A Transfer Learning Approach to Localising a Deep Brain Stimulation Target

Ying-Qiu Zheng[1][(✉)] , Harith Akram[2] , Stephen Smith[1] ,
and Saad Jbabdi[1]

[1] Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK
ying-qiu.zheng@ndcn.ox.ac.uk
[2] UCL Institute of Neurology, Queen Square, London, UK

Abstract. The ventral intermediate nucleus of thalamus (Vim) is a well-established surgical target in magnetic resonance-guided (MR-guided) surgery for the treatment of tremor. As the structure is not identifiable from conventional MR sequences, targeting the Vim has predominantly relied on standardised Vim atlases and thus fails to model individual anatomical variability. To overcome this limitation, recent studies define the Vim using its white matter connectivity with both primary motor cortex and dentate nucleus, estimated via tractography. Although successful in accounting for individual variability, these connectivity-based methods are sensitive to variations in image acquisition and processing, and require high-quality diffusion imaging techniques which are often not available in clinical contexts. Here we propose a novel transfer learning approach to accurately target the Vim particularly on clinical-quality data. The approach transfers anatomical information from publicly available high-quality datasets to a wide range of white matter connectivity features in low-quality data, to augment inference on the Vim. We demonstrate that the approach can robustly and reliably identify the Vim despite compromised data quality, and is generalisable to different datasets, outperforming previous surgical targeting methods.

Keywords: Surgical targeting · Deep brain stimulation · MR-guided surgery · Ventral intermediate nucleus of thalamus · Transfer learning

1 Introduction

Functional neurosurgical techniques, such as deep brain stimulation (DBS), and MR-guided stereotactic ablation, have been used as effective treatments of neurological and psychiatric disorders for decades. By intervening on a target brain structure, a neurosurgical treatment typically modulates brain activity in disease-related circuits and can often mitigate the disease symptoms and restore

brain function when drug treatments are ineffective. An example is the ventral intermediate nucleus of thalamus (Vim), a well-established surgical target in DBS and stereotactic ablation for the treatment of tremor in Parkinson's Disease, essential tremor, and multiple sclerosis [1]. The Vim plays a central role in tremor circuitry [2–9]. It receives efferent fibers from the dentate nucleus of the contralateral cerebellum and projects primarily to the primary motor cortex (M1) [5], as part of the dentato-thalamo-cortical pathway (DTCp).

As the structure is not readily visible on conventional MR images, targeting the Vim has relied primarily on standardised coordinates provided by stereotactic atlases, instead of directly targeting the structure based on subject-specific image-derived features. Such standardised stereotactic coordinates/atlases provide a reproducible way to identify the nucleus. However, the atlas-based Vim targeting falls short of accounting for the inter-individual and often inter-hemispheric anatomical variability, which are often substantial for thalamic nuclei [14], whilst efficacy of stereotactic operations heavily depends on accurate identification of the target nucleus.

To overcome this limitation, several recent studies proposed to localise the Vim using its anatomical connectivity features (e.g., with M1 and the dentate nucleus) *in vivo* on an individual basis [11,15,16], aided by cutting-edge diffusion-weighted imaging (DWI) and tractography techniques. DWI allows estimation of local fiber orientations, based upon which tractography generates streamline samples representative of the underlying white matter pathway, starting from a given seed. Typically, these studies identify Vim by finding the region of maximum connectivity with both M1 and contralateral dentate, and thus better capture individual variations of the nucleus, leading to improved clinical outcome [11,12].

However, the connectivity-driven Vim requires high angular resolution diffusion imaging (HARDI) techniques to allow reconstruction of Vim's connectivity features, which are often not readily available in advanced-care clinical contexts. Furthermore, even with cutting-edge HARDI and higher order diffusion modelling techniques, the connectivity-derived Vim has exhibited considerable variations across different acquisition protocol and processing pipelines, suggesting that they have to be used with caution to ensure that they adhere to the true underlying anatomical variability instead of simply reflecting the methodological confounds erroneously interpreted as variability.

