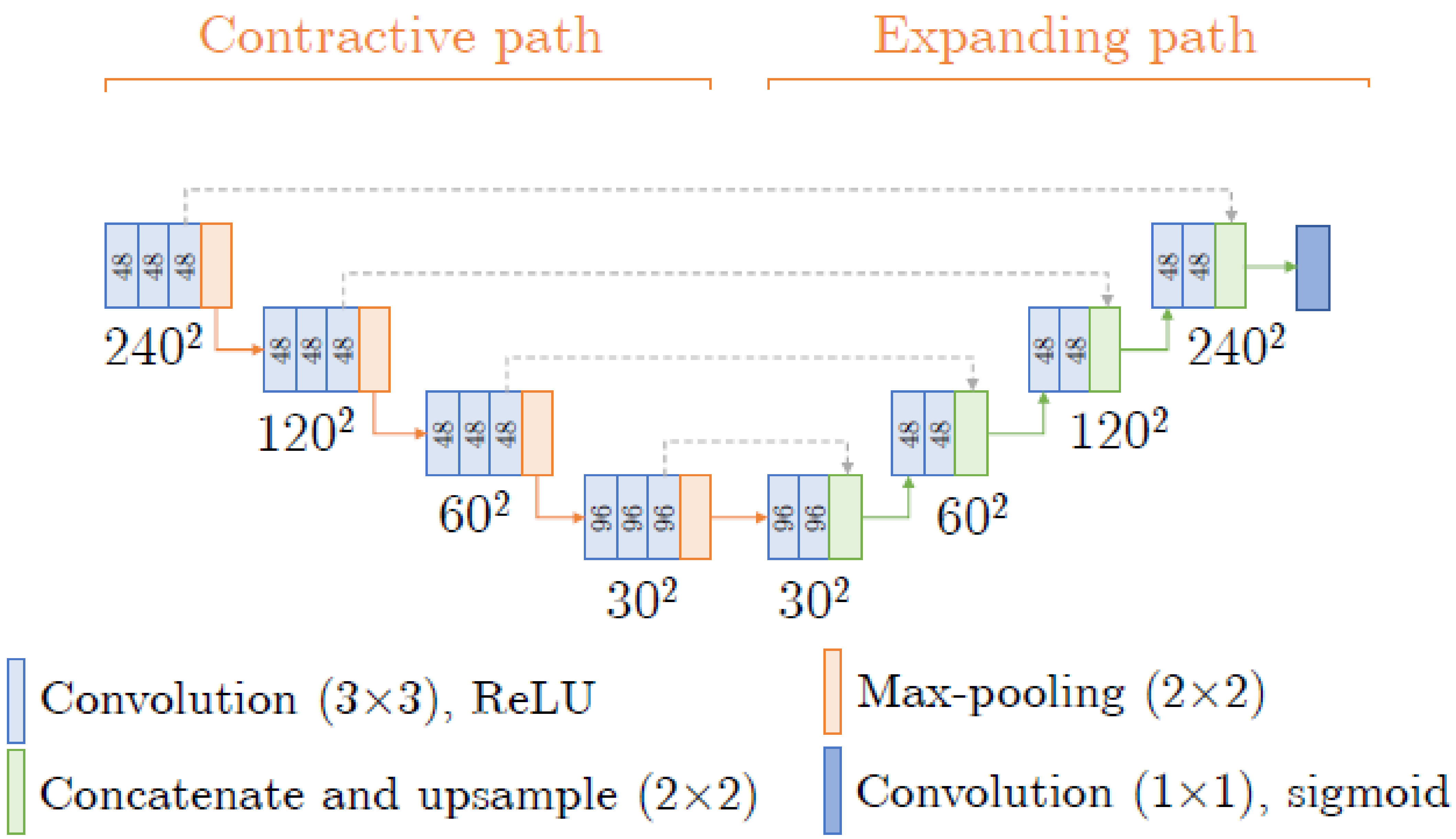
Future Processing HEALTHCARE

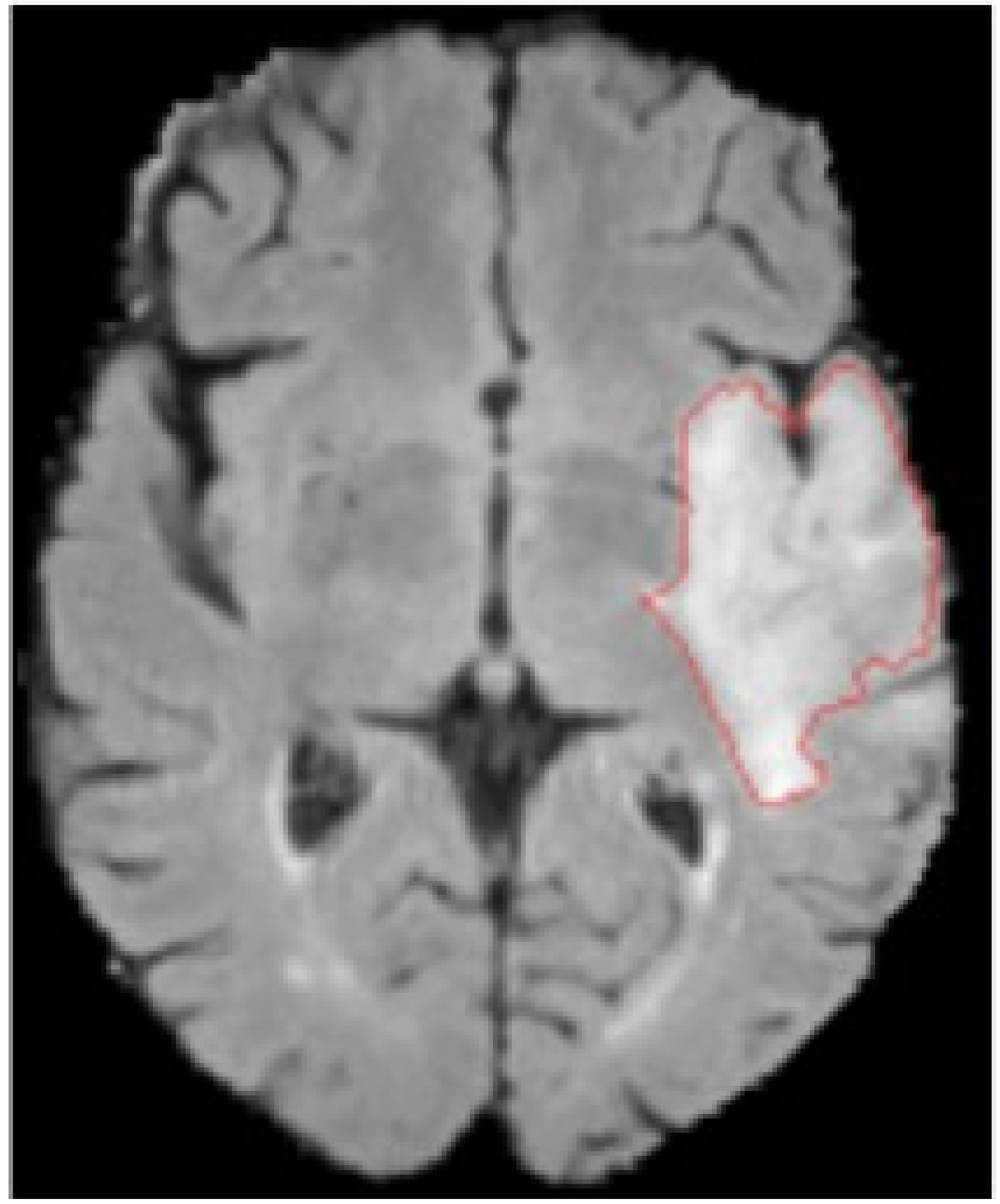
Intended use: This system automatically delineates and calculates the volumetric characteristics of the largest brain lesion (low- and high-grade gliomas) from T2-FLAIR.

