



Automated volumetric assessment of pituitary adenoma

Raffaele Da Mutton¹ · Olivier Zanier¹ · Olga Ciobanu-Carus¹ · Stefanos Voglis² · Michael Hugelshofer² · Athina Pangalu³ · Luca Regli^{1,2} · Carlo Serra^{1,2} · Victor E. Staartjes^{1,2}

Received: 17 April 2023 / Accepted: 7 September 2023 / Published online: 25 September 2023
© The Author(s) 2023

Abstract

Purpose Assessment of pituitary adenoma (PA) volume and extent of resection (EOR) through manual segmentation is time-consuming and likely suffers from poor interrater agreement, especially postoperatively. Automated tumor segmentation and volumetry by use of deep learning techniques may provide more objective and quick volumetry.

Methods We developed an automated volumetry pipeline for pituitary adenoma. Preoperative and three-month postoperative T1-weighted, contrast-enhanced magnetic resonance imaging (MRI) with manual segmentations were used for model training. After adequate preprocessing, an ensemble of convolutional neural networks (CNNs) was trained and validated for preoperative and postoperative automated segmentation of tumor tissue. Generalization was evaluated on a separate holdout set.

Results In total, 193 image sets were used for training and 20 were held out for validation. At validation using the holdout set, our models (preoperative / postoperative) demonstrated a median Dice score of 0.71 (0.27) / 0 (0), a mean Jaccard score of $0.53 \pm 0.21 / 0.030 \pm 0.085$ and a mean 95th percentile Hausdorff distance of $3.89 \pm 1.96 / 12.199 \pm 6.684$. Pearson's correlation coefficient for volume correlation was 0.85 / 0.22 and -0.14 for extent of resection. Gross total resection was detected with a sensitivity of 66.67% and specificity of 36.36%.

Conclusions Our volumetry pipeline demonstrated its ability to accurately segment pituitary adenomas. This is highly valuable for lesion detection and evaluation of progression of pituitary incidentalomas. Postoperatively, however, objective and precise detection of residual tumor remains less successful. Larger datasets, more diverse data, and more elaborate modeling could potentially improve performance.

Keywords Pituitary · Extent of resection · Machine learning · Deep learning · Endocrinology · Neurosurgery

Introduction

Pituitary adenomas (PA) are a frequent type of intracranial tumor [1]. Endonasal transsphenoidal surgery has established itself as the best option for its treatment in most cases [2]. Its outcome varies greatly with different factors like tumor morphology and the surgeon caseload [3–5]. Treatment is

indicated in case of functioning PA other than prolactinomas, in case of symptomatic PA or in case of relevant volumetric progression [6]. If surgery is performed, assessment of residual tumor is relevant in order to determine the extent of resection (EOR) [7], though manual segmentation of tumor volumes is likely highly dependent on the rater, especially postoperatively [8–10]. Automated analysis of pre- and postoperative imaging could consequently have the potential to provide more objective and precise volumetry.

Semantic image segmentation is a classic machine learning application [11, 12], not only due to the fact that manual segmentation requires considerable amounts of expert time [13, 14]. Convolutional neural networks (CNNs)—and specifically U-Nets—have recently been applied successfully for biomedical image segmentation due to their throughput speed and overall good performance in this task [15].

To the best of the authors' knowledge, no automated approaches to segment PA pre- and postoperatively for volumetry and resection assessment exist. We hypothesize

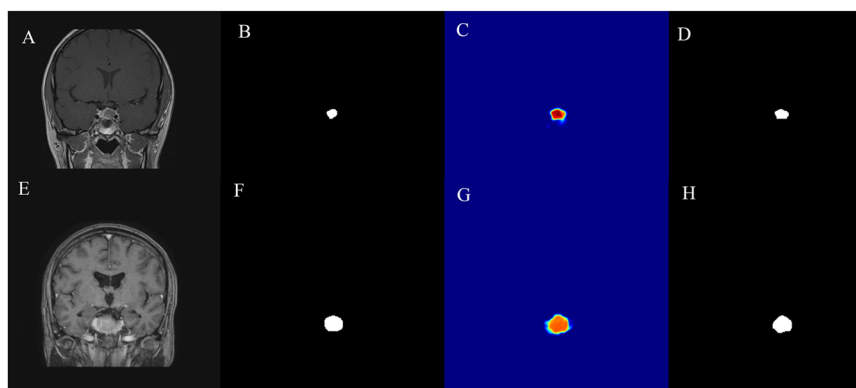
✉ Victor E. Staartjes
victoregon.staartjes@usz.ch