Given the limitations of the standardised approaches and connectivity-based methods, we propose a novel approach, HQ-augmentation, to reliably target the Vim, particularly for clinical (or generally lower-quality) data. We utilised the publicly-available high-quality (HQ) Human Connectome Project (HCP) dataset [18,19] to augment surgical targeting of the Vim on low-quality (LQ) data. More specifically, the approach transfers the anatomical information derived from high-quality data, i.e., the approximate position of the Vim, to a wide range of low-quality white matter connectivity features in order to infer the likelihood that a given voxel is classified as part of Vim on the corresponding low-quality data. We demonstrate that the proposed approach not only yields
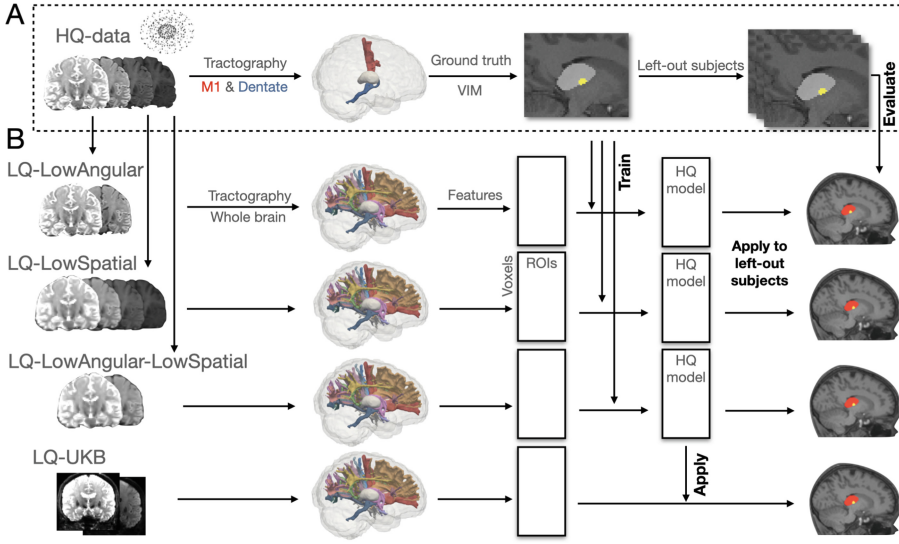
Fig. 1. Schematic illustration of the HQ-augmentation model. The process begins with the use of HQ diffusion data to establish a "ground truth" Vim within the thalamic masks, utilising white matter connectivity with the M1 and contralateral cerebellum. Following this, surrogate low-quality diffusion datasets are generated through intentional degradation of the HQ datasets. The HQ-augmentation model is then trained using the "ground truth" Vim derived from the HCP HQ data as target labels, and a broad set of HCP's low-quality connectivity profiles (in the form of voxel-by-ROI matrices) as input features. After training, the model is applied to unseen low-quality datasets, which include surrogate low-quality HCP and UK Biobank diffusion data, to predict Vim location. The performance of the model is subsequently evaluated against the corresponding HQ "ground truth" data.

consistent Vim targets despite compromised data quality, but also preserves inter-individual anatomical variability of the structure. Furthermore, the approach generalises to unseen datasets with different acquisition protocols, showing potential of translating into a reliable clinical routine for MR-guided surgical targeting.

2   Materials and Methods

HQ Dataset I: Human Connectome Project (HCP) 3T MRI Datasets. We leveraged the 3T diffusion MRI data from HCP [18] as the HQ dataset. The minimally pre-processed T1-, T2-, and diffusion-weighted MRI scans from a total of 1,062 healthy young adults were included, among which 43 subjects were scanned twice. Diffusion-weighted imaging was acquired at isotropic spatial resolution 1.25 mm, with three shells (b-values = 1000, 2000 and $3000\,\text{s/mm}^2$) and approximately 90 unique diffusion directions per shell, acquired twice (total scan time 60 min per subject) [17].

HQ Dataset II: UK Biobank (UKB) 3T MRI Datasets. The UKB 3T MRI datasets [24] were also used as HQ data. The T1, T2 and diffusion-weighted scans from 2,560 subjects with retest (second scan) sessions were included. The diffusion MRI was carried out at isotropic spatial resolution 2 mm with two shells (b-values = 1000 and 2000 s/mm^2), 50 diffusion directions per shell (total scan time 6min per subject).

