Supplementary Information

Reproducibility and clinical validation of automated habenula segmentation via deep learning in major depressive disorder with 7 Tesla MRI

Sang-Heon Lim[1,2,†], Jihyun Yoon[3,†], Young Jae Kim[2], Chang-Ki Kang[4], Seo-Eun Cho[5], Kwang Gi Kim[1,2,*] & Seung-Gul Kang[5,*]

[1]*Department of Health Sciences and Technology, Gachon Advanced Institute for Health Sciences and Technology (GAIHST), Gachon University, Seongnam-si, Republic of Korea;*

[2]*Department of Biomedical Engineering, College of Medicine, Gachon University, Seongnam-si, Republic of Korea;*

[3]*Department of Family Medicine, Chungbuk National University Hospital, Cheongju, Republic of Korea;*

[4]*Department of Radiological Science, College of Health Science, Gachon University, Incheon, Republic of Korea;*

[5]*Department of Psychiatry, Gil Medical Center, Gachon University College of Medicine, Incheon, Republic of Korea;*

Supplementary Table S1. Demographic and clinical characteristics of the participants and their comparison between the MDD and normal control groups.

Properties	Total (n = 69)	MDD (n = 33)	NC (n = 36)	Statistics	
Age, years	37.7 ± 13.1	40.4 ± 14.1	34.7 ± 11.0	t = 1.76	p = 0.084
Sex (male:female)	20:49	8:25	12:24	χ^2 = 0.69	p = 0.406
Education, years	14.0 ± 3.0	12.7 ± 3.4	15.2 ± 1.9	t = –3.37	p = 0.001
HDRS-17 score	8.8 ± 7.9	15.6 ± 5.8	2.5 ± 2.4	t = 13.03	p < 0.001
BDI score	15.1 ± 15.2	27.8 ± 12.7	3.5 ± 3.7	t = 10.62	p < 0.001
CGI score	2.5 ± 1.7	4.0 ± 1.0	1.1 ± 0.2	t = 17.88	p < 0.001

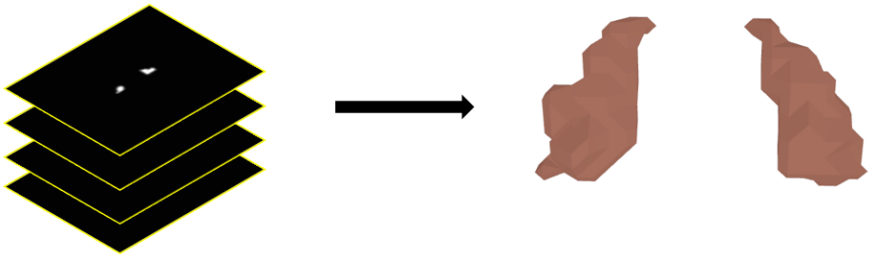
Data are mean±SD or number

Abbreviations: MDD, major depressive disorder; SD, standard deviation; NC, normal control; HDRS-17, Hamilton Depression Rating Scale-17 items; BDI, Beck Depression Inventory; CGI, Clinical Global Impression Scale

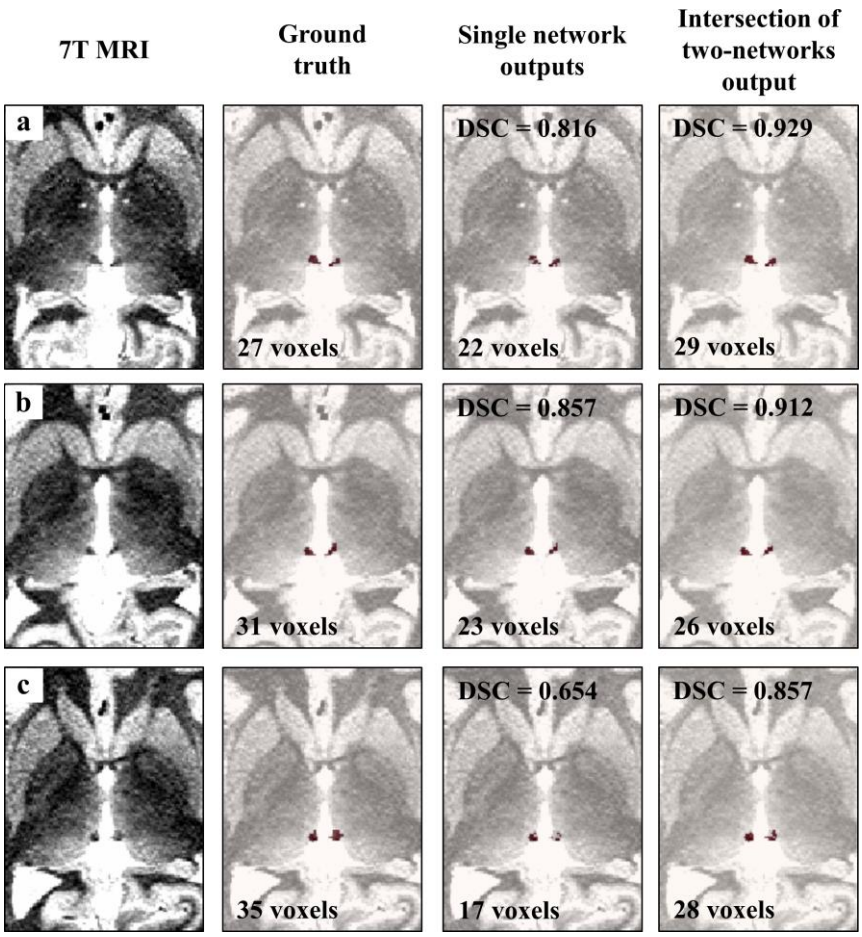
Supplementary Table S2. The automatic segmentation results of a single attention-network trained with intersected ground truth.

	Precision	**Sensitivity**	**DSC**
Fold 1	0.824 ± 0.046	0.750 ± 0.057	0.759 ± 0.202
Fold 2	0.852 ± 0.184	0.782 ± 0.182	0.798 ± 0.156
Fold 3	0.871 ± 0.155	0.808 ± 0.160	0.818 ± 0.123
Fold 4	0.833 ± 0.171	0.861 ± 0.150	0.831 ± 0.131
Fold 5	0.855 ± 0.164	0.744 ± 0.208	0.776 ± 0.164
Average	**0.847 ± 0.143**	**0.789 ± 0.150**	**0.790 ± 0.154**

Abbreviations: DSC, dice similarity coefficient



Supplementary Figure S1. An example of the 3D volume reconstruction for the habenula volume calculation.



Supplementary Figure S2. Ablation study of the two training methods.

From the left, 7 Tesla MRI, gold-standard data (ground truth), segmentation results of a single attention-network, and intersected segmentation results of a fusion attention-network. The single attention-network was trained with the intersected ground truths obtained from two different examiners. The proposed network is a parallel structure that trains two different ground truths separately. The fusion segmentation results were generated using the intersections of the segmentation results of two attention u-net networks. (**a–c**) Axial images from three different participants. Due to the low sensitivity of the single network, the DSCs and predicted region of voxels were lower than the proposed network.

Abbreviations: MRI, magnetic resonance imaging; DSC, dice similarity coefficient