



# Simultaneous brain structure segmentation in magnetic resonance images using deep convolutional neural networks

Tomoko Maruyama<sup>1,2</sup> · Norio Hayashi<sup>3</sup> · Yusuke Sato<sup>2,4</sup> · Toshihiro Ogura<sup>3</sup> · Masumi Uehara<sup>3</sup> · Akio Ogura<sup>3</sup> · Haruyuki Watanabe<sup>3</sup> · Yoshihiro Kitoh<sup>1</sup> · for the Alzheimer's Disease Neuroimaging Initiative

Received: 23 December 2020 / Revised: 24 July 2021 / Accepted: 28 July 2021  
© Japanese Society of Radiological Technology and Japan Society of Medical Physics 2021

## Abstract

In brain magnetic resonance imaging (MRI) examinations, rapidly acquired two-dimensional (2D) T1-weighted sagittal slices are typically used to confirm brainstem atrophy and the presence of signals in the posterior pituitary gland. Image segmentation is essential for the automatic evaluation of chronological changes in the brainstem and pituitary gland. Thus, the purpose of our study was to use deep learning to automatically segment internal organs (brainstem, corpus callosum, pituitary, cerebrum, and cerebellum) in midsagittal slices of 2D T1-weighted images. Deep learning for the automatic segmentation of seven regions in the images was accomplished using two different methods: patch-based segmentation and semantic segmentation. The networks used for patch-based segmentation were AlexNet, GoogLeNet, and ResNet50, whereas semantic segmentation was accomplished using SegNet, VGG16-weighted SegNet, and U-Net. The precision and Jaccard index were calculated, and the extraction accuracy of the six convolutional network (DCNN) systems was evaluated. The highest precision (0.974) was obtained with the VGG16-weighted SegNet, and the lowest precision (0.506) was obtained with ResNet50. Based on the data, calculation times, and Jaccard indices obtained in this study, segmentation on a 2D image may be considered a viable and effective approach. We found that the optimal automatic segmentation of organs (brainstem, corpus callosum, pituitary, cerebrum, and cerebellum) on brain sagittal T1-weighted images could be achieved using SegNet with VGG16.

**Keywords** MRI · Deep learning · Brain · Convolutional neural network · Segmentation

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

✉ Tomoko Maruyama  
tmaruyama@shinshu-u.ac.jp; dr207102@gchs.ac.jp

<sup>1</sup> Division of Radiology, Shinshu University Hospital, 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan

<sup>2</sup> Department of Radiological Technology, Graduate School of Radiological Technology, Gunma Prefectural College of Health Sciences, 323-1 Kamioki, Maebashi, Gunma 371-0052, Japan

<sup>3</sup> Department of Radiological Technology, Gunma Prefectural College of Health Sciences, 323-1 Kamioki, Maebashi, Gunma 371-0052, Japan

<sup>4</sup> Department of Radiology, Gunma University Hospital, 3-39-5 Showamachi, Maebashi, Gunma 371-8511, Japan

## 1 Introduction

Magnetic resonance (MR) images are essential tools and indispensable for diagnosing many medical lesions. In particular, MR imaging (MRI) is excellent for diagnosing brain lesions because it can depict soft tissue with high contrast. MR brain images are often segmented for diagnosis and analysis. Image segmentation is one of the most active areas of research, and the automation of segmentation has recently emerged as a major challenge. When segmentation is performed using conventional semi-automatic methods, machine learning techniques such as random forest or k-means have been applied. In a previous segmentation study using the BRATS dataset (a common dataset for evaluating the accuracy of brain tumor segmentation techniques), Tustison reported that random forest provided the best performance [1].

Due to the increasing popularity of deep learning techniques, this approach has recently become a primary

candidate for automatic segmentation methods. Moreover, deep learning is widely used for classification problems that show limited success with traditional machine learning approaches, and also for detailed object recognition from images. In addition to its general applications, such as automatic recognition of faces and cars, deep learning has been applied in various ways to medical problems and has yielded high-quality results. Several studies have shown that deep learning improves segmentation accuracy in MR images [2–4].