Surrogate Low-Quality (LQ) Datasets. We considered a range of low-quality datasets representing the typical data quality in clinical contexts. This included 1) HCP surrogate low angular resolution diffusion dataset, obtained by extracting the b = 0 s/mm^2 and 32 b = 1000 s/mm^2 volumes from HQ HCP 3T diffusion MRI ("**LQ-LowAngular**"); 2) HCP surrogate low spatial resolution dataset, obtained by downsampling the HCP 3T diffusion MRI to isotropic 2 mm sptial resolution ("**LQ-LowSpatial**"); 3) HCP surrogate low angular and spatial resolution dataset, created by downsampling the surrogate low-angular-resolution dataset to isotropic 2 mm spatial resolution ("**LQ-LowAngular-LowSpatial**"); 4) UKB surrogate low angular resolution dataset, created by extracting the b = 0 s/mm^2 and 32 b = 1000 s/mm^2 volumes from the original UKB diffusion dataset ("**LQ-UKB**").

The Connectivity-Driven Approach. We followed the practice in [11] to generate connectivity-driven Vim. This approach identifies the Vim by finding the maximum probability of connection to M1 and contralateral dentate nucleus within the thalamic mask, generated via probablistic tractography. When this is applied to HQ original data, it generates the ground truth segmentations. When it is applied to LQ surrogate data, it provides results against which other methods (atlas-based and HQ-augmentation) can be compared.

The Atlas-Defined Approach. We used a previously published and validated Vim atlas [11] to find atlas-defined Vim on individual subjects. The atlas was registered into the individual T1 space (also referred to native space) via the warp fields between the corresponding individual T1 scans and the MNI152 standard brain. The warped group-average Vim atlas was thresholded at 50% percentile in the native space as the atlas-defined Vim.

The HQ-Augmentation Approach. The goal of this approach is to leverage anatomical information in HQ data to infer the likelihood of a voxel belonging to the Vim, given a wide range of tract-density features (multiple distinct tract bundles) derived from low-quality data. The HQ-augmentation model was trained on the HCP dataset for each type of low-quality dataset. Using the HCP HQ data, we first generated the connectivity-driven Vim (referred to as HQ-Vim) as the "ground truth" location of the nucleus, serving as training labels in the model. Next, for each low-quality counterpart, we generated an extended set of tract-density features, targeting a wide range of region-of-interests (ROIs), as the input features of the model. The richer set of connectivity features serves

to compensate for the primary tract-density features (with M1 and dentate), when those are compromised by less sufficient spatial or angular resolution in low-quality diffusion MRI, thus making Vim identification less reliant on the primary tract-density features used in the connectivity-driven approach and more robust to variations in data quality. During training, the model learns to use the extended set of low-quality connectivity features to identify the Vim that is closest to the one that can be otherwise obtained from its HQ counterpart.

Specifically, assume $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ... \mathbf{x}_V]^T$ is a $V \times d$ connectivity feature matrix for a given subject, where $\mathbf{x}_i$ is a $d \times 1$ vector representing the connectivity features in voxel i, V is the total number of voxels of the thalamus (per hemisphere) for this subject; $\mathbf{y} = [y_1, y_2, ... y_V]^T$ is a $V \times 1$ vector containing the HQ-Vim labels for the V voxels, in which y_i is the label of voxel i. Given the low-quality features $\mathbf{X}$, we seek to maximise the probability of reproducing the exact same HQ-Vim label assignment $\mathbf{y}$ on its low-quality counterparts, across the training subjects

$$\log P(\mathbf{y}|\mathbf{X}) = \log[\frac{1}{Z(\mathbf{X})} \exp(-E(\mathbf{y}|\mathbf{X}))] \tag{1}$$