¹ Machine Intelligence in Clinical Neuroscience (MICN) Laboratory, Department of Neurosurgery, Clinical Neuroscience Center, University Hospital Zurich, University of Zurich, Zurich, Switzerland

² Department of Neurosurgery, University Hospital Zurich, Frauenklinikstrasse 10, 8091 Zurich, Switzerland

³ Department of Neuroradiology, University Hospital Zurich, Frauenklinikstrasse 10, 8091 Zurich, Switzerland

Fig. 1 Examples of images and predictions. A/E preoperative input image in coronal orientation, B/F ground truth, C/G unthresholded predictions, D/H thresholded predictions



that a CNN can generate segmentations of PA faster and more objective while maintaining quality of segmentation.

Methods

Data and preprocessing

Patients undergoing transsphenoidal surgery for PA at University Hospital Zurich in the period of October 2012 to May 2021 with available preoperative and 3-month postoperative 3-Tesla magnetic resonance imaging (MRI) were included. After identifying the closest T1-weighted contrast-enhanced MRI scan prior to surgery and a three month follow up those were assigned a Study ID and exported. In order to account for different manufacturers and acquisition protocols at referring hospitals, the images were converted to NIfTI format [16], reshaped to $256 \times 256 \times 256$ voxels, voxel size normalized to $1.0 \times 1.0 \times 1.0$, and images were reoriented using a right-anterior-superior affine matrix. Tumor tissue was subsequently manually labeled for training and residual volume assessment. After creating a holdout set of 20 patients for assessment of model generalization, the remaining 386 studies (two per patient) had its pixel intensities normalized for each study individually, and were then sliced in the coronal plane [17].

Model Development

We used a 2D-U-Net as model architecture, with a binary cross entropy loss function, Adam optimizer, and a sigmoid activation function [15]. It was built using the following platforms: Python 3.9.0 [18], Keras 2.5.0 [19], SimpleITK [20–22] and nibabel [23]. Training was carried out on a Nvidia RTX 3090 graphical processing unit. Separate preoperative and postoperative models were trained in five-fold cross validation.

The five resulting models were subsequently used to create ensemble segmentations by averaging their respective predictions. To binarize the predicted probabilities ranging from zero to one, a threshold of 0.6 for preoperative and 0.44 for postoperative scans was used as illustrated in Fig. 1. For the postoperative predictions, the following automated postprocessing steps were implemented: Coherent volumes smaller than fifty pixels were removed, holes within the segmentation were filled and a dilate function was used to smoothen the corners which is closer to natural tumor growth.

For the postoperative models, we additionally implemented transfer learning by initializing the postoperative models with the parameters of the fully trained preoperative models, and image augmentation was performed with a sampling ratio of 1/255 and rotation of between 0 and 90 degrees as well as zoom of 0% to 30%.

Evaluation

Manual and automated segmentations were compared using the Dice score, Jaccard score and the 95th percentile of the Hausdorff distance [24–27]. Dice and Jaccard evaluate similarity and overlap and range from zero—indicating no overlap—to one for perfect congruence. The Hausdorff distance analyzes the distances between two sets of points made up from the edges of two segmentations. Smaller values thus represent better performance. We opted for the 95th percentile instead of maximal Hausdorff distance to decrease the importance of outliers. Volumes were calculated in mm^3 from the segmented masks, and their correlation with the manually segmented volumes was assessed using Pearson's product-moment correlation. Automated and manual EOR were similarly correlated. Finally, we assessed the model's performance in detecting gross total resection (GTR, defined as an EOR of 100%) using a confusion matrix.