Most previous studies on brain organ segmentation have been performed using three-dimensional assessment (3D images) [5, 6]. Mlynarski et al. segmented eight regions, including the pituitary gland and brainstem, from axial images using a dataset of 3D contrast-enhanced T1-weighted images and a network based on U-Net [7]. Chen et al. similarly used 3D images and U-Net to segment six regions, including the brainstem [8]. Segmentation using 3D images has produced satisfactory accuracy in previous studies because 3D images provide superior spatial robustness versus 2D images [9]. However, depending on the GPU specifications, assessments based on 3D images involve large amounts of data and prolonged calculation times [9]. Furthermore, to be clinically useful, 3D images of the brain often require longer scan times than 2D images [10].

Most routine clinical assessments performed with brain MRI scans use 2D images, which can be acquired in a short time. For example, 2D T1-weighted sagittal slices are typically used to confirm brainstem atrophy and the presence of signals in the posterior pituitary gland. Although sagittal images contain a substantial amount of information that is useful for diagnosing brain stem and pituitary gland lesions, few studies have focused on these images. The available datasets for deep learning studies of brain segmentation mainly contain axial images [7, 8]. Using conventional methods without deep learning, Lee et al. used thresholding to separate the cerebrum (the largest region on the sagittal images), from other brain regions in a 3D image [11]. In addition, Rohni et al. reported that it is possible to classify Alzheimer's disease by segmenting the brainstem from a 2D sagittal image, and that 2D images can be segmented relatively simply, quickly, and easily [12]. Therefore, we might expect that these types of segmentations could be more easily and accurately performed using deep learning, and would consequently be useful in clinical practice. For example, establishing a computerized technique to evaluate the pituitary gland and brainstem objectively and automatically would reduce the interpretation load for radiologists making such assessments.

Accordingly, the purpose of our study was to use deep learning to automatically recognize internal organs (brainstem, corpus callosum, pituitary, cerebrum, and cerebellum) on one midsagittal slice of 2D T1-weighted images. By

applying several networks for segmentation, we investigated the most suitable network for simultaneous segmentation.

## 2 Methods

### 2.1 Overview

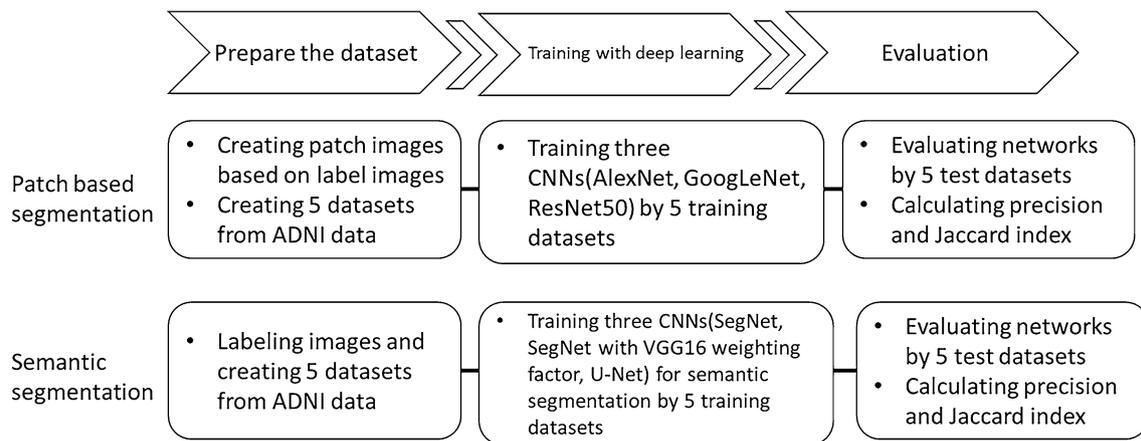
T1-weighted midsagittal images were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)) [13]. ADNI is a database created to test whether the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) can be measured by combining MRI, positron emission tomography (PET), and clinical assessment (for up-to-date information, see [www.adni-info.org](http://www.adni-info.org)). Deep learning for automatic segmentation of seven regions (brainstem, corpus callosum, pituitary, cerebrum, cerebellum, other, and air) was performed on the images using two different methods: patch-based segmentation and semantic segmentation. The networks used for the patch-based segmentation method were AlexNet [14], GoogLeNet [15], and ResNet50 [16]. Three types of semantic segmentation methods were used: SegNet [17], SegNet with a VGG16-weighting factor [18], and U-Net [19]. The precision and Jaccard index were calculated after segmentation was completed, and the extraction accuracy of the six DCNN systems was evaluated. A flowchart of the study protocol is shown in Fig. 1. MATLAB R2019a 9.6.0 (MathWorks, Massachusetts, USA) software was used for the analysis. The graphics processing unit (GPU) was a GEFORCE GTX 1070-TI (NVIDIA, California, USA).