Here $E(\mathbf{y}|\mathbf{X})$ is the cost of the label assignment $\mathbf{y}$ given the features $\mathbf{X}$, whilst $Z(\mathbf{X})$ serves as an image-dependent normalising term. Maximising the posterior $P(\mathbf{y}|\mathbf{X})$ across subjects is equivalent to minimising the cost of the label assignment $\mathbf{y}$ given the features $\mathbf{X}$. Suppose $\mathcal{N}_i$ is the set of voxels neighbouring voxel i, the cost $E(\mathbf{y}|\mathbf{X})$ is modelled as

$$E(\mathbf{y}|\mathbf{X}) = \sum_i \psi_u(y_i|\mathbf{x}_i) + \sum_i \sum_{j \in \mathcal{N}_i} \psi_p(y_i, y_j|\mathbf{x}_i, \mathbf{x}_j) + \lambda_1 ||\mathbf{W}||_1 + \lambda_2 ||\mathbf{W}||_2^2 \tag{2}$$

The first component $\psi_u(y_i)$ measures the cost (or inverse likelihood) of voxel i taking label y_i. Here $\psi_u(y_i)$ takes the form $\psi_u(y_i) = \mathbf{w}_{y_i}^T \phi(\mathbf{x}_i)$, where $\phi(\cdot)$ maps a feature vector $\mathbf{x}_i = [x_1, x_2, ... x_d]$ to a further expanded feature space in order to provide more flexibility for the parameterisation. $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2]$ is the coefficient matrix, each column containing the coefficients for the given class (i.e., belonging to the HQ-Vim or not). Here we chose a series of polynomials along with the group-average Vim probability (registered into native space) to expand the feature space, i.e.,

$$\phi(\mathbf{x}_i) = [x_1, x_2, ... x_d, x_1^{p_1}, x_2^{p_1}, ... x_d^{p_1}, x_1^{p_2}, x_2^{p_2}, ... x_d^{p_2}, x_1^{p_3}, x_2^{p_3}, ... x_d^{p_3}, g_i]$$

where $p_1 = 2, p_2 = 0.5, p_3 = 0.2$ are the power of the polynomials, and g_i is the group-average probability of voxel i classified as Vim. The second pairwise cost encourages assigning similar labels to neighbouring voxels, particularly for those sharing similar connectivity features. We modelled this component as $\psi_p(\mathbf{x}_i, \mathbf{x}_j) = \mu(y_i, y_j) \rho_m k_m(\phi(\mathbf{x}_i), \phi(\mathbf{x}_j))$. Here $k_m(\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)) = \exp(-\gamma_m ||\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)||^2)$ is a kernel function modelling the similarity between voxel i and j in the extended feature space, with length scale γ_m, chosen via cross-validation. $\mu(\cdot)$ is a label compatibility function where $\mu(y_i, y_j) = 0$ if $y_i = y_j$ or $\mu(y_i, y_j) = 1$ if $y_i \neq y_j$.

Therefore, in a local neighbourhood, the kernel function penalises inconsistent label assignment of voxels that have similar features, thus allowing modelling local smoothness. ρ_m controls the relative strength of this pairwise cost weighted by $k_m(\cdot)$. Lastly, the $L1$ and $L2$ penalty terms serve to prevent overfitting of the model. We used a mean-field algorithm to iteratively approximate the maximum posterior $P(\mathbf{y}|\mathbf{X})$ [29] summed across the subjects. The approximated posterior is maximised via gradient descent in a mini-batch form, where the connectivity feature matrix of each subject serves as a mini-batch, demeaned and normalised, and sequentially fed into the optimisation problem.