Table 1 Summary of the patient, tumor and radiological characteristics

Variable	Cohort	
	Test / Train Set	Holdout Set
Baseline		
Number of Patients	193	20
Age years, mean \pm SD	55 \pm 19	52 \pm 20
Male gender, n (%)	107 (55)	13 (65)
Biochemical activity		
NFPA, n (%)	113 (58.55)	15 (0.75)
GH, n (%)	45 (23.32)	4 (20)
PRL, n (%)	19 (9.84)	0 (0)
ACTH, n (%)	8 (4.14)	0 (0)
Others, n (%)	8 (4.14)	1 (5)
ZPS score		
ZPS I, n (%)	61 (31.61)	0 (0)
ZPS II, n (%)	82 (42.49)	9 (45)
ZPS III, n (%)	31 (16.06)	8 (40)
ZPS IV, n (%)	11 (5.70)	1 (5)
Score missing, n (%)	8 (4.14)	2 (10)
Knosp score		
Knosp 0, n (%)	44 (22.80)	1 (5)
Knosp 1, n (%)	43 (22.28)	7 (0.35)
Knosp 2, n (%)	47 (24.35)	6 (0.3)
Knosp 3, n (%)	35 (18.13)	4 (20)
Knosp 4, n (%)	16 (8.29)	1 (5)
Score missing, n (%)	8 (4.14)	1 (5)
EOR		
EOR mean \pm SD	92.84 \pm 29.21	97.77 \pm 3.60
EOR median (IQR)	100.00 (4.18)	99.57 (3.04)
Tumor volume		
Preoperative, mean \pm SD (ml)	7.42 \pm 9.93	9.12 \pm 8.86
Preoperative, median (IQR) (ml)	4.79 (7.1)	6.37 (7.75)
Postoperative, mean \pm SD (ml)	0.35 \pm 1.61	0.67 \pm 0.84
Postoperative, median (IQR) (ml)	0 (0.05)	0.06 (1.36)
MRI data		
Number of scans	386	40
Vendor		
Siemens	171	0
Philips Healthcare	144	36
GE Medical Systems	71	4
Acquisition orientation		
2D	317	31
3D	69	9
Field strength		
1 T	1	1
1.5 T	45	6
3 T	340	33
Voxel dimensions		
Pixel Spacing, mean \pm SD (mm)	0.47 \pm 0.14	0.51 \pm 0.14
Pixel Spacing, median (IQR) (mm)	0.45 (0.19)	0.47 (0.09)
Slice Thickness, mean \pm SD (mm)	1.98 \pm 0.53	2.02 \pm 0.67
Slice Thickness, median (IQR) (mm)	2 (0)	2 (0)

SD standard deviation, *IQR* interquartile range, *NFPA* non functioning pituitary adenoma, *GH* growth hormone, *ACTH* adrenocorticotropic hormone, *PRL* prolactin, *GTR* gross total resection, *T* Tesla

Results

Cohort

In total, 213 patients were included retrospectively, of which 193 were applied for development of the model. Validation was performed in 20 held-out patients. Summary demographics and radiological information are displayed in Table 1.

Preoperative segmentation performance

Segmentation performance is summarized in Table 2. In terms of preoperative tumor segmentation, our ensemble model achieved a mean Dice score of 0.62 ± 0.22 , with automatically rated volumes correlating well with manually segmented volumes ($r = 0.85$). Figure 2 shows metric performance and tumor volume with a linear regression. Exact Jonckheere-Terpstra-Test for a trend did not reach significance (JT, p value for Dice score: 110, 0.1757; JT, p value for Jaccard score: 114, 0.1166) [28].

Postoperative segmentation performance

For postoperative segmentations, a mean Dice score of 0.046 ± 0.125 was observed. The correlation of manually segmented tumors and predicted tumor masks correlated less satisfactory than in preoperative scans ($r = 0.22$). Introduction of transfer learning techniques and image augmentation did not improve performance (Table 3).

Resection assessment performance

Table 4 summarizes performance in terms of resection assessment. Our model's predictions generated EOR values that only poorly correlated with manual segmentations ($r = -0.14$). Automatically detected EOR differed from the ground truth manual segmentation by $18.65\% \pm 31.10\%$ on average. GTR was detected with an accuracy of 50.00%, sensitivity of 66.67% and specificity of 36.36%.