J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

Deep learning in medical products

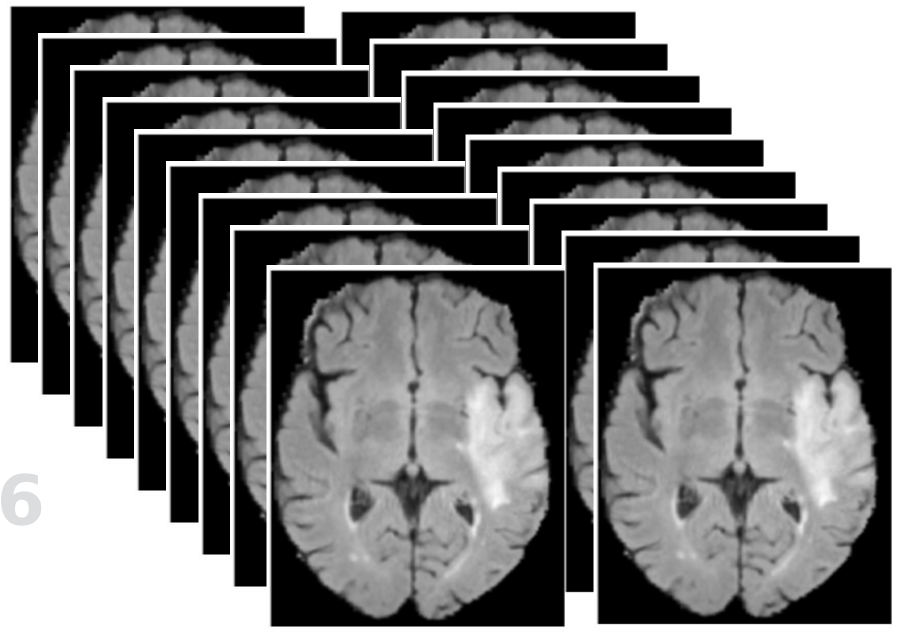




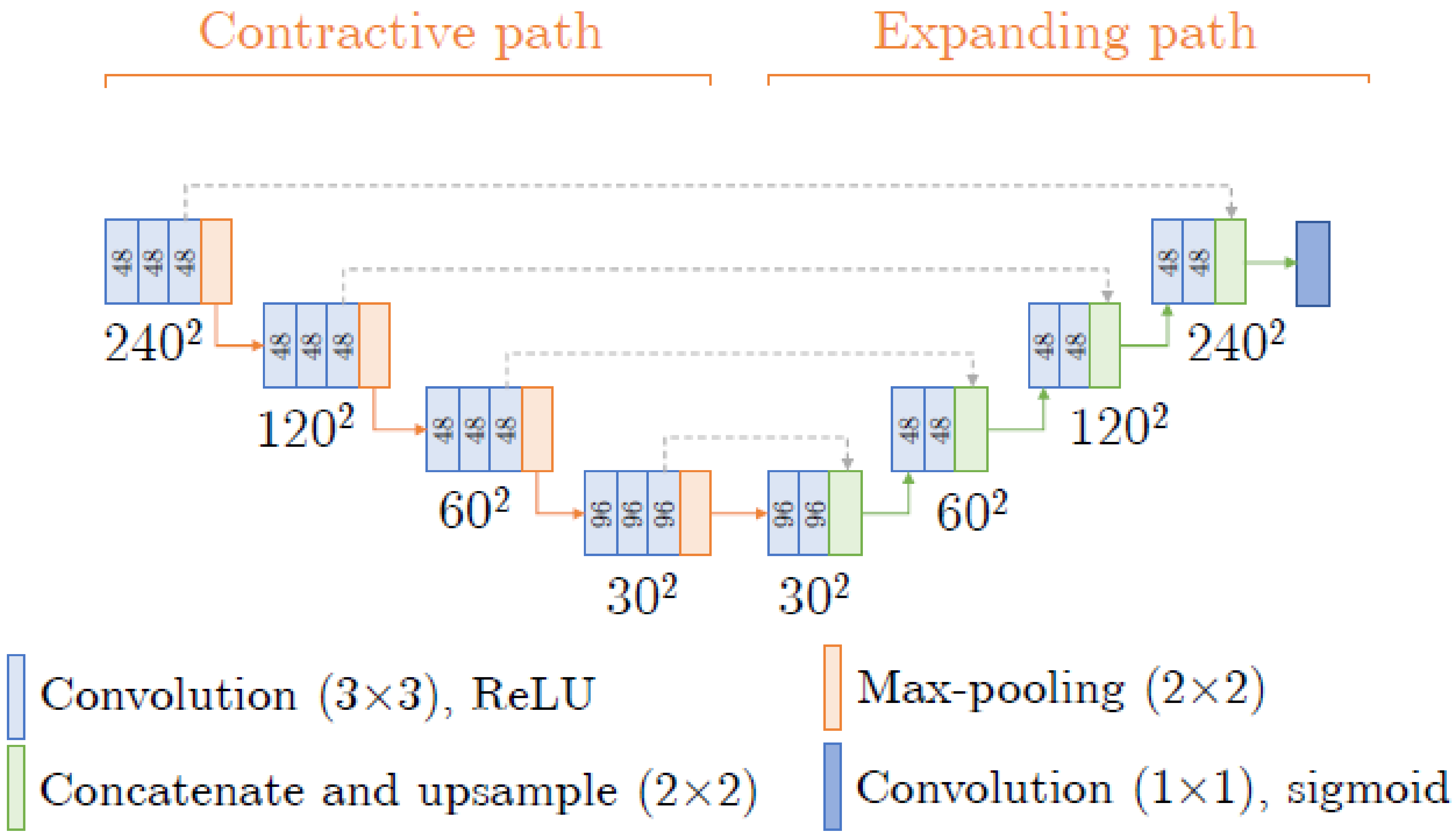


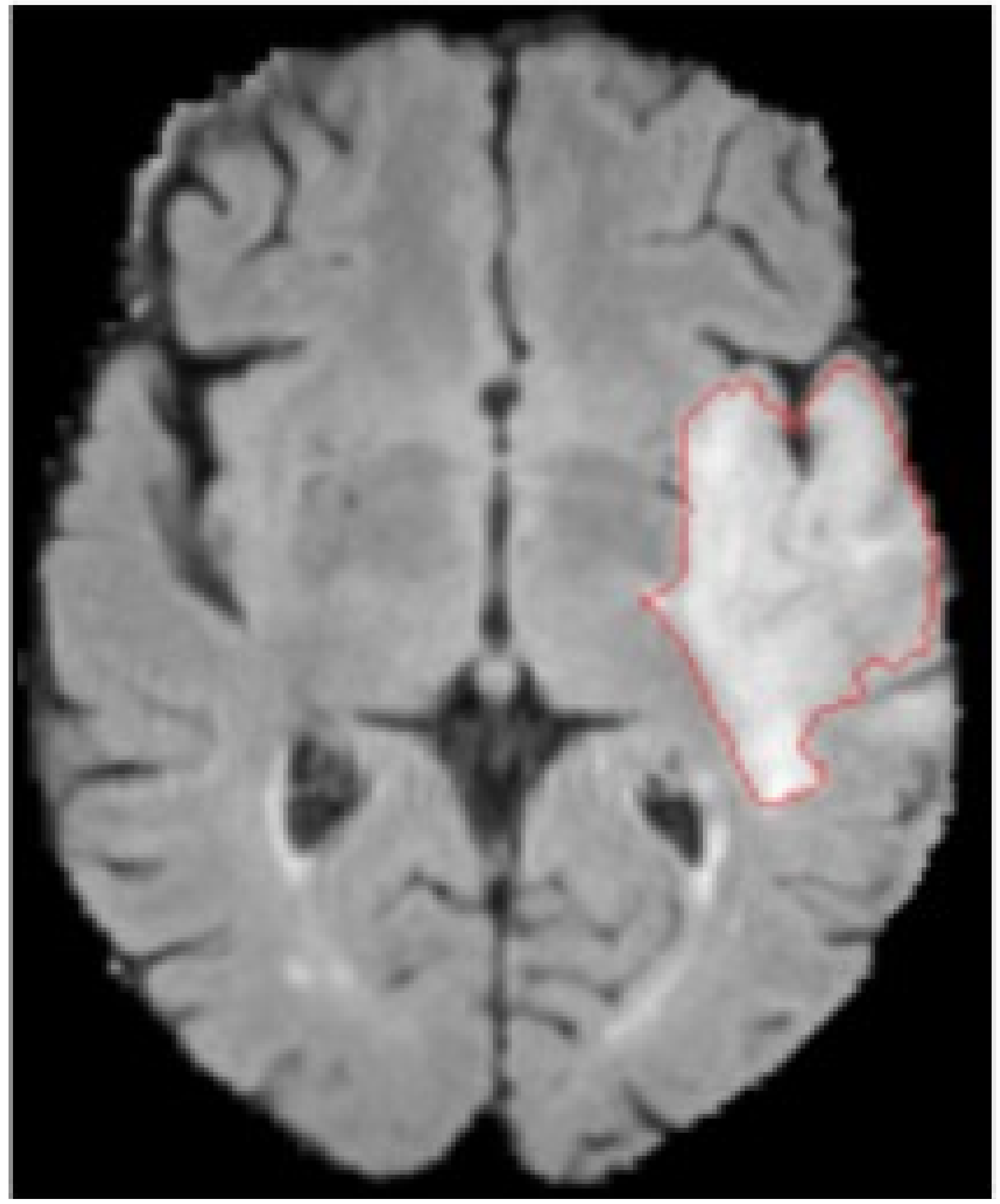
J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

Deep learning in medical products



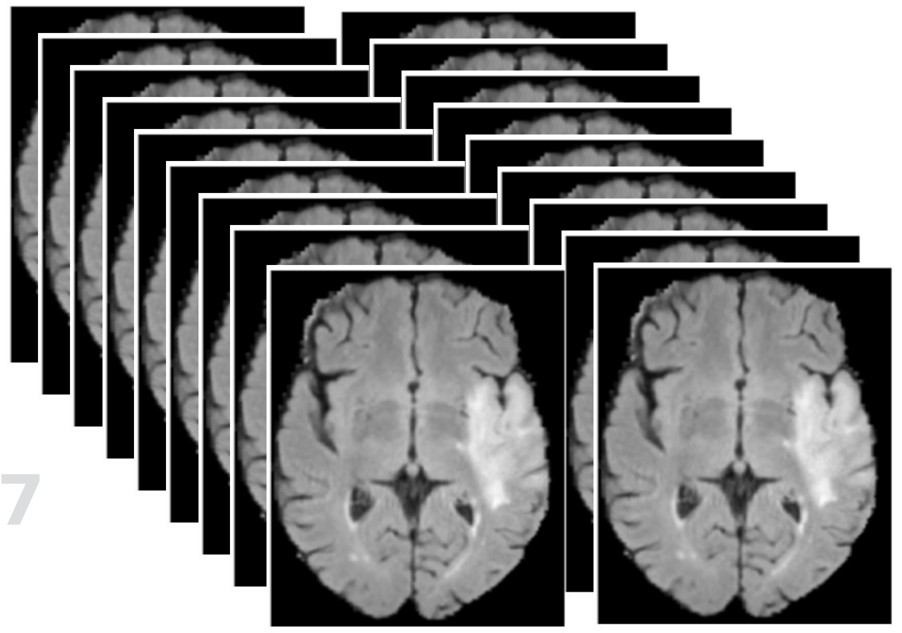
Training set T





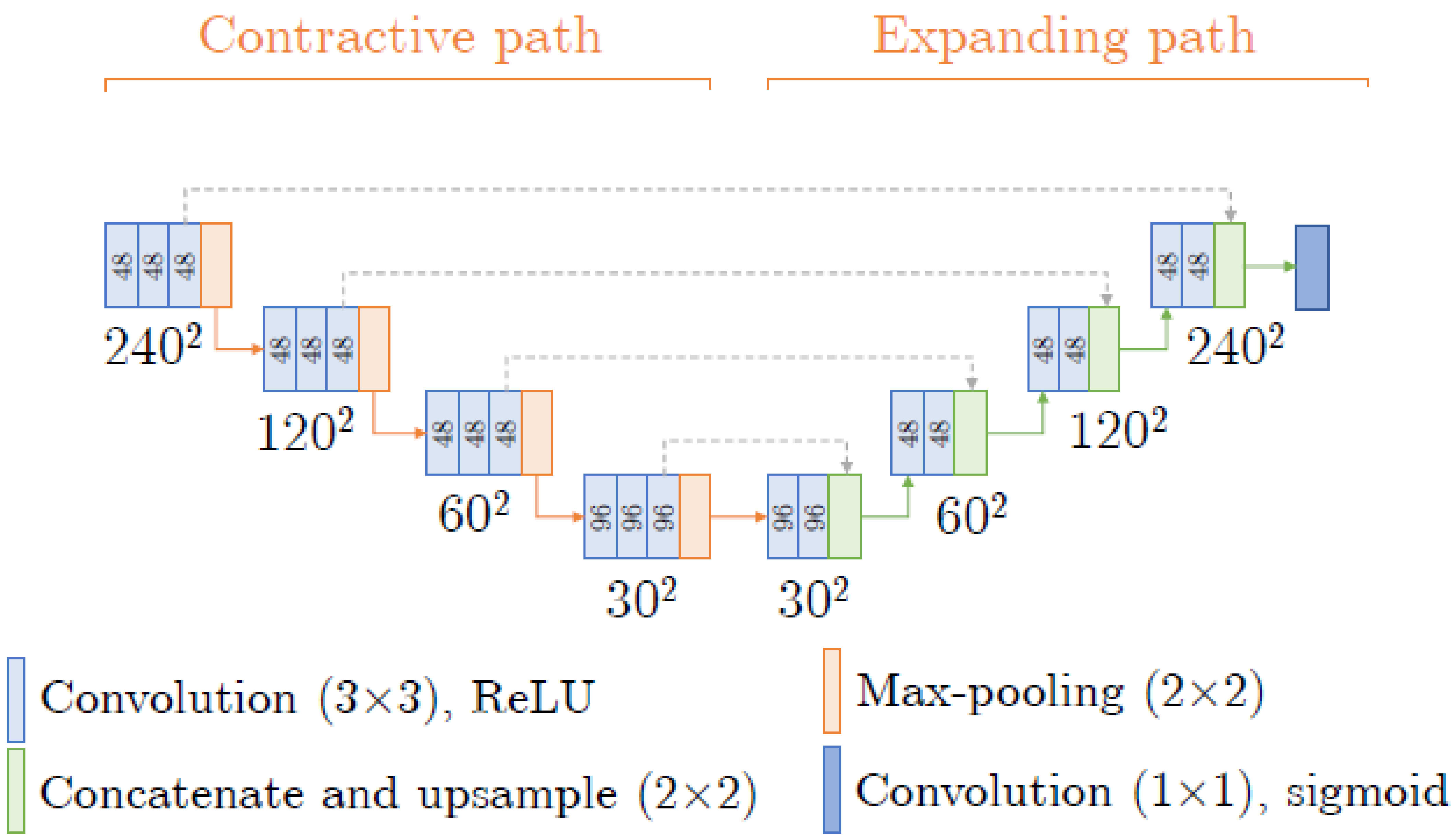
J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