3   Experiments and Discussions

Accuracy of the HQ-Augmentation Model on the HCP Surrogate Low-Quality Datasets. The HQ-augmentation model was trained using HQ-Vim as labels and the LQ connectivity profiles as features, and then tested on left-out subjects from the same LQ dataset. Its accuracy was evaluated against the HQ-Vim counterpart obtained from the HQ HCP data of these left-out subjects, which served as ground truth. The accuracies of the HQ-augmentation Vim were also compared with those of the atlas-defined Vim and the connectivity-driven Vim, the latter obtained using the low-quality M1 and dentate tract-density features. We considered two accuracy metrics: Dice coefficient, which measures the overlap with the ground truth, and centroid displacement, which calculates the Euclidean distance between the predicted and ground truth centers-of-mass. As the connectivity-driven approach may even fail on HQ data, resulting in unreliable HQ-Vim, the training and evaluation of HQ-augmentation model were conducted only on a subset of "good subjects", whose HQ-Vim's center-of-mass was within 4 mm from the atlas's center-of-mass in native space. For the subjects which HQ-Vim regarded as unreliable (i.e., deviating too much from the atlas), the accuracy of HQ-augmentation model was evaluated against the atlas-defined Vim. The HQ-augmentation model produced Vim predictions that are closest to the HQ-Vim (i.e., higher Dice coefficient and smaller centroid displacement) than the LQ connectivity-driven approach and the atlas-defined Vim, evaluated on the reliable subset (see Fig. 2A for the results on LQ-LowSpatial-LowAngular data; see supplementary materials for results on the other low-quality datasets). When the HQ-vim failed to serve as approximate ground truth, the HQ-augmentation model on low-quality data yielded Vim predictions that were closer to the atlas-defined Vim than the HQ-Vim derived from HQ data (Fig. 2B).

Generalisability of the HQ-Augmentation Model to UK Biobank. We also tested whether the HQ-augmentation models trained on HCP were generalisable to other datasets collected under different protocols. It is crucial for a model to be generalisable to unseen protocols, as collecting large datasets for training purposes in clinical contexts can often be impractical. We therefore applied the HQ-augmentation models trained on different HCP LQ datasets to UK Biobank surrogate low-quality data (LQ-UKB) and averaged the outputs to give a single
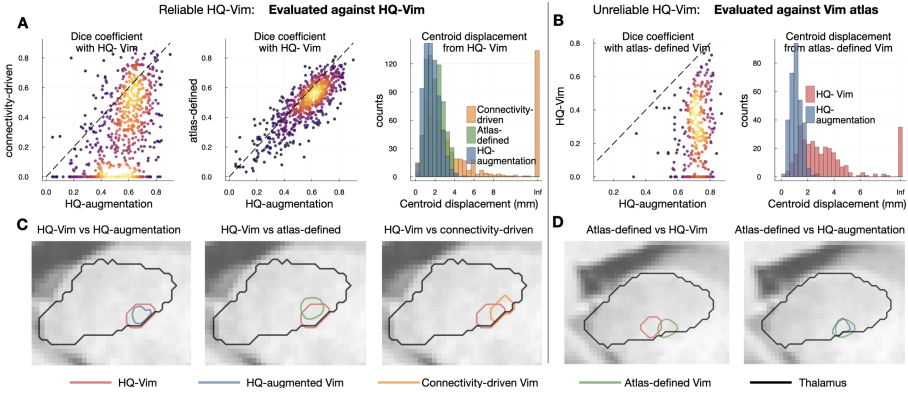
Fig. 2. Accuracy of HQ-augmentation on LQ-HCP-LowSpatial-LowAngular data. (A) When using the HQ-Vim as ground truth, the HQ-augmentation approach using low-quality features (here LQ-LowAngular-LowSpatial features) gives higher Dice coefficient and smaller centroid displacement with the HQ-Vim, than the atlas-defined Vim (green) and the low-quality connectivity-driven Vim (orange). **(B)** When using the atlas-defined Vim as ground truth (because, in these "unreliable" subjects, the HQ-connectivity method was considered to have failed), the HQ-augmentation model using low-quality features even gives more reliable results than the HQ-Vim (red). **(C)** and **(D)**. Example contours of Vim identified by each approach. (Color figure online)
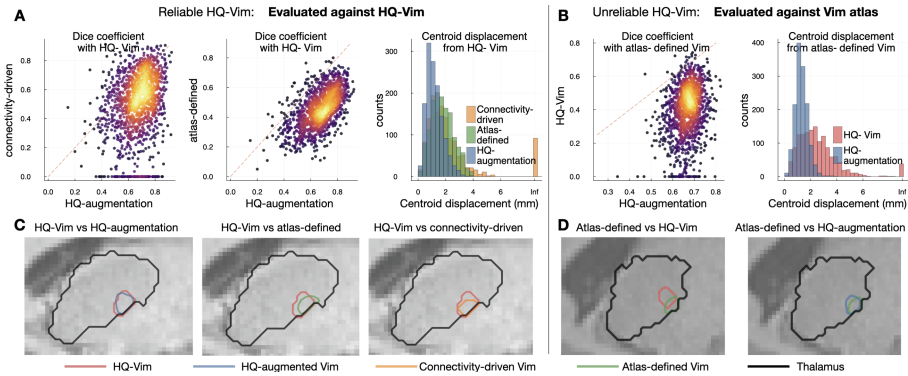