Discussion

We have developed and validated an automated PA segmentation pipeline based on deep learning. We demonstrate that our approach performs favorably when it comes to segmentation and volumetric assessment of preoperative images. Generating objective and precise postoperative segmentations of residual tumor remains a challenge, even with the application of advanced machine learning techniques.

Table 2 Segmentation performance of the fully trained preoperative and postoperative models

Metric	Preoperative U-Net		Postoperative U-Net	
	Resampled training	Holdout set	Resampled training	Holdout set
Dice				
Mean ± SD	0.50 ± 0.05	0.62 ± 0.22	0 ± 0	0.046 ± 0.125
Median (IQR)	0.49 (0.05)	0.71 (0.27)	0 (0)	0 (0)
Jaccard				
Mean ± SD	0.43 ± 0.05	0.53 ± 0.21	0 ± 0	0.030 ± 0.085
Median (IQR)	0.43 (0.05)	0.61 (0.30)	0 (0)	0 (0)
95th percentile Hausdorff distance				
Mean ± SD	6.55 ± 0.70.	3.89 ± 1.96	20.810 ± 14.604	12.199 ± 6.684
Median (IQR)	4.50 (1.13)	3.39 (1.66)	15.769 (12.152)	0 (6.655)

Resampled training performance as well as generalization towards the held-out data is reported
SD standard deviation, *IQR* interquartile range

Fig. 2 Metric performance and manually segmented tumor volume of the holdout set with a linear regression

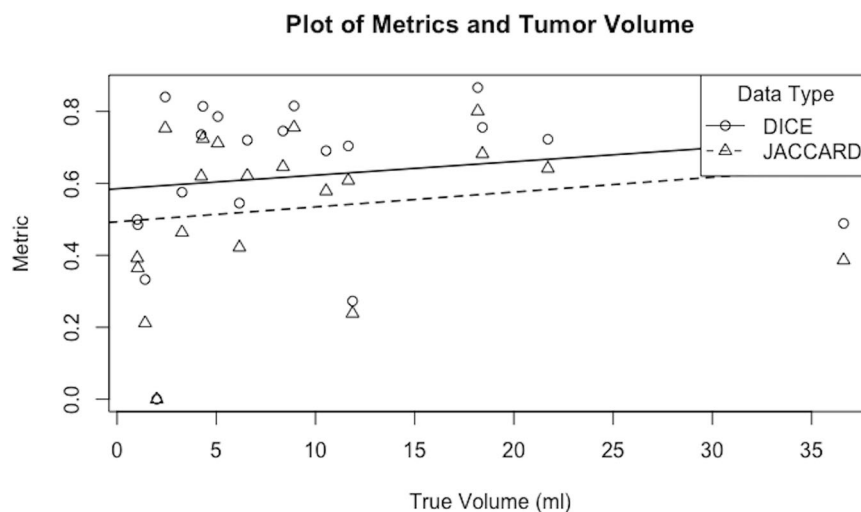


Table 3 Image augmentation and transfer learning based on the preoperative models were applied as an attempt to improve postoperative segmentation performance

Metric	Postoperative U-Net + Augmentation	Postoperative U-Net + Transfer Learning	Postoperative U-Net + Augmentation + Transfer Learning
Performance Type	Holdout Set	Holdout Set	Holdout Set
Dice			
Mean ± SD	0 ± 0	0.040 ± 0.087	0.013 ± 0.035
Median (IQR)	0 (0)	0 (0.007)	0 (0.007)
Jaccard			
Mean ± SD	0 ± 0	0.026 ± 0.062	0.008 ± 0.022
Median (IQR)	0 (0)	0 (0.004)	0 (0.004)
95th percentile Hausdorff distance			
Mean ± SD	NaN	15.051 ± 8.131	14.876 ± 5.021
Median (IQR)	NaN	13.045 (7.836)	15.953 (8.050)

Generalization towards the held-out data is reported
SD standard deviation, *IQR* interquartile range

As neuroimaging has become much more frequent, also detection of incidentalomas is prone to increase since especially nonfunctioning incidentaloma of the pituitary are

highly prevalent (1.4–27% in autopsy and 3.7–37% in imaging) [29]. Hence automated segmentation of incidental PA would tackle a frequent issue. Incorporating this into a