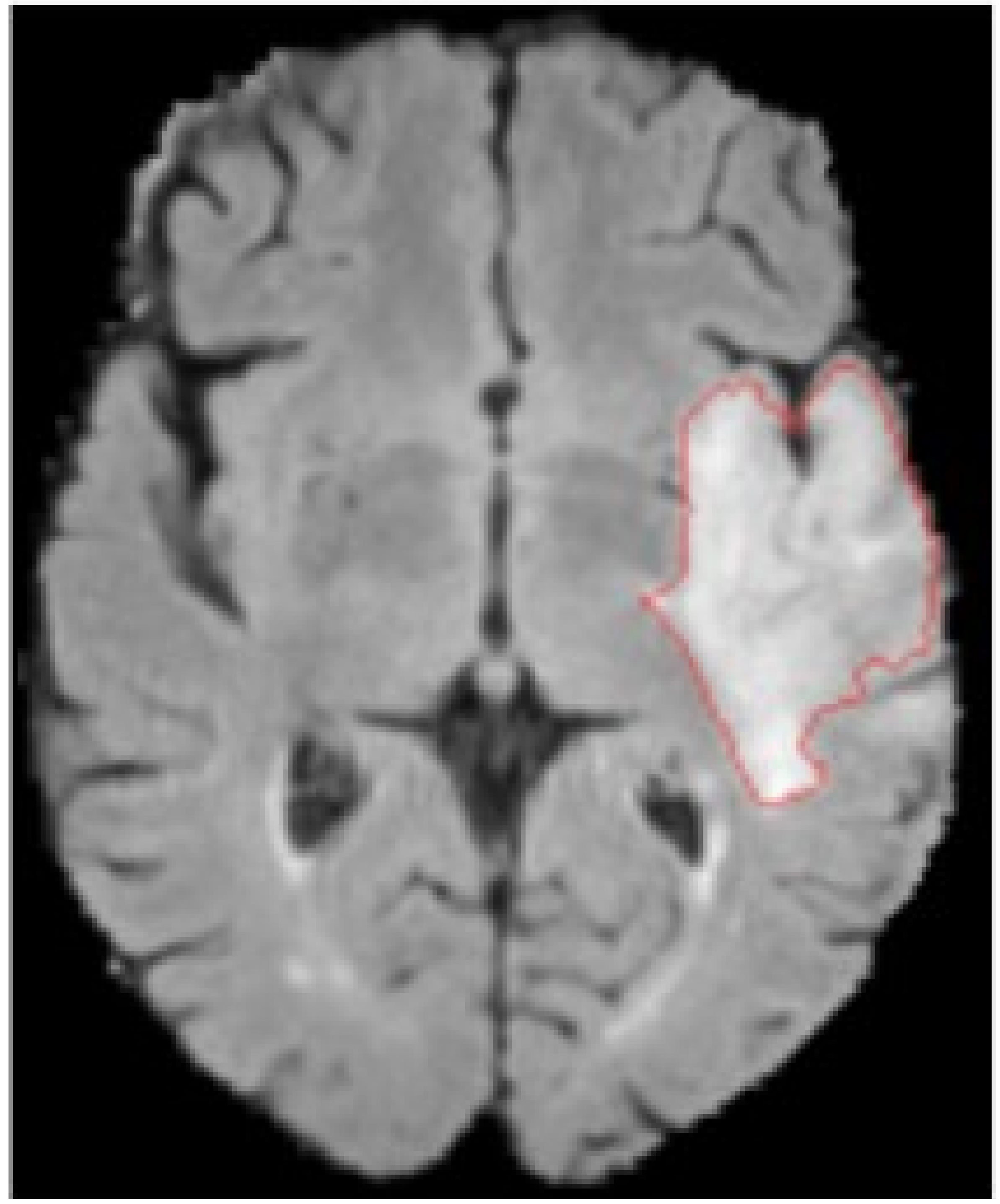
Deep learning in medical products



Training set T

BraTS
Multimodal Brain Tumor Segmentation Challenge

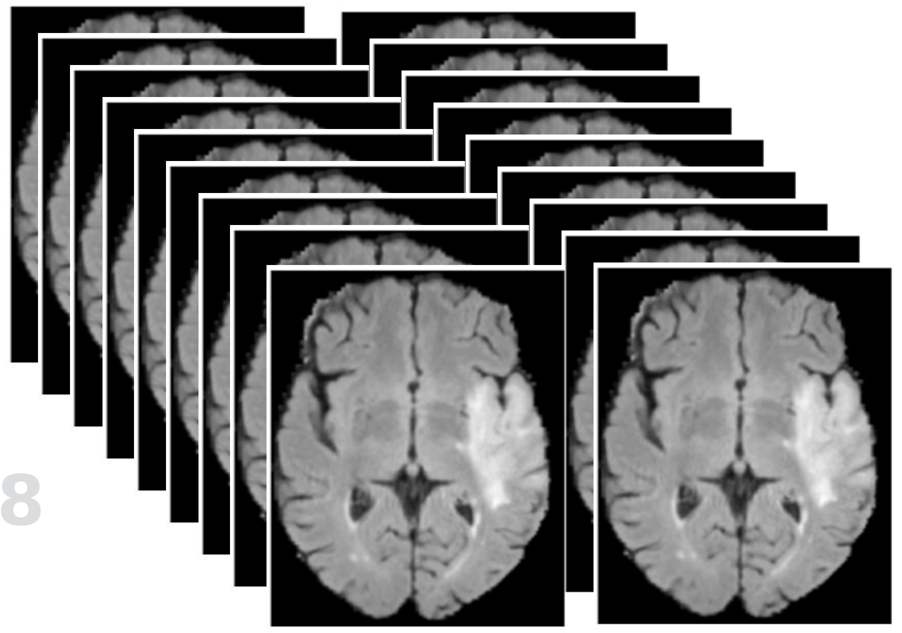




J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

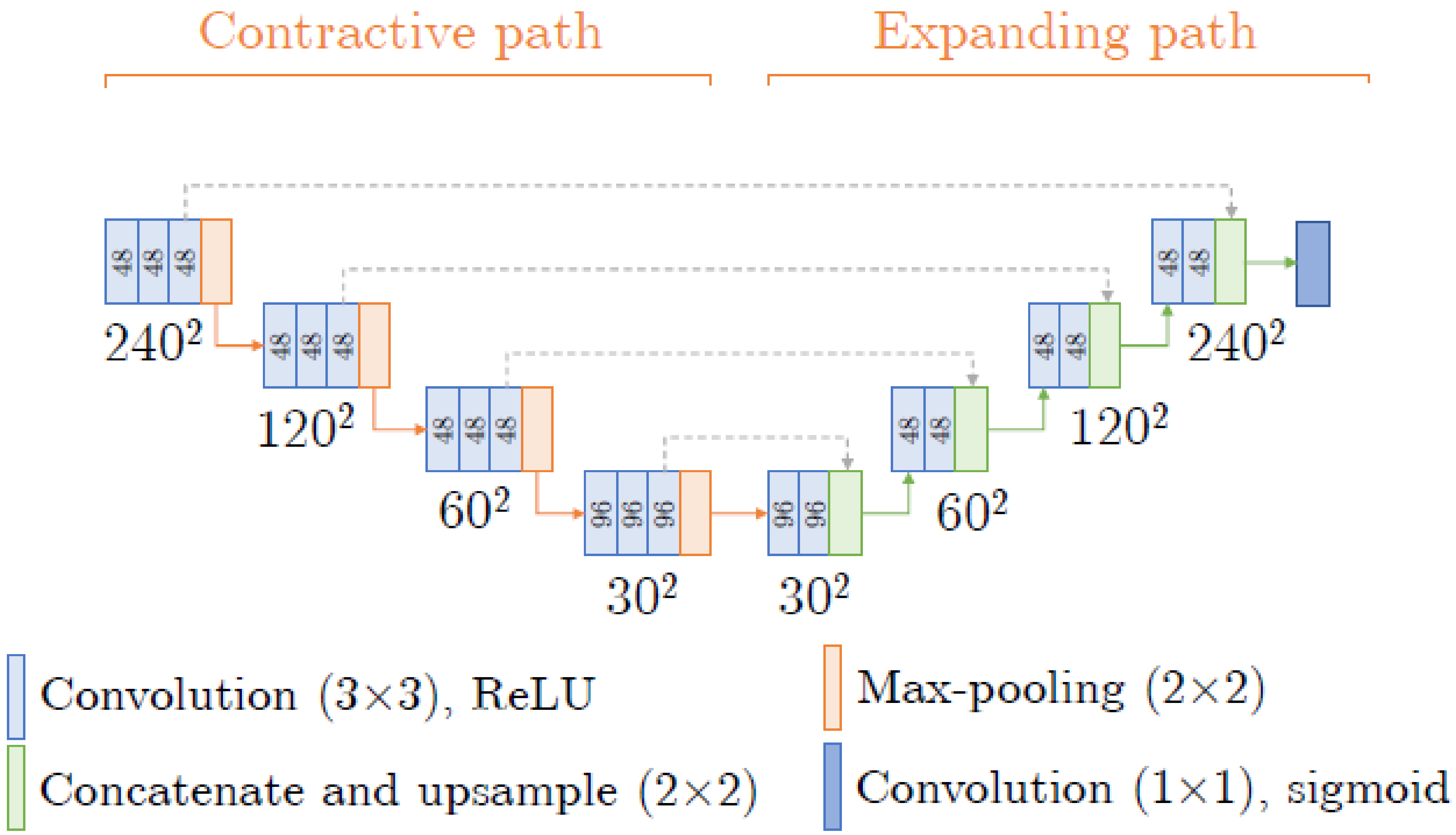
Deep learning in medical products

Sens.Ai

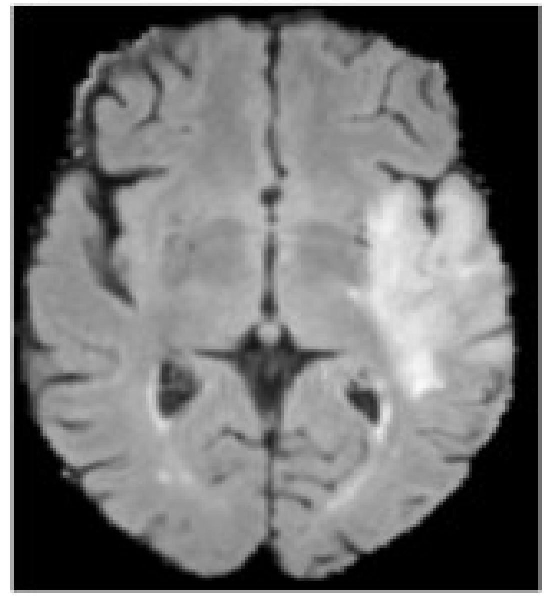


Training set T

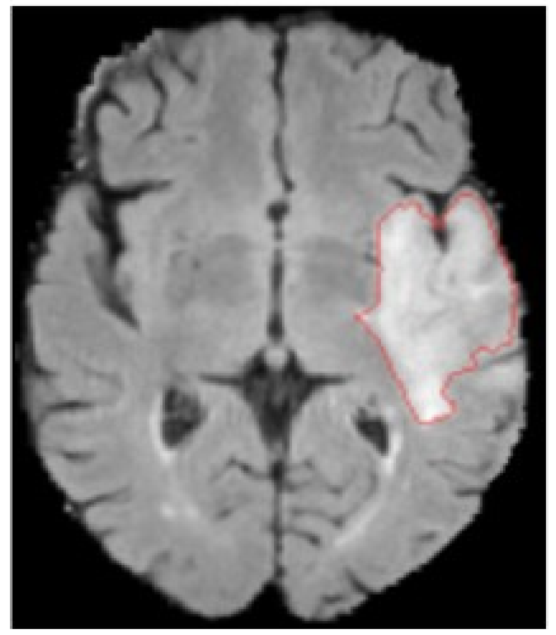
BraTS Multimodal Brain Tumor Segmentation Challenge