Fig. 3. Generalisability of HQ-Augmentation to UKB. The HQ-augmentation model trained on HCP was applied to the UKB low-quality features (blue) as HQ-augmented Vim. **(A)**. When using the UKB HQ-Vim as ground truth, the HQ-augmentated Vim has higher Dice coefficient and smaller centroid displacement with the UKB HQ-Vim, than the atlas-defined Vim (green) and the connectivity-driven Vim using low-quality features (orange). **(B)** When using the atlas-defined Vim as ground truth, the HQ-augmentation model using low-quality features even gave more reliable Vim than the UKB HQ-Vim (red), which used HQ features to target Vim. **(C)** and **(D)**. Contours of Vim, identified by each approach. (Color figure online)
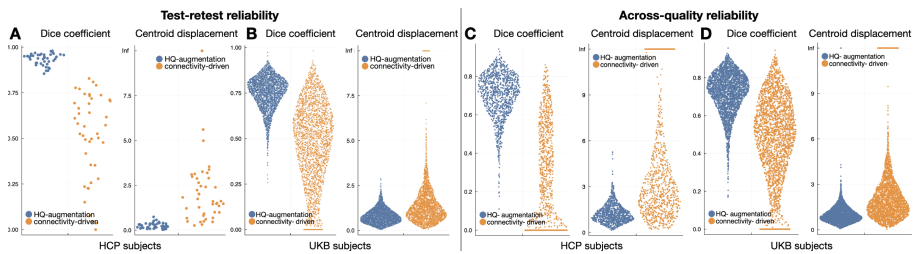
Fig. 4. The HQ-augmentation model (blue) gives more reproducible Vim identification than the connectivity-driven approach (orange), giving higher Dice coefficient and smaller centroid displacement across scanning sessions (i.e. test-retest reliability), **A** and **B**) and across datasets of different quality (across-quality reliability, **C** and **D**). (Color figure online)

ensembled Vim prediction per UKB subject in the LQ-UKB dataset. (Note that this step did not involve retraining or fine-tuning for LQ-UKB.) To evaluate its performance, similarly, we derived the HQ-Vim from the original (HQ) UK Biobank data via the connectivity-driven approach, and split it into a reliable subset, in which the UKB HQ-Vim served as the ground truth, and an unreliable subset, in which the atlas-defined Vim (warped into the UKB individual native space) served as the ground truth. On the reliable subset, the evaluations were conducted against the UKB HQ-Vim (Fig. 3A), whilst on the unreliable subset, the evaluations were conducted against the atlas-defined Vim (Fig. 3B). Despite being trained on HCP, the HQ-augmentation model produced Vim predictions that were closer to the corresponding HQ-Vim on the UKB subjects, outperforming the connectivity-driven approach on the UKB surrogate low-quality data and the atlas-defined approach. Even when using the atlas-defined Vim as the ground truth, the HQ-augmentation Vim on the UKB low-quality data showed higher accuracy than the HQ-Vim derived from the HQ data.

Reliability Analysis of the HQ-Augmentation Model. We also conducted a reliability analysis for the HQ-augmentation model and the connectivity-driven approach to assess their consistency in providing results despite variations in data quality (across-quality reliability) and across scanning sessions (test-retest reliability). To evaluate the HQ-augmentation model's across-quality reliability, we trained it using HQ tract-density maps to produce an "HQ version" of HQ-augmented Vim. The similarity between HQ-augmentation outputs using high- or low-quality features was assessed using the Dice coefficient and centroid displacement measures. Test-retest reliability was determined by comparing the outputs of the HQ-augmentation model on first-visit and repeat scanning sessions. Similarly, we assessed the across-quality reliability and test-retest reliability for the connectivity-driven approach, applied to high- and low-quality data accordingly. The HQ-augmentation model consistently provided more reliable results than the connectivity-driven approach, not only across datasets of different quality but also across different scanning sessions (Fig. 4).