### 2.2 Dataset

T1-weighted brain images were obtained from the ADNI database. Midsagittal slices were reconstructed from 3D axial image data. The number of original images was 450, and the image size was  $256 \times 256$  pixels. Five training and test datasets were created by changing the image combinations. The 450 images were divided into 400 training images and 50 test images. Five sets of 50 test images were selected from the 450 images so that the images did not overlap. The remaining 400 images were used as the training images.

### 2.3 Region segmentation by deep learning

Region segmentation using deep learning can be performed using two different approaches. In patch-based segmentation, multiple patch images are extracted from one image, and a patch set is learned by classification networks, such as AlexNet, VGG16, or GoogLeNet. Patch-based methods have proven effective for labeling brain structures [20, 21]. Segmentation can be performed by returning the trained



**Fig. 1** A flowchart of the method outline is shown below. Deep learning for automatic recognition of the different regions in images was performed using two different methods: patch-based segmentation

and semantic segmentation. After segmentation, the precision and Jaccard index were calculated, and the extraction accuracy of the six dynamic conditional convolutional network systems was evaluated

network's judgment based on the patches of the test images to the original image. The second approach is semantic segmentation [22, 23], which is the predominant method used for segmentation. Networks such as SegNet and U-Net were developed for semantic segmentation. These architectures can classify all the pixels that constitute the image and determine the regions to which they belong.

### 2.3.1 Patch-based segmentation

- 1) Patch images of  $61 \times 61$  pixels were extracted from the original images of  $256 \times 256$  pixels. The patch image interval was set to seven pixels. The optimal patch size was selected by considering the maintenance of information around the target organ and the number of images for segmentation.
- 2) Based on the central pixel, the patch images were classified into seven categories (brainstem, corpus callosum, pituitary gland, cerebrum, cerebellum, other, and air) and stored in their respective folders. When all patch images extracted from the original training images were used, we found a 200-fold difference between the pituitary region with the smallest number and other regions with the largest number. We determined that this result impeded efficient learning; therefore, we used 1000–2000 random patch images for each category (brainstem, corpus callosum, pituitary gland, cerebrum, cerebellum, other, and air). The patch images were randomly selected from all extracted patches in each category using a computer program. However, the pituitary area yielded fewer than 1000 patch images; hence, all images obtained for this region were used. All the patch images extracted from the test dataset images were evaluated.

- 3) The three networks (AlexNet, GoogLeNet, and ResNet50) that were previously demonstrated effective for classification were fine-tuned using five training datasets. The pre-trained networks AlexNet, GoogLeNet, and ResNet50 were trained with tens of thousands of general images and distributed using MATLAB.
- 4) The trained networks were evaluated using the five datasets. The numbers of patch images in the five datasets are listed in Table 1.

### 2.3.2 Semantic segmentation

- 1) The supervised images were created by manually labeling seven regions (brainstem, corpus callosum, pituitary, cerebrum, cerebellum, other, and air) on the original image. Supervised images were created manually using a tool called Image Labeler, which is provided in MATLAB. Image Labeler is an application for defining pixel ROI labels that pair with images. The labeling of supervised images was confirmed by a Radiographer with more than 20 years of experience. Augmentation (random shift of 10 pixels in the x and y directions and scaling of 0.8- and 1.2 times) was performed to compensate for the small number of trainings with 400 images.
- 2) The five training datasets were trained on a segmentation network (SegNet, U-Net, and SegNet with a VGG16-weighting factor). Pre-training of SegNet and U-Net was not performed. SegNet with a VGG16-weighting factor is a network in which the initialized weights of a pre-trained VGG16 are applied to the SegNet encoder.
- 3) The trained networks were evaluated using five test datasets.