Test data

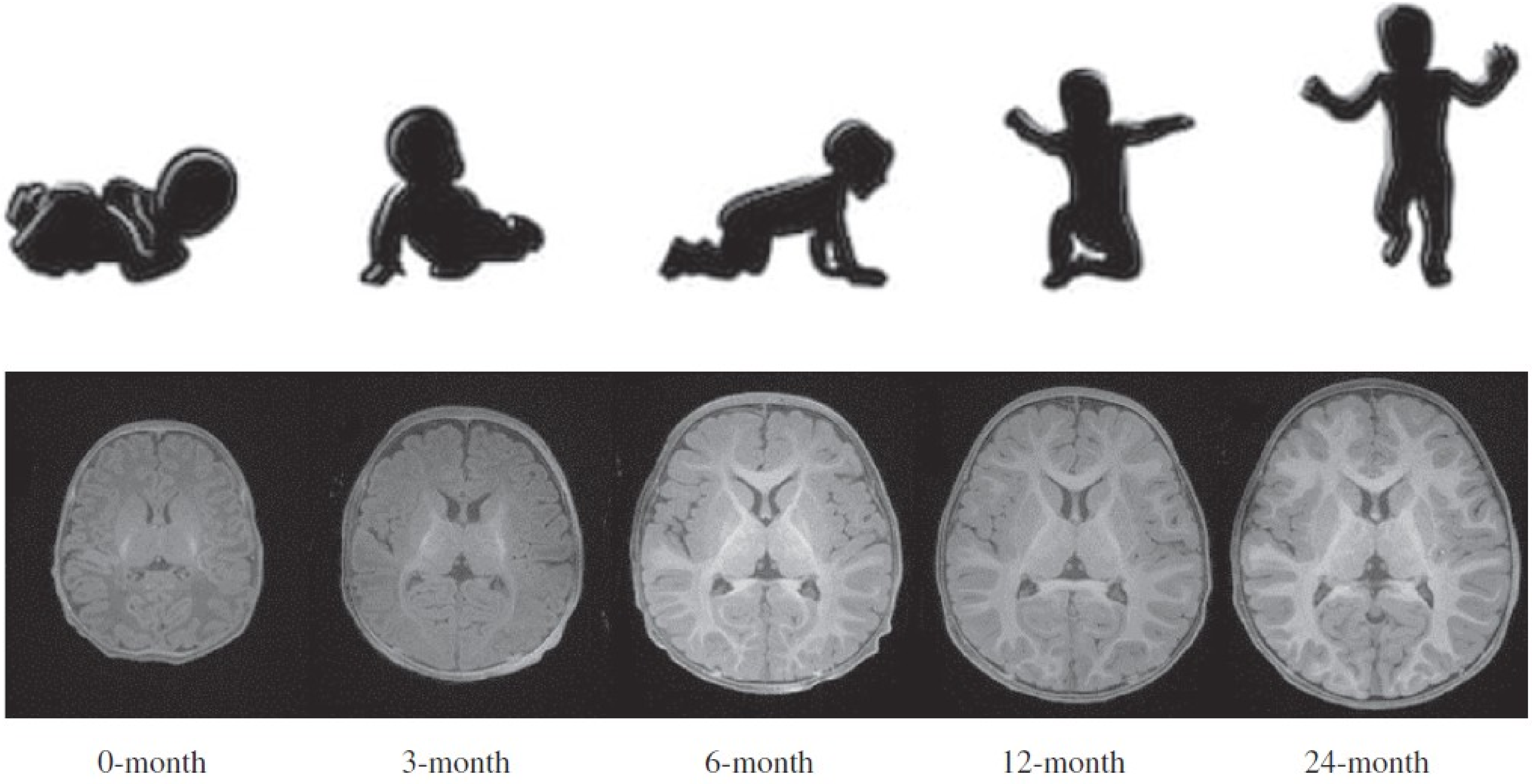


Brain tumor annotated by Sens.AI



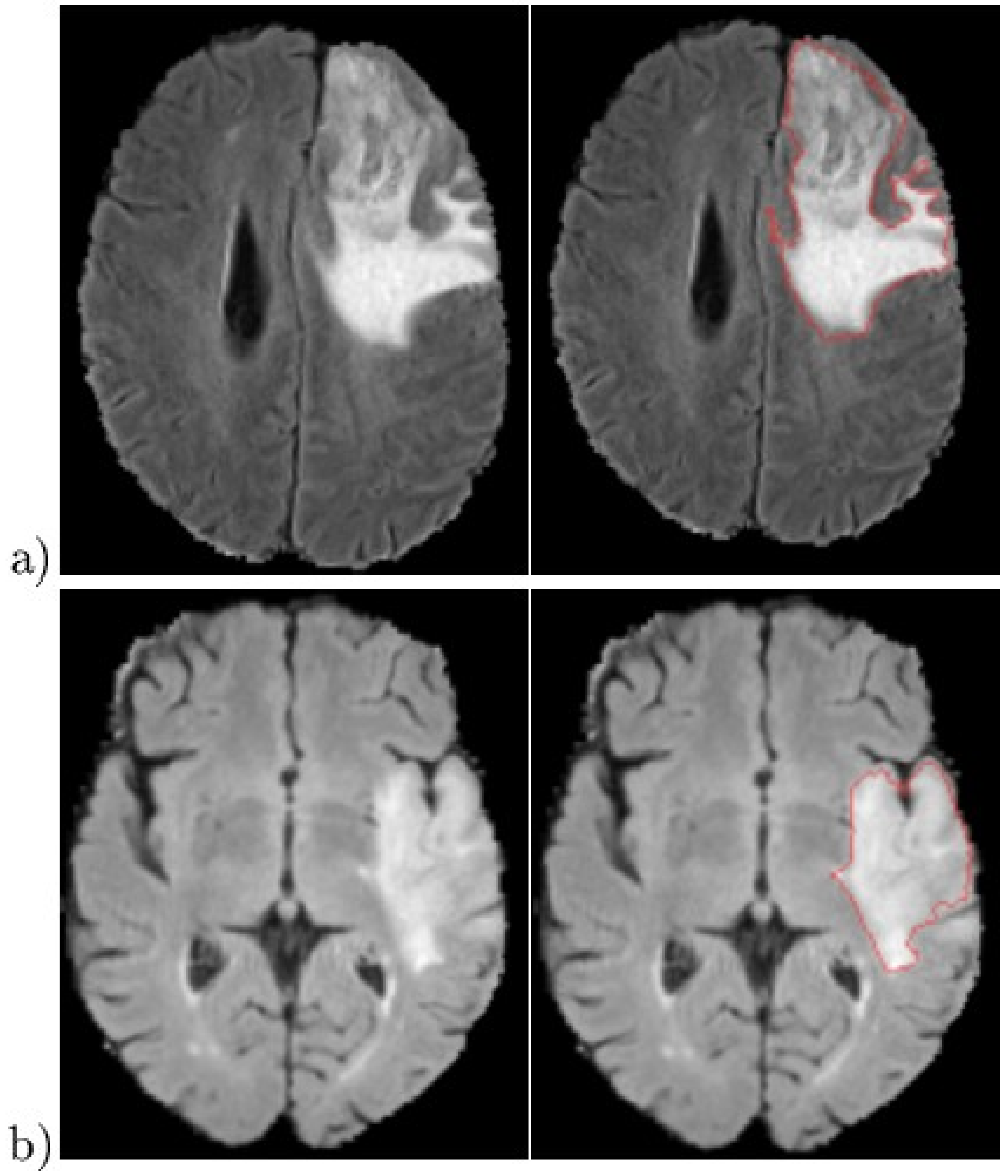
J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

S. McKinney et al., International evaluation of an AI system for breast cancer screening, Nature volume 577, pages 89–94 (2020)



Mostapha and Styner, Role of deep learning in infant brain MRI analysis, Magnetic Resonance Imaging Volume 64, December 2019, Pages 171-189, 2019.

Is our **T** *of sufficient quality?*

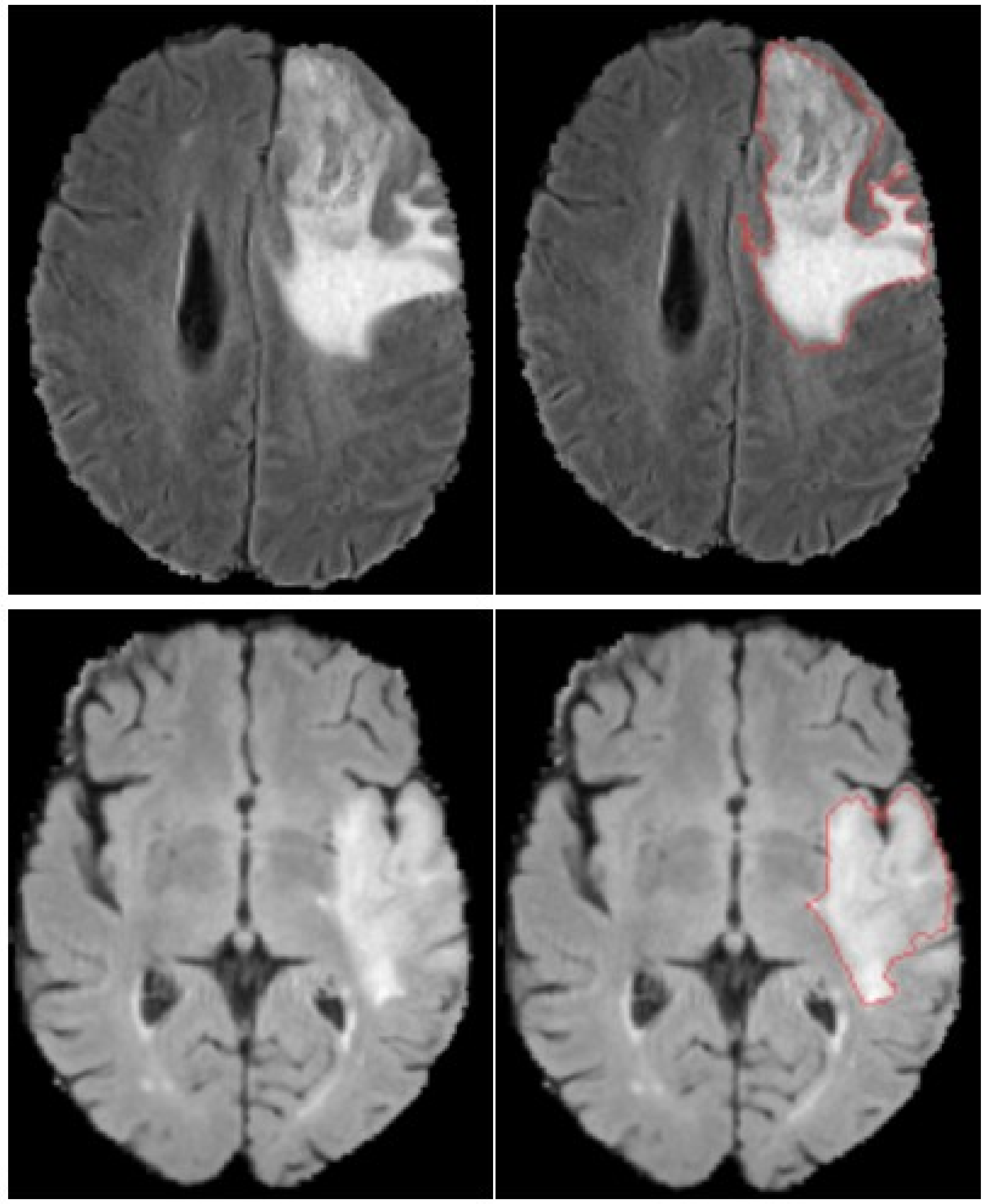


Score	Description	Outcome
1	**Very low quality** segmentation	**No**, I would not use it to support diagnosis
2	**Low quality** segmentation	**No**, I would not use it to support diagnosis
3	**Acceptable** segmentation	**Yes**, I would use it to support diagnosis
4	**Very high-quality** segmentation	**Yes**, I would use it to support diagnosis

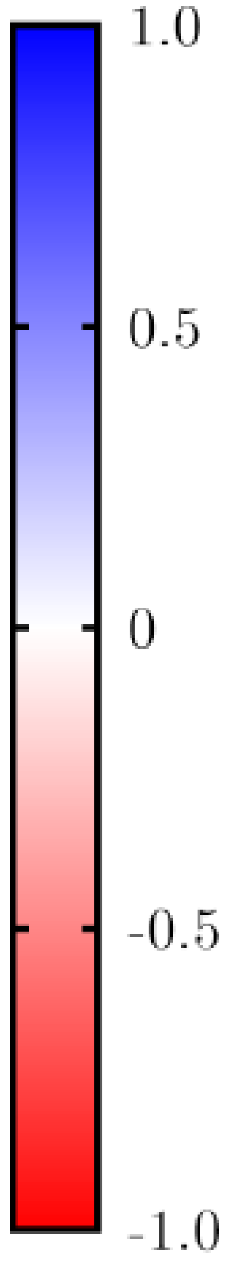
Example ground-truth (GT) segmentations for patients a-b in BraTS 2017 (left image), alongside the Mean Opinion Score scale

Is our **T** *of sufficient quality?*

T2-FLAIR Ground truth

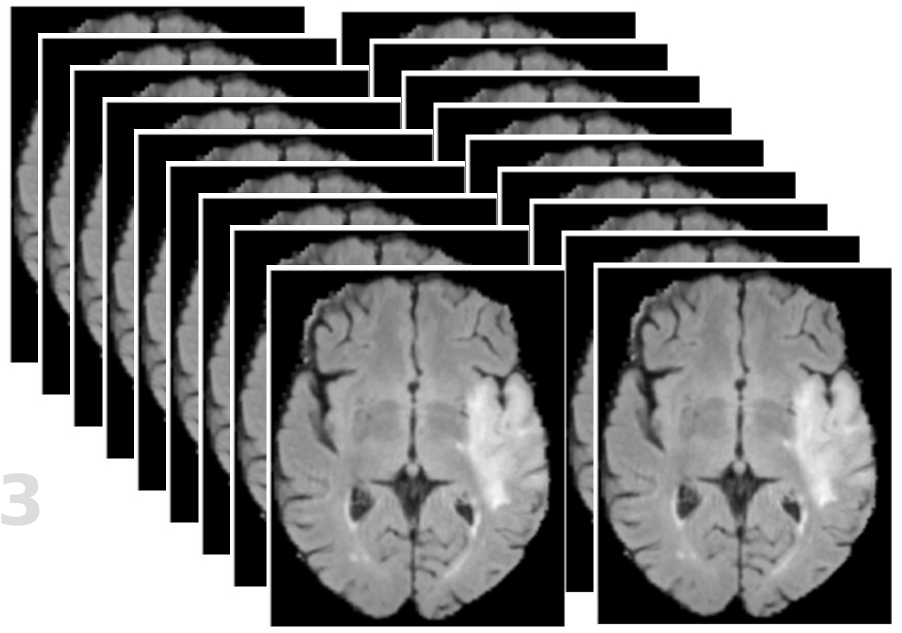


	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6	Reader 7	Reader 8	Reader 9	Reader 10	Reader 11	Reader 12
Reader 1	1.000	0.363	0.385	0.090	0.809	-0.044	0.408	0.296	0.552	0.320	0.457	0.400
Reader 2	0.363	1.000	0.405	0.132	0.504	0.150	0.326	0.254	0.327	0.124	0.250	0.305
Reader 3	0.385	0.405	1.000	0.319	0.373	0.233	0.369	0.386	0.512	0.313	0.457	0.516
Reader 4	0.090	0.132	0.319	1.000	0.223	-0.003	0.243	0.381	0.273	-0.030	0.348	0.105
Reader 5	0.809	0.504	0.373	0.223	1.000	-0.098	0.405	0.361	0.387	0.200	0.430	0.365
Reader 6	-0.044	0.150	0.233	-0.003	-0.098	1.000	0.057	0.304	0.141	0.291	0.150	0.241
Reader 7	0.408	0.326	0.369	0.243	0.405	0.057	1.000	0.484	0.418	0.107	0.280	0.366
Reader 8	0.296	0.254	0.386	0.381	0.361	0.304	0.484	1.000	0.185	0.259	0.585	0.426
Reader 9	0.552	0.327	0.512	0.273	0.387	0.141	0.418	0.185	1.000	0.545	0.160	0.227
Reader 10	0.320	0.124	0.313	-0.030	0.200	0.291	0.107	0.259	0.545	1.000	0.222	0.201
Reader 11	0.457	0.250	0.457	0.348	0.430	0.150	0.280	0.585	0.160	0.222	1.000	0.470
Reader 12	0.400	0.305	0.516	0.105	0.365	0.241	0.366	0.426	0.227	0.201	0.470	1.000



Example ground truth (GT) for patients a-b in BraTS – **165 HGG i 63 LGG** (left image) together with the inter-rater agreement