Table 4 Volumetric performance of the automated resection assessment pipeline

Measurement	U-Net Performance vs. Manual Segmentation
Preoperative Volume [mm ³]	
Correlation (Pearson)	0.85
Postoperative Volume [mm ³]	
Correlation (Pearson)	0.22
EOR [%]	
Correlation (Pearson)	−0.14
Difference in EOR [Median (IQR)]	3.31% (17.75%)
Difference in EOR [Mean ± SD]	18.65% ± 31.10%
GTR	
Accuracy	50.00%
Sensitivity	66.67%
Specificity	36.36%
PPV	46.15%
NPV	57.14%

EOR extent of resection, *GTR* gross total resection, *PPV* positive predictive value, *NPV* negative predictive value

SD standard deviation, *IQR* interquartile range

diagnostic software to detect suspicious lesions of the pituitary gland would be valuable for clinical routine. To address this, we developed a fully automated graphical user interface that tackles this issue (<https://micnlab.com/download-the-zurich-neurosurgical-toolkit/>).

Furthermore, automated, objective volumetric measurement for assessing progression of incidentalomas, especially microadenomas, yields a clinical benefit since volumetric progression is crucial for surgical indication [30]. Last, prognostic scores like the Zurich pituitary or Knosp score could be automatically implemented, ultimately giving the clinician further information and saving his time. These three options combined have the potential to standardize and speed up the clinical workflow of PA.

When indicated, the transsphenoidal approach is very effective in most cases of PA with comparatively low surgical morbidity and mortality [3]. When evaluating surgical oncology results—not only in individual patients, but also when comparing cohorts, surgeons, and departments and for research purposes—volumetric assessment of residual adenoma volume is of paramount importance. However, segmenting tumors on each slice accurately is time-consuming and often not possible in daily clinical practice. Furthermore, as in other tumor segmentation tasks such as gliomas, the interrater agreement likely is low, especially for postoperative residual tumor tissue [10, 31]. In this light, it must be considered that even morphological grading at the preoperative stage using the Hardy and Knosp classification

already suffers from relatively low interrater agreement [8, 9].

Volumetric rating of post-resection sellar scans thus presents a particular challenge when it comes to objectivity. Automated semantic segmentation – the machine learning task that deals with detecting and outlining structures on images—could prove a viable option to improve the speed and objectivity with which volumes are assessed pre- and postoperatively.

To some extent, the poor performance on postoperative images is to be expected: In the end, supervised learning techniques can only ever be as good as the “ground truth” data they are trained on, and with disagreements in labeling of small or debatable residual tumor, this has certainly been the case in this study. Even in the much more intensely studied task of glioma and glioblastoma segmentation, which has been fueled by the yearly international BraTS challenges [32], performance overall appears mediocre and demonstrates that – especially for low grade glioma—it appears to be difficult to generate any improvements in segmentation compared to human raters, apart from the increased speed and objective nature with which automated segmentations can be produced. Even in BraTS 2014/2015, where postoperative images were also segmented, the ground truth labels for postoperative images eventually had to be generated by learning algorithms.

Our ensemble method is—to the best of the authors’ knowledge—the only attempt at automatically segmenting post-transsphenoidal surgery scans. The deep learning approach was able to accurately outline the tumors preoperatively, but struggled with small residual tumors. Certainly, this can be explained at least partially by the computationally necessary downsampling of the images, resulting at times in voxels that appear almost equally as large as the residual tumor itself. Also, pituitary adenomas do not appear constantly with the same relative intensity making them hard to identify at times [33].

Limitations

Although we included scans from many institutions and all major scanner manufacturers, our study remains single-center and external validation would be necessary before any kind of clinical application. Furthermore, we applied 2D segmentation—which has demonstrated reliable results previously for other similar applications—although 3D segmentation potentially could further increase performance. A larger dataset of subtotally resected pituitary adenomas would most likely also allow improvements, since most postoperative images showed no tumor for the model to learn to recognize. A more heterogeneous dataset as we did allows for better generalization and reduces the risk of overfitting to a particular manufacturer or hospital

protocol. On the other hand, a reduction in performance is to be expected with this approach.