**Table 1** The number of patch images in the dataset for patch-based segmentation

		Brainstem	Corpus Callosum	Pituitary	Cerebrum	Cerebellum	Other	Air
Training ( <i>n</i> = 400)	Set 1	1569	1575	1061	1465	1510	1887	1819
	Set 2	1619	1574	1109	1585	1369	1833	1681
	Set 3	1655	1509	1048	1594	1544	1816	1648
	Set 4	1517	1312	1040	1558	1437	1735	1720
	Set 5	1542	1512	1055	1423	1624	1452	1580
Test ( <i>n</i> = 50)	Set 1	1461	832	168	10,341	1508	20,871	4019
	Set 2	1467	774	120	10,156	1480	21,784	3419
	Set 3	1550	979	181	10,662	1505	21,154	3169
	Set 4	1389	1030	189	10,289	1516	20,656	4131
	Set 5	1417	902	174	10,308	1447	20,933	4017

**Table 2** Learning parameters of MATLAB for each network

	Max epochs	Mini batch size	Initial learn rate
AlexNet	5	8	0.0001
GoogLeNet	5	8	0.0001
ResNet50	5	8	0.0001
SegNet	100	4	0.001
SegNet with VGG16-weighting factor	100	4	0.001
U-Net	100	4	0.05

## 2.4 Parameters for deep learning

Table 2 lists the parameters used for the network learning. The optimization function utilized a stochastic gradient descent for all networks. The stochastic gradient descent method optimizes and updates the parameters sequentially. The number of epochs, mini-batch size, and initial learning rate were adjusted on the basis of the standard values specified in MATLAB to fit the datasets used in this experiment.

**Table 3** The average precision values for each network

	Brainstem	Corpus Callosum	Pituitary	Cerebrum	Cerebellum	All image
AlexNet	0.693	0.566	0.350	0.811	0.740	0.632
GoogLeNet	0.663	0.468	0.271	0.919	0.727	0.610
ResNet50	0.499	0.341	0.147	0.896	0.645	0.506
SegNet	0.973	0.958	0.979	0.917	0.978	0.961
SegNet with VGG16-weighting factor	0.981	0.972	0.984	0.958	0.983	0.976
U-Net	0.880	0.584	0.001	0.746	0.626	0.567

## 2.5 Evaluation

The test dataset was evaluated using 50 images. The overall classification accuracy was evaluated using the confusion matrix, and the precision and Jaccard index were calculated. The precision (1) and Jaccard index (2) were obtained using the following equations,

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (1)$$

$$\text{Jaccard index} = \frac{TP}{TP + FP + FN} \quad (2)$$

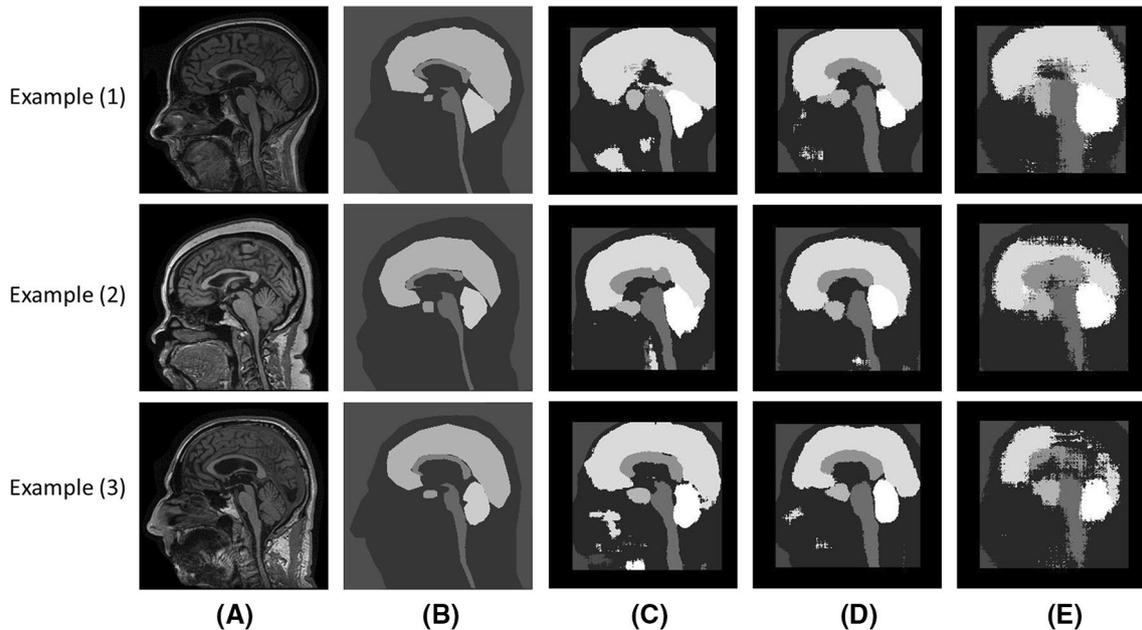
where TP, FP, and FN are true positive, false positive, and false negative.