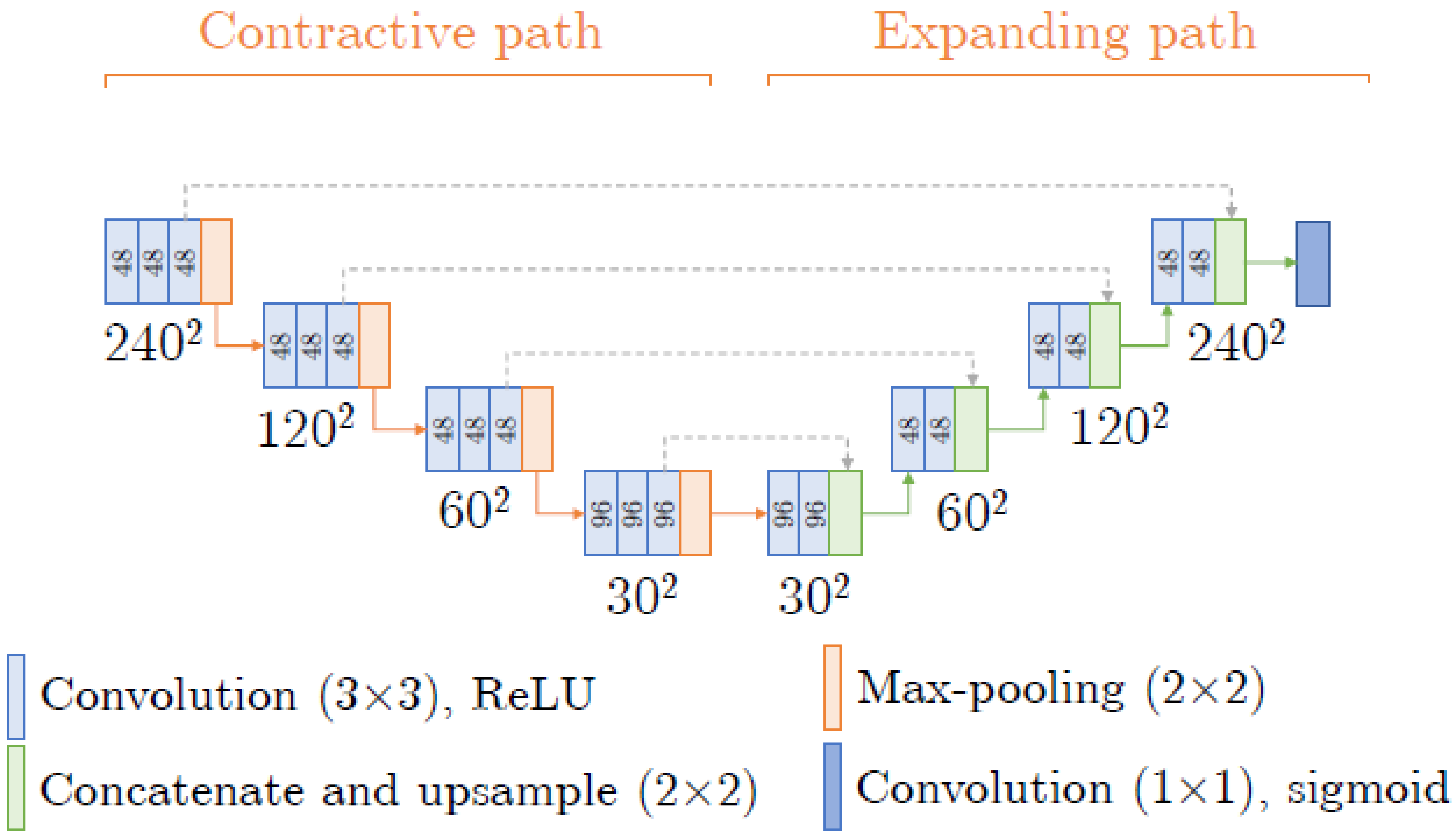
Deep learning in medical products

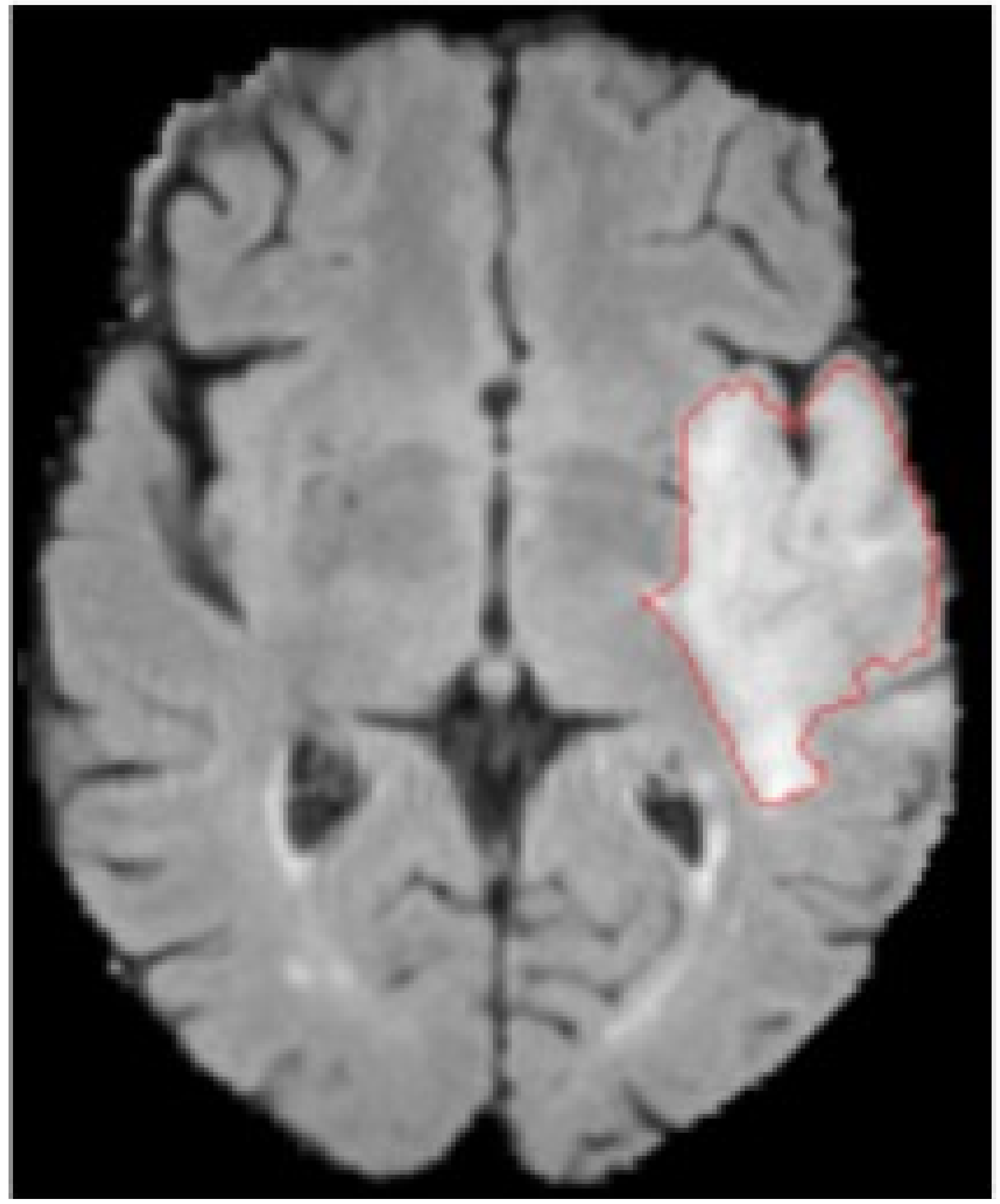


Training set T

~**1250 MRIs** for brain extraction (four data sources, various modalities, reviewed by 3 readers with 11, 7 and 5 YOE)

228 MRIs for BT segmentation (165 HGG and 63 LGG)

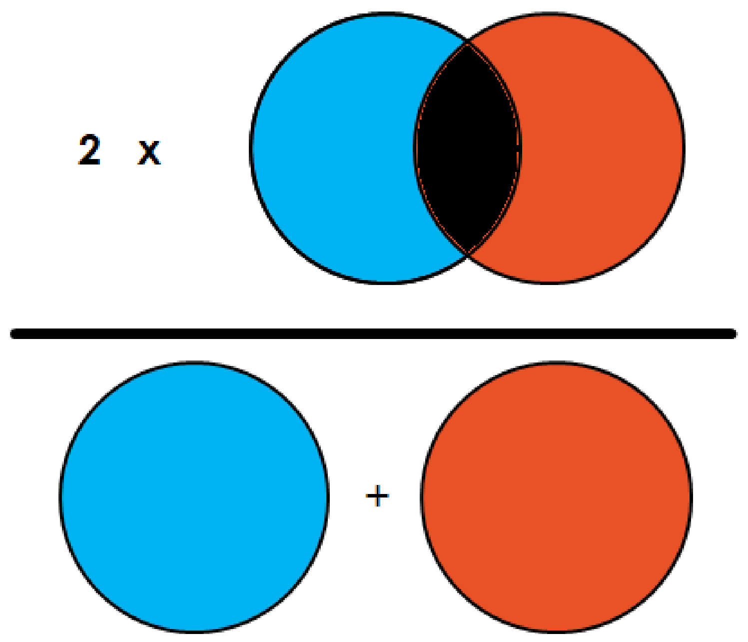


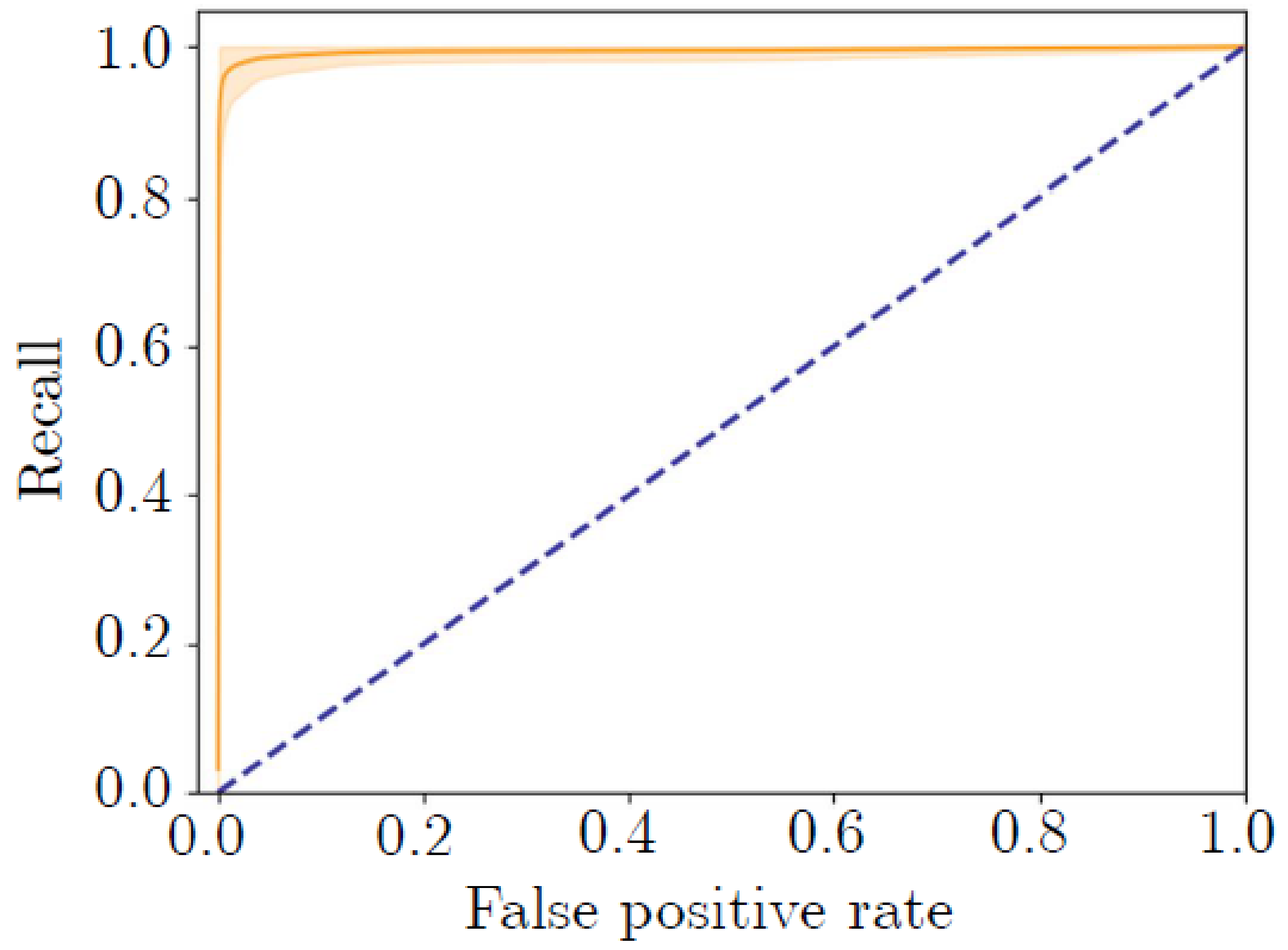


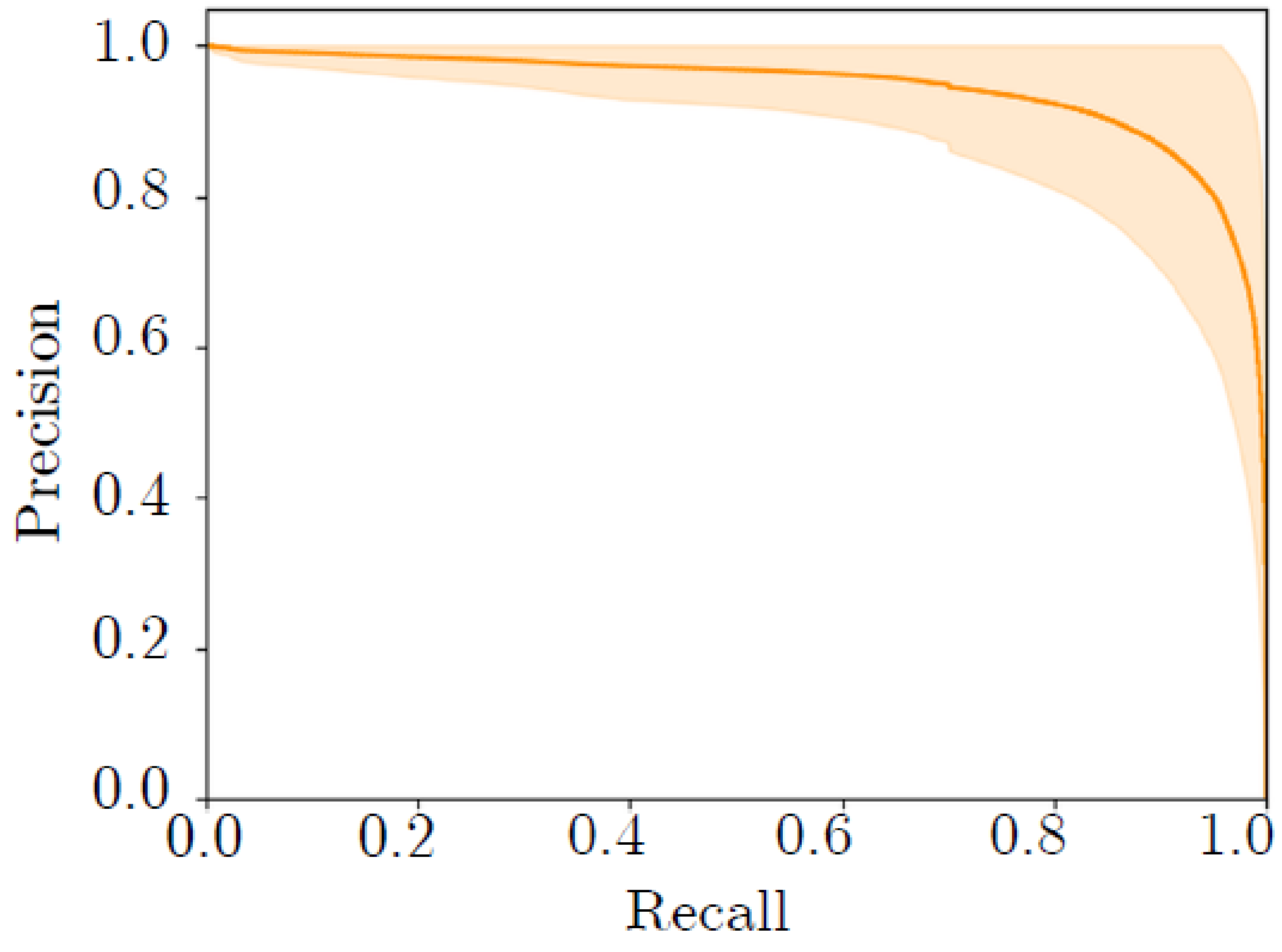
J. Nalepa et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artificial Intelligence in Medicine: 102: 101769 (2020)