Conclusion

Our volumetry pipeline demonstrated its ability to accurately and automatically segment pituitary adenoma. This is highly valuable for lesion detection and evaluation progression of pituitary incidentalomas. Postoperatively, however, objective and precise detection of residual tumor remains less successful. Larger datasets, more diverse data, and more elaborate modeling could potentially improve performance. Yet, focusing on use cases for preoperative segmentations seems more promising.

Data availability

The data in support of our findings can be obtained upon reasonable request from the corresponding author. The models can be downloaded and applied using our graphical user interface, available at “<https://micnlab.com/download-the-zurich-neurosurgical-toolkit/>”.

Acknowledgements The results of this paper have been presented at the EANS 2022 Congress in Belgrade by Raffaele Da Mutten in a seven-minute presentation on the 18th of October 2022. Consequently, the abstract is available under: <https://doi.org/10.1016/j.bas.2022.101287>.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by R.D.M. and O.C.-C. The code was written and the results were analyzed by R.D.M., O.Z. and V.E.S. The first draft of the manuscript was written by R.D.M. and all authors commented on it. All authors read and approved the final manuscript.

Funding Open access funding provided by University of Zurich.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Ethics approval The ethics committee of the canton of Zurich approved this project under the BASEC number 2021-01147. No identifying images or data are contained in this text. In this retrospective work, no additional consent was obtained. Before the procedure, consent to the surgery was given.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this

article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. K.A. McNeill, Epidemiology of Brain Tumors. *Neurologic Clin.* **34**, 981–998 (2016). <https://doi.org/10.1016/j.ncl.2016.06.014>
2. R. Pratheesh, S. Rajaratnam, K. Prabhu et al. The current role of transcranial surgery in the management of pituitary adenomas. *Pituitary* **16**, 419–434 (2013). <https://doi.org/10.1007/s11102-012-0439-z>
3. F.G. Barker, A. Klibanski, B. Swearingen, Transsphenoidal surgery for pituitary tumors in the United States, 1996–2000: mortality, morbidity, and the effects of hospital and surgeon volume. *J. Clin. Endocrinol. Metab.* **88**, 4709–4719 (2003). <https://doi.org/10.1210/jc.2003-030461>
4. V.E. Staartjes, C. Serra, M. Zoli et al. Multicenter external validation of the Zurich Pituitary Score. *Acta Neurochir.* **162**, 1287–1295 (2020). <https://doi.org/10.1007/s00701-020-04286-w>
5. A.S.G. Micko, A. Wöhrer, S. Wolfsberger, E. Knosp, Invasion of the cavernous sinus space in pituitary adenomas: Endoscopic verification and its correlation with an MRI-based classification. *J. Neurosurg.* **122**, 803–811 (2015). <https://doi.org/10.3171/2014.12.JNS141083>
6. M. Buchfelder, S. Schlaffer, Surgical treatment of pituitary tumours. *Best. Pract. Res. Clin. Endocrinol. Metab.* **23**, 677–692 (2009). <https://doi.org/10.1016/j.beem.2009.05.002>
7. C. Serra, J.-K. Burkhardt, G. Esposito et al. Pituitary surgery and volumetric assessment of extent of resection: a paradigm shift in the use of intraoperative magnetic resonance imaging. *FOC* **40**, E17 (2016). <https://doi.org/10.3171/2015.12.FOCUS15564>
8. M.A. Mooney, D.A. Hardesty, J.P. Sheehy et al. Interrater and intrarater reliability of the Knosp scale for pituitary adenoma grading. *J. Neurosurg.* **126**, 1714–1719 (2016). <https://doi.org/10.3171/2016.3.JNS153044>
9. M.A. Mooney, D.A. Hardesty, J.P. Sheehy et al. Rater reliability of the hardy classification for pituitary adenomas in the magnetic resonance imaging era. *J. Neurol. Surg. B Skull Base* **78**, 413–418 (2017). <https://doi.org/10.1055/s-0037-1603649>
10. M. Visser, D.M.J. Müller, R.J.M. van Duijn et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. *NeuroImage: Clin.* **22**, 101727 (2019). <https://doi.org/10.1016/j.nicl.2019.101727>
11. Z. Akkus, A. Galimzianova, A. Hoogi et al. Deep learning for brain MRI segmentation: State of the art and future directions. *J. Digit. Imaging* **30**, 449–459 (2017). <https://doi.org/10.1007/s10278-017-9983-4>
12. Z. Obermeyer, E.J. Emanuel, Predicting the future — big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **375**, 1216–1219 (2016). <https://doi.org/10.1056/NEJMp1606181>
13. B.F. King, artificial intelligence and radiology: What will the future hold. *J. Am. Coll. Radiol.* **15**, 501–503 (2018). <https://doi.org/10.1016/j.jacr.2017.11.017>
14. J. Egger, T. Kapur, C. Nimsky, R. Kikinis, Pituitary Adenoma Volumetry with 3D Slicer. *PLoS ONE* **7**, e51788 (2012). <https://doi.org/10.1371/journal.pone.0051788>
15. Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv: 150504597 [cs]