## 3 Results

Table 3 lists the precision values, and Table 4 lists the Jaccard indices of the six networks. Both are the average values for the learning and evaluation of the five datasets. Typical images of the segmentation results are shown in Figs. 2, 3. The indicated images were randomly chosen and tested using trained networks. The network with the

**Table 4** The average Jaccard indices in each network

	Brainstem	Corpus Callosum	Pituitary	Cerebrum	Cerebellum	All image
AlexNet	0.652	0.538	0.306	0.850	0.673	0.604
GoogLeNet	0.649	0.459	0.269	0.847	0.703	0.585
ResNet50	0.678	0.517	0.310	0.865	0.733	0.481
SegNet	0.637	0.599	0.333	0.861	0.663	0.618
SegNet with VGG16-weighting factor	0.681	0.652	0.415	0.894	0.770	0.682
U-Net	0.613	0.512	0.001	0.727	0.376	0.446

**Fig. 2** Three examples of patch-based segmentation results are shown. **A** is the original image and **(B)** is the teacher image corresponding to the original image, whereas **(C, D, and E)** are the result images obtained with AlexNet, GoogLeNet, and ResNet50, respectively

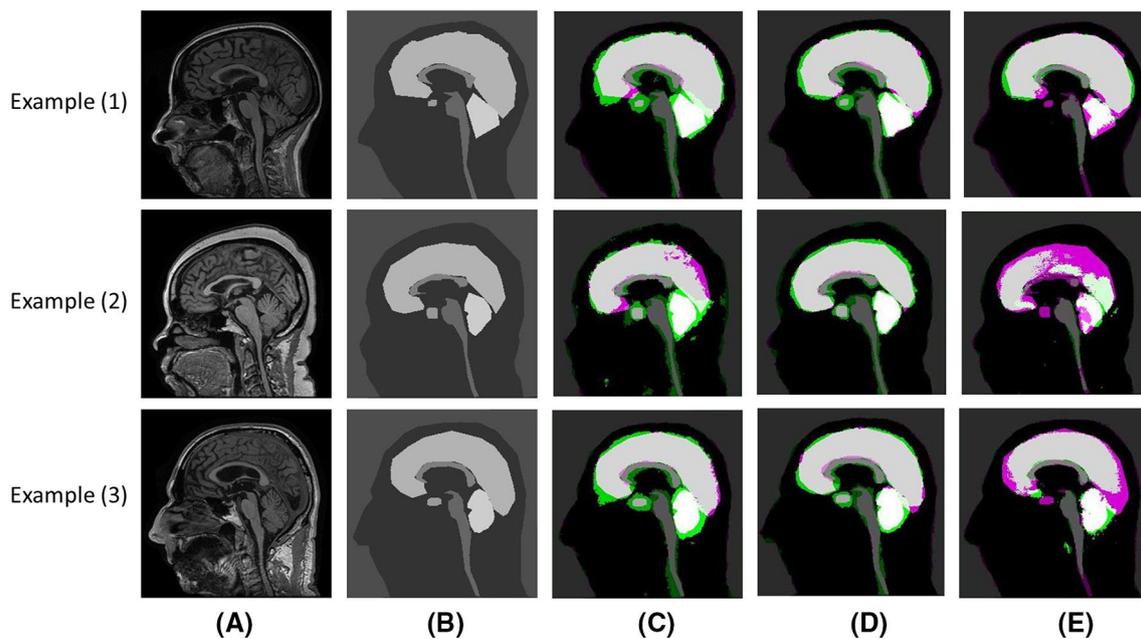
highest precision (0.974) was SegNet with the VGG16-weighting factor, whereas the lowest precision (0.506) was obtained with ResNet50. Assessments for the other regions demonstrated the highest precision, whereas those for the pituitary region demonstrated the lowest precision. The network with the highest Jaccard index (0.682) was SegNet with the VGG16-weighting factor, and the network with the lowest Jaccard index (0.446) was U-Net. The air region exhibited the highest Jaccard index, whereas the pituitary region exhibited the lowest Jaccard index.