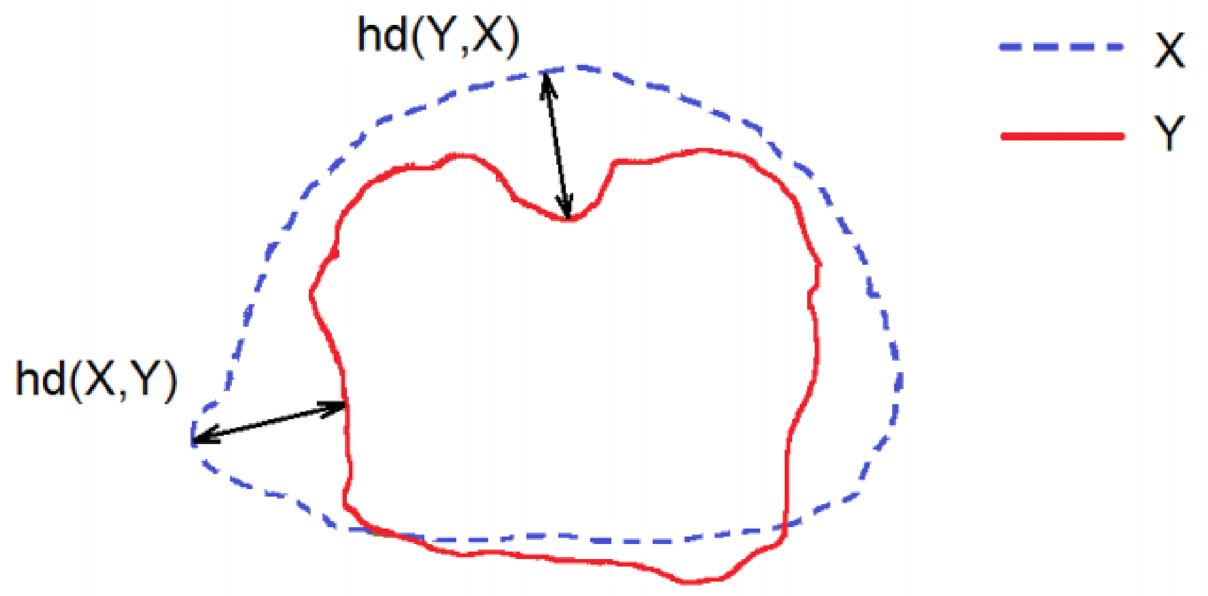
Is the algorithm „ready”?

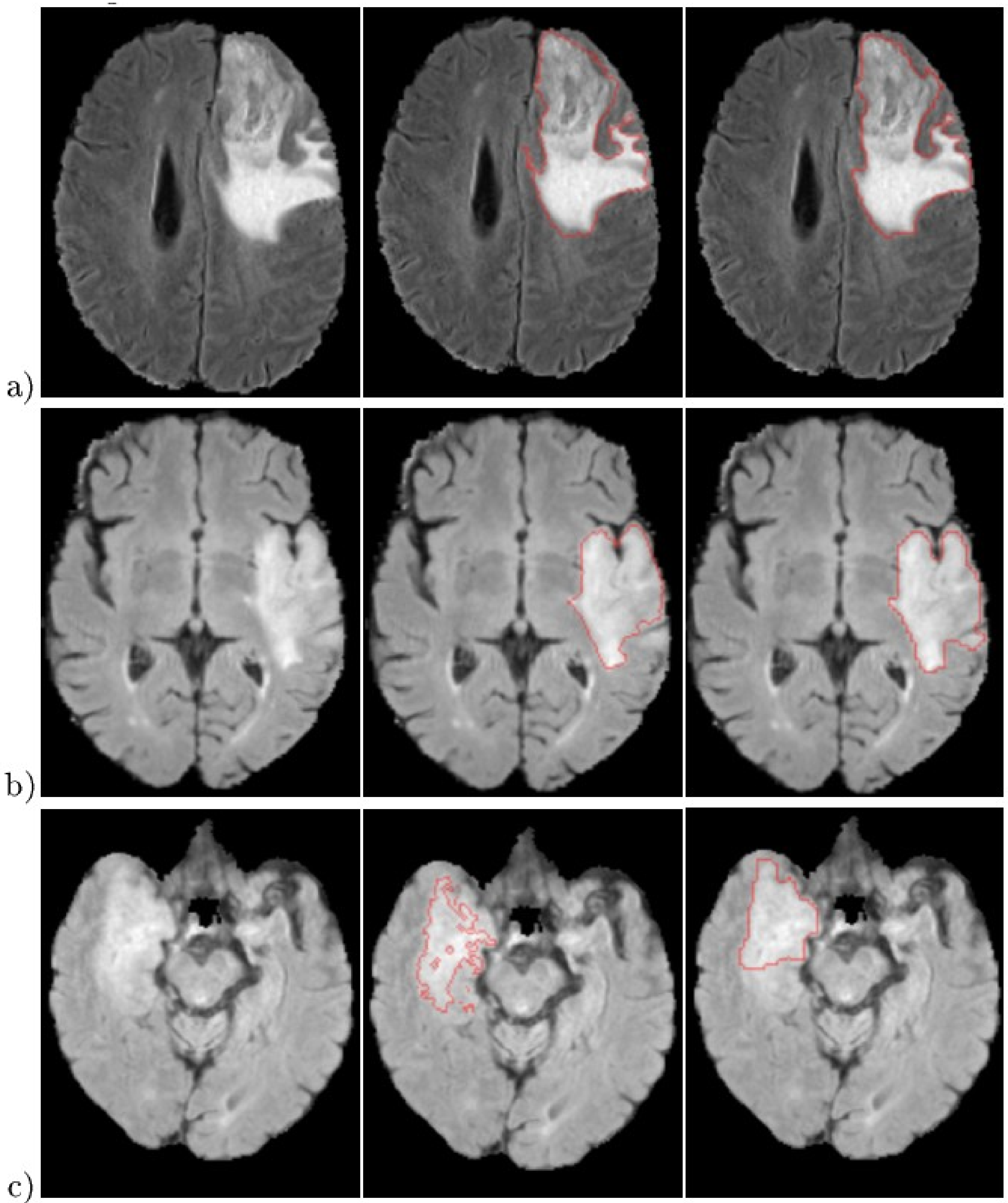
	DICE	GT Volume [mm^3]	Volume [mm^3]	Volume error [%]	AUC (ROC)	AUC (P – R)	HD	HD (95)
Min.	0.609	32051	29223	0.508	0.945	0.643	6.164	1.000
Mean	0.882	112120	105891	13.339	0.994	0.896	23.884	8.807
Max.	0.968	232098	212323	58.094	1.000	0.979	71.127	41.713
Std. dev.	0.095	59859	53784	15.794	0.014	0.095	20.217	11.769