16. X. Li, P.S. Morgan, J. Ashburner et al. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J. Neurosci. Methods* **264**, 47–56 (2016). <https://doi.org/10.1016/j.jneumeth.2016.03.001>
17. Zettler N., Mastmeyer A. (2021) Comparison of 2D vs. 3D U-Net Organ Segmentation in abdominal 3D CT images
18. Van Rossum G., Drake Jr F.L. (1995) Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam
19. Chollet F. (2015) Keras. <https://github.com/fchollet/keras>
20. Beare R., Lowekamp B., Yaniv Z. (2018) Image Segmentation, Registration and Characterization in R with SimpleITK. *J Stat Soft* **86**. <https://doi.org/10.18637/jss.v086.i08>
21. Z. Yaniv, B.C. Lowekamp, H.J. Johnson, R. Beare, SimpleITK Image-Analysis Notebooks: a collaborative environment for education and reproducible research. *J. Digit Imaging* **31**, 290–303 (2018). <https://doi.org/10.1007/s10278-017-0037-8>
22. Lowekamp B.C., Chen D.T., Ibáñez L., Blezek D. (2013) The Design of SimpleITK. *Front Neuroinform* **7**. <https://doi.org/10.3389/fninf.2013.00045>
23. Brett, M., Markiewicz Christopher J., Hanke, M., et al (2020) nipy/nibabel: 3.2.1
24. A.A. Taha, A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* **15**, 29 (2015). <https://doi.org/10.1186/s12880-015-0068-x>
25. L.R. Dice, Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302 (1945). <https://doi.org/10.2307/1932409>
26. P. Jaccard, The distribution of the flora in the alpine zone. I. *N. Phytol.* **11**, 37–50 (1912). <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
27. A. Ralescu, Probability and fuzziness. *Inf. Sci.* **34**, 85–92 (1984). [https://doi.org/10.1016/0020-0255\(84\)90018-5](https://doi.org/10.1016/0020-0255(84)90018-5)
28. A.R. Jonckheere, A distribution-free k-sample test against ordered alternatives. *Biometrika* **41**, 133 (1954). <https://doi.org/10.2307/2333011>
29. Thapar K., Kovacs K., Scheithauer B., Lloyd R.V. (2001) Diagnosis and management of pituitary tumors. Humana Press
30. F. Galland, M.-C. Vantyghem, L. Cazabat et al. Management of nonfunctioning pituitary incidentaloma. *Annales d'Endocrinologie* **76**, 191–200 (2015). <https://doi.org/10.1016/j.ando.2015.04.004>
31. O. Rodríguez, B. Mateos, R. de la Pedraja et al. Postoperative follow-up of pituitary adenomas after trans-sphenoidal resection: MRI and clinical correlation. *Neuroradiology* **38**, 747–754 (1996). <https://doi.org/10.1007/s002340050341>
32. B.H. Menze, A. Jakab, S. Bauer et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med Imaging* **34**, 1993–2024 (2015). <https://doi.org/10.1109/TMI.2014.2377694>
33. K. Karimian-Jazi, Hypophysentumoren. *Radiologe* **59**, 982–991 (2019). <https://doi.org/10.1007/s00117-019-0570-1>