The training time was approximately 20 min for AlexNet, 65 min for GoogLeNet, 120 min for ResNet50, 55 min for SegNet, 70 min for U-Net, and 90 min for SegNet with the VGG16-weighting factor. AlexNet had eight layers, GoogLeNet had 22 layers, and ResNet50 had 50 layers. The processing time increased depending on the model size. SegNet with VGG16 set as the encoder

exhibited a longer processing time than SegNet without VGG16 and U-Net.

## 4 Discussion

Among the methods compared in this study, SegNet with the VGG16-weighting factor demonstrated the highest precision and Jaccard index. VGG16 uses a pre-trained network and applies its weight to the SegNet encoder utility. The transfer learning function of VGG16 can be used to obtain high scores, even with a small amount of learning data. This finding is consistent with a previous study [24], in which Swati et al. used pre-trained VGG19 to discriminate brain tumor types in MR images and obtained higher accuracy than that shown by other methods. Guo et al. [25] also reported that applying Vgg weights to SegNet for polyp segmentation



**Fig. 3** Three examples of semantic segmentation results are shown. **A** is the original image and **(B)** is the teacher image corresponding to the original image. **C**, **D**, **E** are the result images obtained with SegNet, SegNet with VGG16-weighting factor, and U-Net

improved its accuracy. SegNet and U-Net, which have not been pre-trained, may not be able to adapt to a small number of trainings.

In assessments based on patch segmentation, no significant differences were observed in the precision and Jaccard index when AlexNet and GoogLeNet were used, and ResNet50 demonstrated lower values than the other two networks. Generally, the deeper the network, the better the extraction accuracy. However, in studies using medical images, networks with a shallower layer yielded superior results when compared to those with a deeper layer. Saikia et al. compared networks for automatic diagnosis based on fine-needle aspiration cytology (FNAC) findings. In their study, GoogLeNet, which has the shallowest layer among the networks considered in this study, yielded higher accuracy than ResNet50 and VGG16 [26]. In addition, Lee et al. reported a comparative study of networks for automatically distinguishing AWLwvf (angiomyolipoma without visible fat) from renal cell carcinoma (RCC). They found that AlexNet demonstrated the highest accuracy, followed by VGGNet, ResNet, and GoogLeNet [27]. Moreover, in that study, AlexNet had the smallest number of layers but the best score, and both precision and Jaccard index tended to decrease as the number of middle layers increased. Deep-layer networks require sufficient training data to learn features. The original data number in our study was 450 cases, while the original data numbers in the studies carried out by Saikia et al. [26] and Lee et al. [27] were 212 and 80 cases, respectively. Thus, when the number of cases is

small, networks with shallow layers may be more suitable for appropriate learning of features than networks with deep layers. In other words, it is necessary to use a suitable network for a given dataset.

In our study, the precision and Jaccard indices of the pituitary region were the lowest of all regions in all six networks. The pituitary region is the smallest visible organ and constitutes the lowest percentage of pixels in the image. Previous studies have reported similar results. Mlynarski et al. adapted the U-Net network to their dataset and segmented eight regions: eye, lens, optic nerve, optic chiasm, pituitary gland, hippocampus, brainstem, and brain (including cerebrum, cerebellum, and brainstem). Although the comparison is not straightforward because segmentation in that study was based on 3D axial images, the region with the highest Dice score was the brain (including the cerebrum, cerebellum, and brainstem), at 0.97. The smallest region, the pituitary gland, had the lowest Dice score at 0.58 [7]. Therefore, the pituitary gland region is apparently more susceptible to false positives and false negatives than other regions. In addition, accurate region segmentation is more difficult if the target is smaller, and methods should be improved to address this limitation.