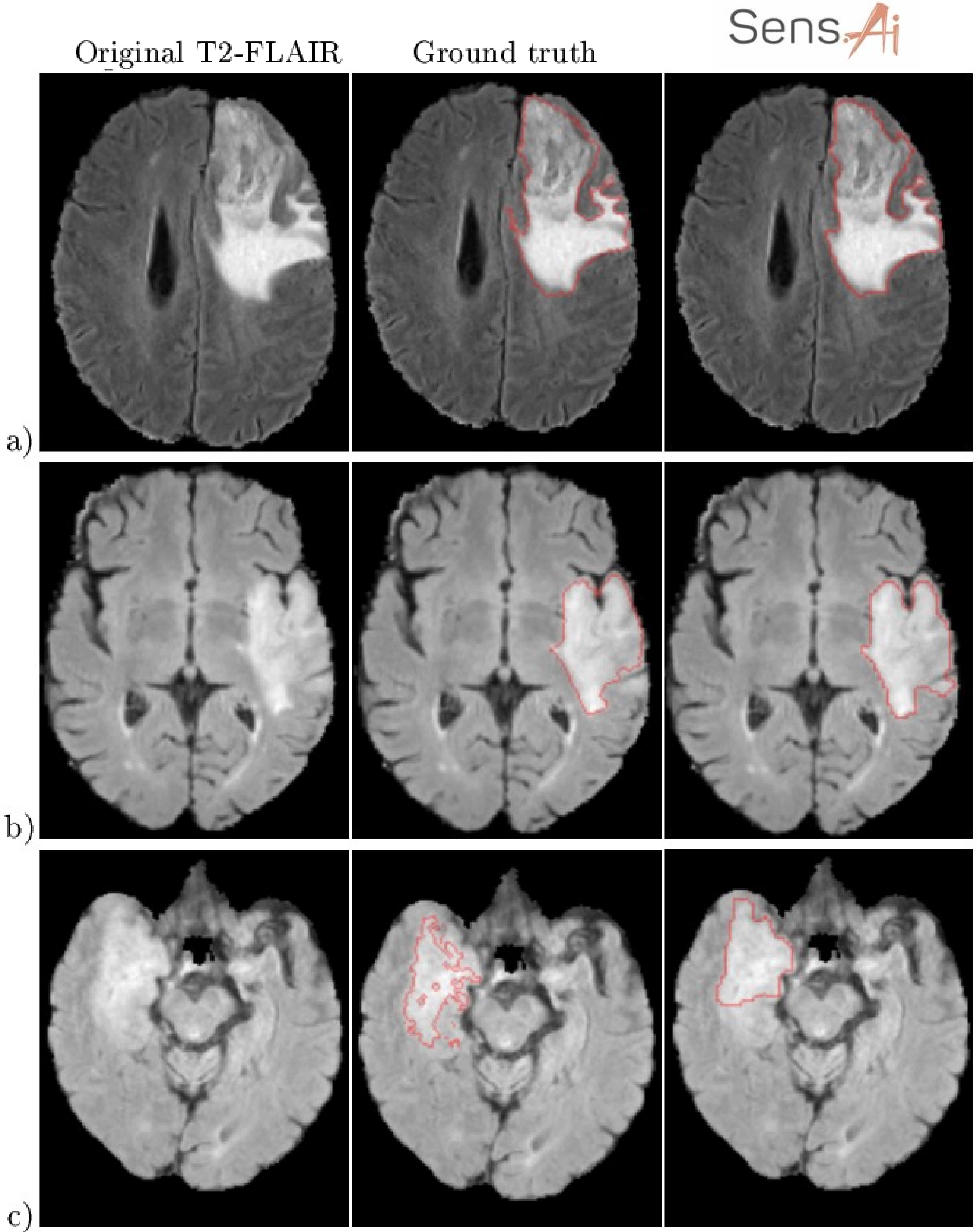




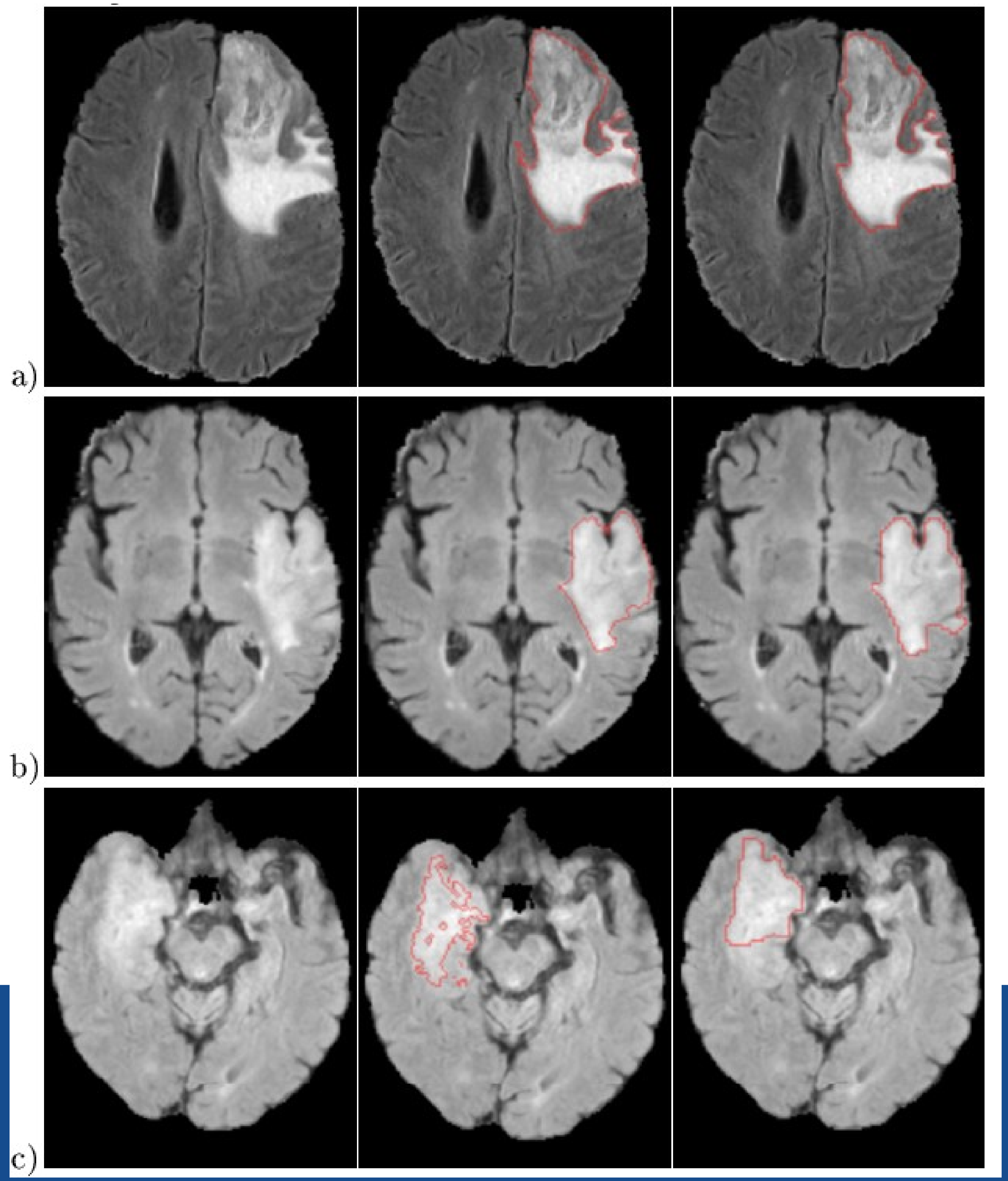





a)
b)
c)


Original T2-FLAIR
Ground truth
Sens.Ai
a)
b)
c)

Is the algorithm „ready”?



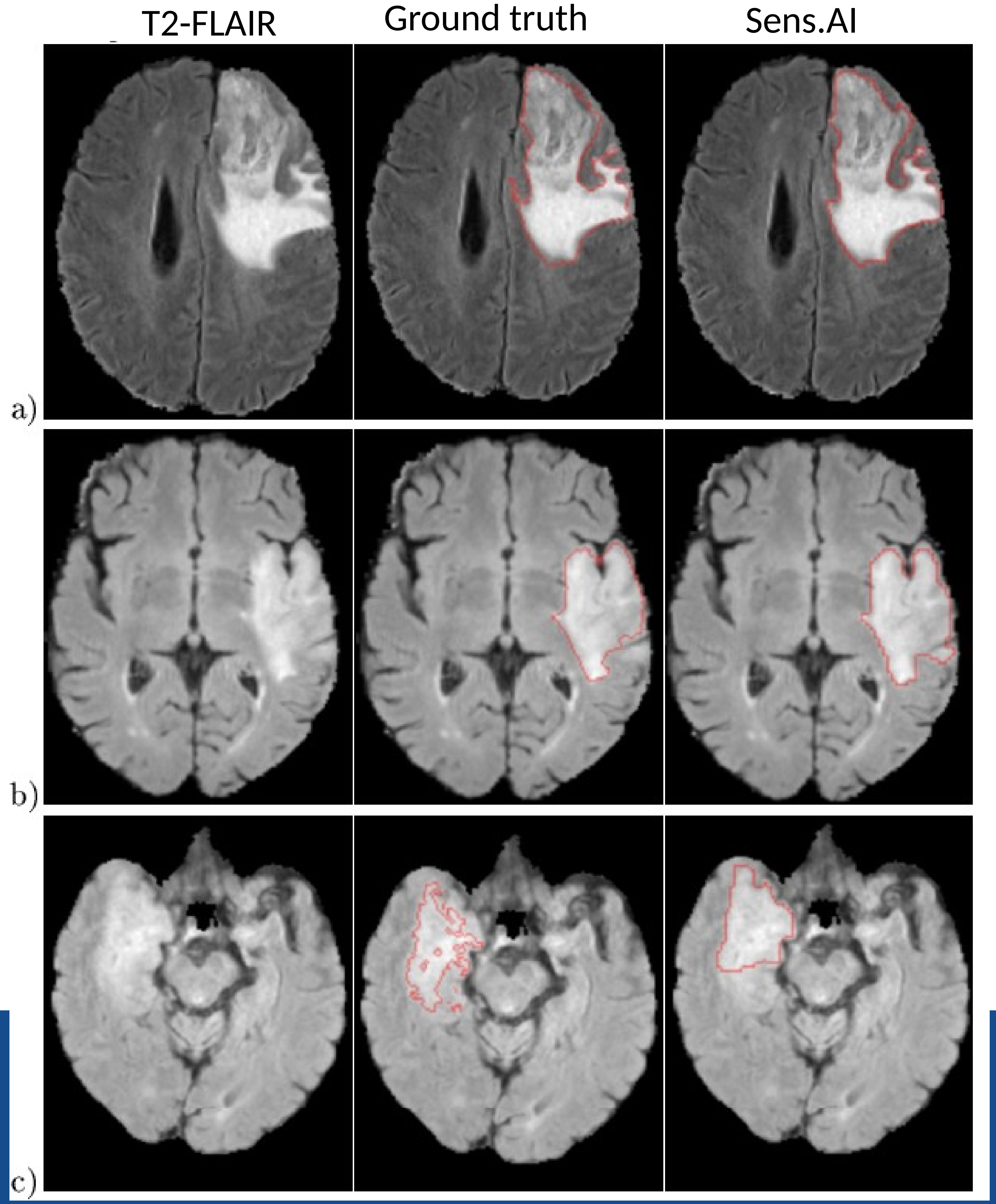
MOS from the a) training set from BraTS and unseen clinical data (CT – 25 HGG i 25 LGG).

Is the algorithm „ready”?

ID	Institution	YOE	BraTS (GT)	BraTS (Sens.AI)	CD
Reader 1	Institution A	3	3.000	2.697	2.300
Reader 2	Institution A	3	3.273	2.909	2.720
Reader 3	Institution A	3	2.152	1.909	1.660
Reader 4	Institution A	4	2.848	2.758	2.740
Reader 5	Institution A	4	3.273	2.939	2.620
Reader 6	Institution A	4	3.697	3.333	2.600
Reader 7	Institution A	5	2.242	2.455	1.940
Reader 8	Institution B	5	3.333	3.152	2.300
Reader 9	Institution B	7	2.909	2.758	2.040
Reader 10	Institution B	9	3.182	2.788	2.000
Reader 11	Institution B	11	2.848	2.636	2.080
Reader 12	Institution B	11	3.424	3.273	2.540
		Average score	3.015	2.801	2.330
		Weighted average score	3.050	2.842	2.318

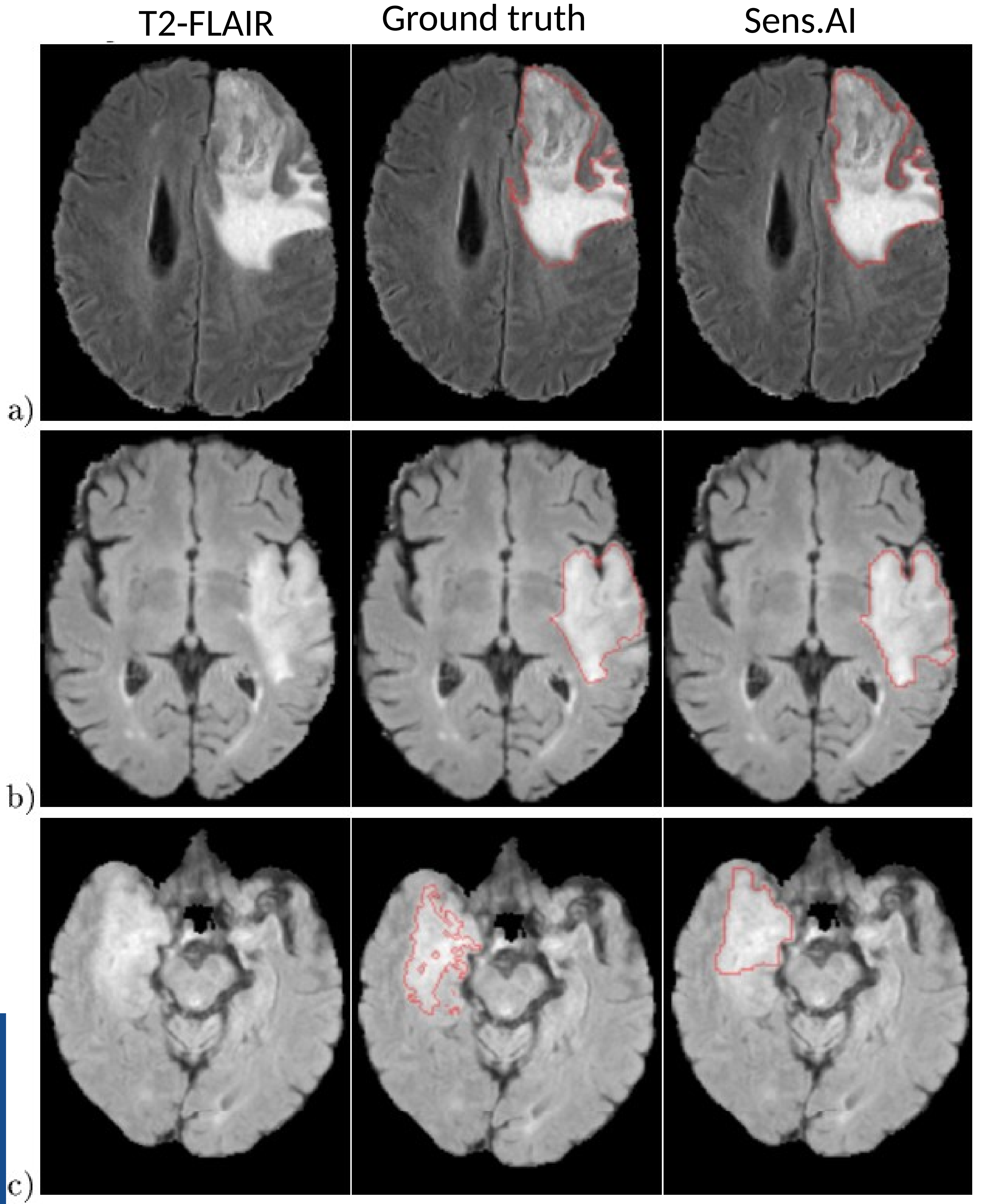
Score	Description	Outcome
1	**Very low quality** segmentation	**No**, I would not use it to support diagnosis
2	**Low quality** segmentation	**No**, I would not use it to support diagnosis
3	**Acceptable** segmentation	**Yes**, I would use it to support diagnosis
4	**Very high-quality** segmentation	**Yes**, I would use it to support diagnosis

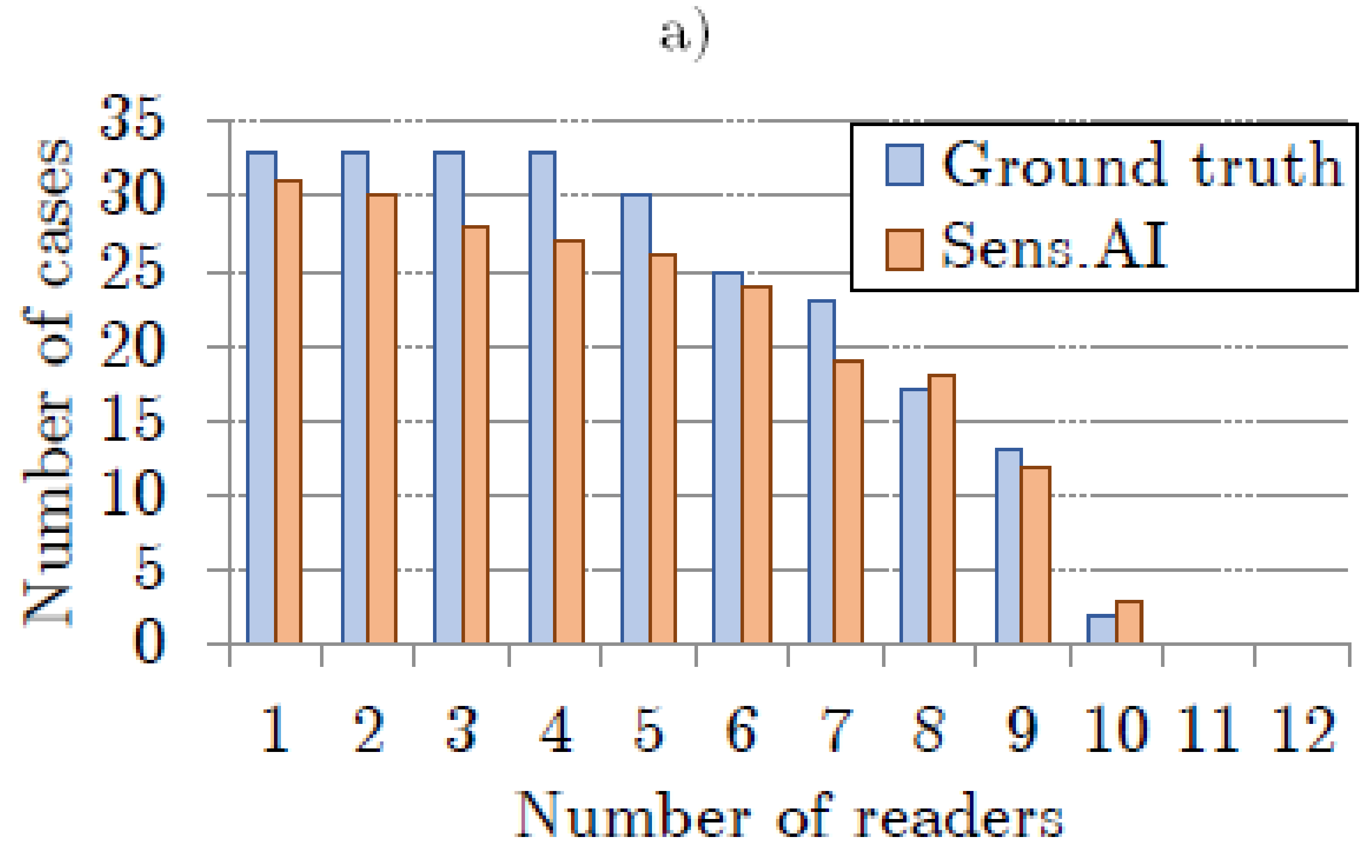
Is the algorithm „ready”?