In this study, segmentation of midsagittal brain MRI scans could extract the approximate contours and regions of organs from 2D images. This method cannot be applied to volume measurement and is not suitable for computer-aided diagnosis, but it is straightforward and can deliver general 2D positional information within a short time. Based on the

data, calculation times, and Jaccard indices obtained in this study, segmentation from 2D images can be considered a viable and effective approach for clinical use. Just as Rohini et al. applied brainstem segmentation to the diagnosis of Alzheimer's disease [6], simultaneous segmentation will enable automatic evaluation of pituitary signals and abnormalities in important organs such as the brainstem and corpus callosum.

All datasets used in our study were open database images, and we did not perform evaluations using clinical images. The accuracy of deep learning is strongly affected by the image quality of the dataset; therefore, testing on clinical images could yield lower accuracy. We used 450 images in this study, but more images should be used to eliminate bias due to the lack of datasets. In addition, it is necessary to investigate in more detail why U-Net, which produced acceptable results in previous studies, did not behave as expected in this study.

## 5 Conclusion

SegNet with the VGG16-weighting factor provided superior automatic segmentation of organs (cerebrum, cerebellum, corpus callosum, brainstem, and pituitary gland) on brain sagittal T1-weighted images. Our study demonstrated a network that appears to be the most suitable for segmenting organs from 2D images of midsagittal brain slices. This finding may enable future automatic evaluation of important organs depicted in sagittal images. In addition, we acquired new knowledge facilitating organ segmentation in 2D image slices versus segmentation in 3D images.

**Acknowledgements** Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Araclon Biotech, BioClinica Inc., Biogen, Bristol-Myers Squibb Company, CereSpir Inc., Cogstate, Eisai Inc., Elan Pharmaceuticals Inc., Eli Lilly and Company, EuroImmun F, Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc., Fujirebio, GE Healthcare, IXICO Ltd., Janssen Alzheimer Immunotherapy Research & Development LLC., Johnson & Johnson Pharmaceutical Research & Development LLC., Lumosity, Lundbeck, Merck & Co. Inc., Meso Scale Diagnostics LLC., NeuroRx Research, Neurotrack Technologies, Novartis Pharmaceuticals Corporation, Pfizer Inc., Piramal Imaging, Servier, Takeda Pharmaceutical Company, and Transition Therapeutics. The Canadian Institutes of Health Research provides funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data were

disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

**Funding** None.

## Declarations

**Conflict of interest** All authors declare that they have no conflict of interest.

**Ethical approval** This article does not include any studies with human participants or animals that were performed by any of the authors.