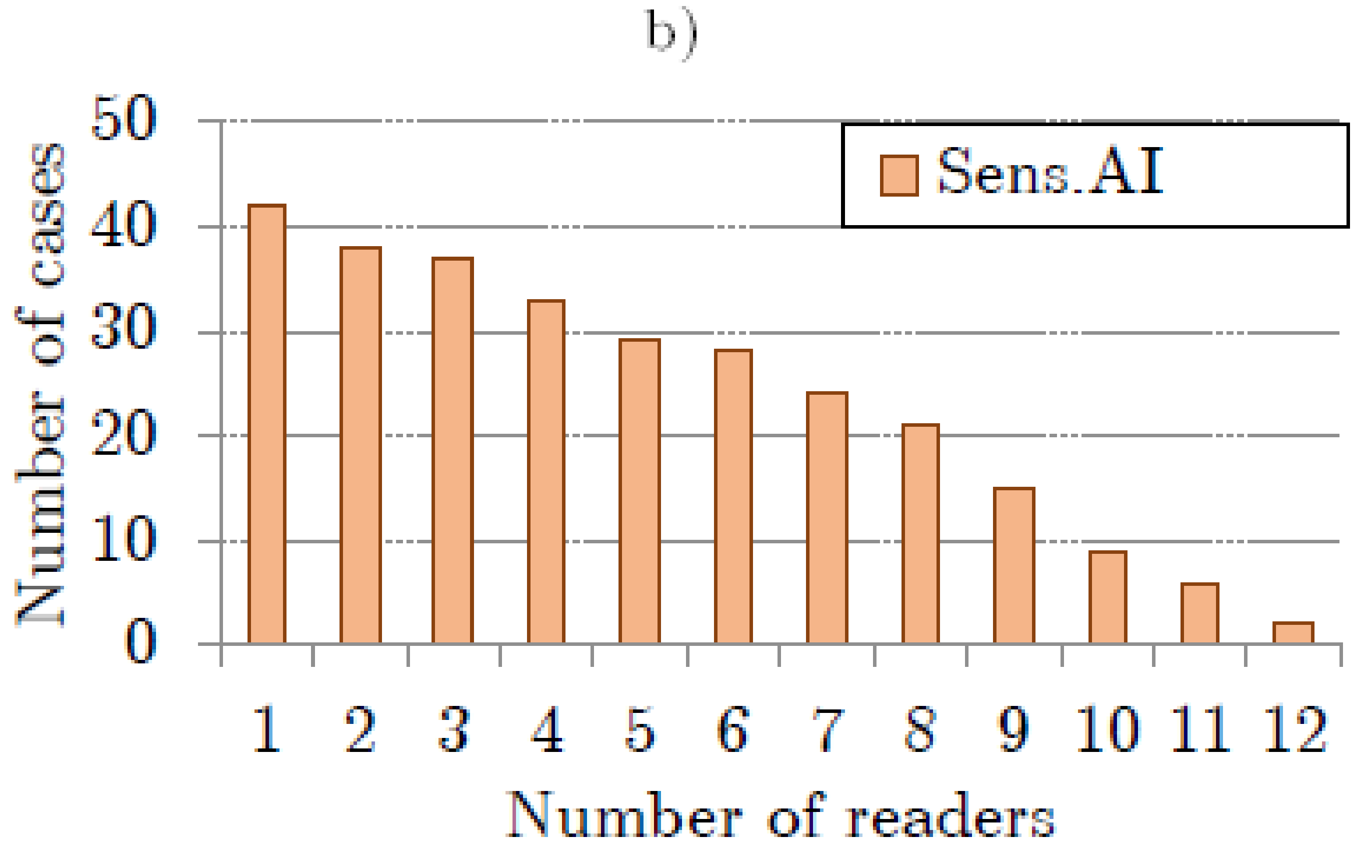


MOS from the a) training set from BraTS and unseen clinical data (CT – 25 HGG i 25 LGG).

Is the algorithm „ready”?







	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6	Reader 7	Reader 8	Reader 9	Reader 10	Reader 11	Reader 12
Reader 1	1.000	0.706	0.452	0.735	0.588	0.594	0.158	0.738	0.540	0.593	0.578	0.566
Reader 2	0.706	1.000	0.418	0.510	0.648	0.393	0.207	0.680	0.403	0.321	0.447	0.430
Reader 3	0.452	0.418	1.000	0.506	0.426	0.370	0.292	0.514	0.593	0.546	0.315	0.390
Reader 4	0.735	0.510	0.506	1.000	0.633	0.564	0.379	0.591	0.500	0.581	0.569	0.253
Reader 5	0.588	0.648	0.426	0.633	1.000	0.611	0.430	0.584	0.622	0.514	0.450	0.315
Reader 6	0.594	0.393	0.370	0.564	0.611	1.000	0.506	0.524	0.690	0.611	0.547	0.611
Reader 7	0.158	0.207	0.292	0.379	0.430	0.506	1.000	0.344	0.474	0.167	0.638	0.380
Reader 8	0.738	0.680	0.514	0.591	0.584	0.524	0.344	1.000	0.552	0.406	0.511	0.518
Reader 9	0.540	0.403	0.593	0.500	0.622	0.690	0.474	0.552	1.000	0.614	0.529	0.500
Reader 10	0.593	0.321	0.546	0.581	0.514	0.611	0.167	0.406	0.614	1.000	0.396	0.336
Reader 11	0.578	0.447	0.315	0.569	0.450	0.547	0.638	0.511	0.529	0.396	1.000	0.552
Reader 12	0.566	0.430	0.390	0.253	0.315	0.611	0.380	0.518	0.500	0.336	0.552	1.000

	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6	Reader 7	Reader 8	Reader 9	Reader 10	Reader 11	Reader 12
Reader 1	1.000	0.653	0.646	0.464	0.542	0.682	0.397	0.552	0.479	0.725	0.826	0.681
Reader 2	0.653	1.000	0.569	0.457	0.520	0.699	0.295	0.594	0.546	0.738	0.745	0.684
Reader 3	0.646	0.569	1.000	0.448	0.494	0.670	0.327	0.548	0.480	0.544	0.623	0.447
Reader 4	0.464	0.457	0.448	1.000	0.450	0.443	0.320	0.265	0.335	0.521	0.492	0.441
Reader 5	0.542	0.520	0.494	0.450	1.000	0.762	0.361	0.604	0.676	0.685	0.566	0.651
Reader 6	0.682	0.699	0.670	0.443	0.762	1.000	0.395	0.743	0.628	0.746	0.687	0.698
Reader 7	0.397	0.295	0.327	0.320	0.361	0.395	1.000	0.479	0.427	0.463	0.435	0.421
Reader 8	0.552	0.594	0.548	0.265	0.604	0.743	0.479	1.000	0.642	0.687	0.568	0.601
Reader 9	0.479	0.546	0.480	0.335	0.676	0.628	0.427	0.642	1.000	0.762	0.623	0.554
Reader 10	0.725	0.738	0.544	0.521	0.685	0.746	0.463	0.687	0.762	1.000	0.834	0.784
Reader 11	0.826	0.745	0.623	0.492	0.566	0.687	0.435	0.568	0.623	0.834	1.000	0.680
Reader 12	0.681	0.684	0.447	0.441	0.651	0.698	0.421	0.601	0.554	0.784	0.680	1.000

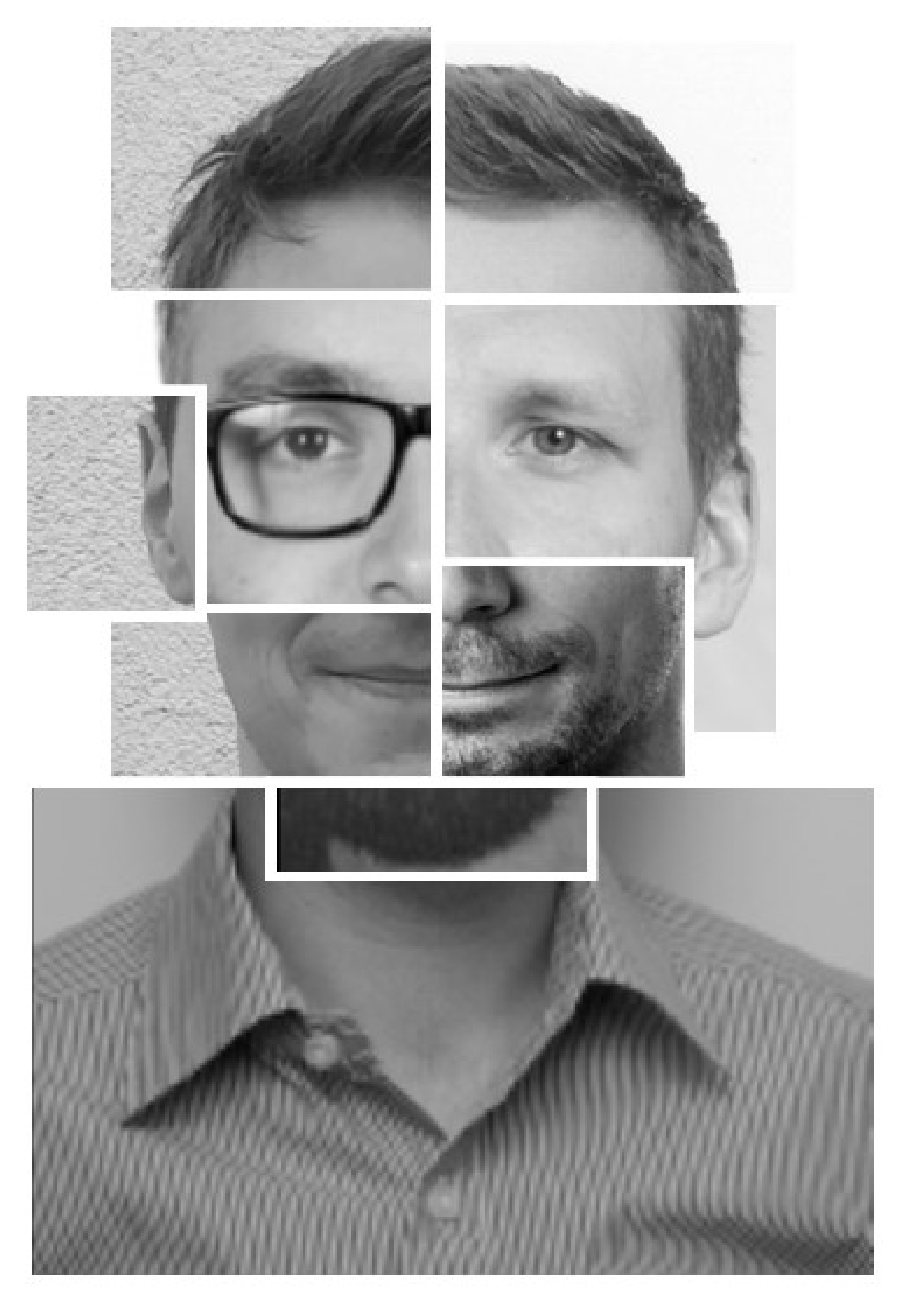
MOS from the a) training set from BraTS and unseen clinical data (CT – 25 HGG i 25 LGG).

„Machine learning algorithms will transform clinical imaging practice over the next decade."

Langlotz et al., Radiology, Vol. 291, No. 3, 2019.

CAN AI MAKE US SEE BEYOND THE VISIBLE: TOWARD CE MARKED DEEP LEARNING SOFTWARE FOR MEDICAL IMAGE ANALYSIS

Jakub Nalepa



Politechnika Śląska

GRAYLIGHT
MEDICAL IMAGING SOFTWARE

E-mail
jnalepa@ieee.org