## References

1. Tustison NJ, Shrinidhi KL, Wintermark M, Durst CR, Kandel BM, Gee JC, Grossman MC, Avants BB. Optimal symmetric multi-modal templates and concatenated random forests for supervised brain tumor segmentation (Simplified) with ANTsR. *Neuroinformatics*. 2015;13(2):209–25.
2. Laukamp KR, Thiele F, Shakirin G, Zopfs D, Faymonville A, Timmer M, Maintz D, Perkuhn M, Borggrefe J. Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. *Eur Radiol*. 2019;29(1):124–32.
3. Naceur MB, Saouli R, Akil M, Kachouri R. Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images. *Comput Methods Programs Biomed*. 2018;166:39–49.
4. Perea Malla CU, Valdes Hernandez MD, Rachmadi MF, Komura T. Evaluation of enhanced learning techniques for segmenting ischaemic stroke lesions in brain magnetic resonance perfusion images using a convolutional neural network scheme. *Front Neuroinform*. 2019;13:33.
5. Staartjes VE, Serra C, Muscas G, Maldaner N, Akeret K, van Niftrik CHB, Fierstra J, Holzmann D, Regli L. Utility of deep neural networks in predicting gross-total resection after transsphenoidal surgery for pituitary adenoma: a pilot study. *Neurosurg Focus*. 2018;45(5):E12.
6. Staartjes VE, Zattra CM, Akeret K, Maldaner N, Muscas G, van BasNiftrik CH, Fierstra J, Regli L, Serra C. Neural network-based identification of patients at high risk for intraoperative cerebrospinal fluid leaks in endoscopic pituitary surgery. *J Neurosurg*. 2019;133(2):329–35.
7. Mlynarski P, Delingette H, Alghamdi H, Bondiau P, Ayache N. Anatomically consistent CNN-based segmentation of organs-at-risk in cranial radiotherapy. *J Med Imaging (Bellingham)*. 2020;7(1):014502.
8. Chen H, Lu W, Chen M, Zhou L, Timmerman R, Tu D, Nedzi L, Wardak Z, Jiang S, Zhen X, Gu X. A recursive ensemble organ segmentation (REOS) framework: application in brain radiotherapy. *Phys Med Biol*. 2019;64(2):025015.
9. Huo Y, Xu Z, Xiong Y, Aboud K, Parvathaneni P, Bao S, Bermudez C, Resnick SM, Cutting LE, Landman BA. 3D whole brain segmentation using spatially localized atlas network tiles. *Neuroimage*. 2019;194:105–19.
10. Kang S, Shin S, Seo S, Byun M, Lee D, Kim Y, Lee D, Lee J. Deep learning-based 3D inpainting of brain MR images. *Scientific Reports*. 2021;11(1):1–1.
11. Lee C, Huh S, Ketter TA, Unser M. Unsupervised connectivity-based thresholding segmentation of midsagittal brain MR images. *Comput Biol Med*. 1998;28(3):309–38.

12. Rohini P, Sundar S, Ramakrishnan S. Characterization of Alzheimer conditions in MR images using volumetric and sagittal brainstem texture features. *Comput Methods Programs Biomed.* 2019;173:147–55.
13. Wyman BT, Harvey DJ, Crawford K, Bernstein MA, Carmichael O, Cole PE, Crane PK, DeCarli C, Fox NC, Gunter JL, Hill D, Killiany RJ, Pachai C, Schwarz AJ, Schuff N, Senjem ML, Suhy J, Thompson PM, Weiner M, Jack CR, Initiative AsDN. Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimers Dement.* 2013;9(3):332–7.
14. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. *Proc Adv Neural Inf Process Syst.* 2012;1106–14.
15. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
16. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–8.
17. Badrinarayanan V, Kendall A, Coipolla R. SegNet: a Deep convolutional Encoder-Decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39:2481–95.
18. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint.* 2014. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
19. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Med Image Comput and Computer-Ass Interv (MICCAI)*. 2015;9351:234–41.
20. Mechrez R, Goldberger J, Greenspan H. Patch-based segmentation with spatial consistency: application to MS lesions in brain MRI. *Int J Biomed Imaging.* 2016;7952541.
21. Coup P, Manj JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *Neuroimage.* 2011;54(2):940–54.
22. Kuzina A, Egorov E, Burnaev E. Bayesian generative models for knowledge transfer in MRI semantic segmentation problems. *Front Neurosci.* 2019;13:844.
23. Cui S, Mao L, Jiang J, Liu C, Xiong S. Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network. *J Healthc Eng.* 2018;4940593.
24. Swati Z, Zhao Q, Kabir M, Ali F, Ali Z, Ahmed S, Lu J. Brain tumor classification for MR images using transfer learning and fine-tuning. *Comput Med Imaging Graph.* 2019;75:34–46.
25. Guo X, Zhang N, Guo J, et al. Automated polyp segmentation for colonoscopy images: A method based on convolutional neural networks and ensemble learning. *Med Phys.* 2019;46:12.
26. Saikia AR, Bora K, Mahanta LB, Das AK. Comparative assessment of CNN architectures for classification of breast FNAC images. *Tissue Cell.* 2019;57:8–14.
27. Lee H, Hong H, Kim J, Jung DC. Deep feature classification of angiomyolipoma without visible fat and renal cell carcinoma in abdominal contrast-enhanced CT images with texture image patches and hand-crafted feature concatenation. *Med Phys.* 2018;45(4):1550–61.

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