Discussion. Our study presents the HQ-augmentation technique as a robust method to improve the accuracy of Vim targeting, particularly in scenarios where data quality is less than optimal. Compared to existing alternatives, our approach exhibits superior performance, indicating its potential to evolve into a reliable tool for clinical applications. The enhanced accuracy of this technique has significant clinical implications. During typical DBS procedures, one or two electrodes are strategically placed near predetermined targets, with multiple contact points to maximise the likelihood of beneficial outcomes and minimise severe side effects. Greater accuracy translates into improved overlap with the target area, which consequently increases the chances of successful surgical outcomes. Importantly, the utility of our method extends beyond Vim targeting. It can be adapted to target any area in the brain that might benefit from DBS, thus expanding its clinical relevance. As part of our ongoing efforts, we are developing a preoperative tool based on the HQ-augmentation technique. This tool aims to optimise DBS targeting tailored to individual patients' conditions, thereby enhancing therapeutic outcomes and reducing patient discomfort associated with suboptimal electrode placement.

References

1. Cury, R.G., et al.: Thalamic deep brain stimulation for tremor in Parkinson disease, essential tremor, and dystonia. Neurology **89**(13), 1416–1423 (2017)
2. Muthuraman, M., et al.: Oscillating central motor networks in pathological tremors and voluntary movements. What makes the difference? Neuroimage **60**(2), 1331–1339 (2012)
3. Baker, K.B., et al.: Deep brain stimulation of the lateral cerebellar nucleus produces frequency-specific alterations in motor evoked potentials in the rat in vivo. Exp. Neurol. **226**(2), 259–264 (2010)
4. Dum, R.P., Strick, P.L.: An unfolded map of the cerebellar dentate nucleus and its projections to the cerebral cortex. J. Neurophysiol. **89**(1), 634–639 (2003)
5. Gallay, M.N., et al.: Human pallidothalamic and cerebellothalamic tracts: anatomical basis for functional stereotactic neurosurgery. Brain Struct. Funct. **212**, 443–463 (2008)
6. Darian-Smith, C., Darian-Smith, I., Cheema, S.S.: Thalamic projections to sensorimotor cortex in the macaque monkey: use of multiple retrograde fluorescent tracers. J. Comparat. Neurol. **299**(1), 17–46 (1990)
7. Calzavara, R., et al.: Neurochemical characterization of the cerebellar-recipient motor thalamic territory in the macaque monkey. Eur. J. Neurosci. **21**(7), 1869–1894 (2005)
8. McIntyre, C.C., Hahn, P.J.: Network perspectives on the mechanisms of deep brain stimulation. Neurobiol. Dis. **38**(3), 329–337 (2010)
9. Helmich, R.C., et al.: Cerebral causes and consequences of parkinsonian resting tremor: a tale of two circuits? Brain **135**(11), 3206–3226 (2012)
10. Hirai, T., Jones, E.G.: A new parcellation of the human thalamus on the basis of histochemical staining. Brain Res. Rev. **14**(1), 1–34 (1989)
11. Akram, H., et al.: Connectivity derived thalamic segmentation in deep brain stimulation for tremor. NeuroImage: Clin. **18**, 130–142 (2018)

12. Su, J.H., et al.: Improved Vim targeting for focused ultrasound ablation treatment of essential tremor: a probabilistic and patient-specific approach. Hum. Brain Mapp. **41**(17), 4769–4788 (2020)
13. Elias, G.J.B., et al.: Probabilistic mapping of deep brain stimulation: insights from 15 years of therapy. Ann. Neurol. **89**(3), 426–443 (2021)
14. Morel, A., Magnin, M., Jeanmonod, D.: Multiarchitectonic and stereotactic atlas of the human thalamus. J. Comparat. Neurol. **387**(4), 588–630 (1997)
15. Ferreira, F., et al.: Ventralis intermedius nucleus anatomical variability assessment by MRI structural connectivity. NeuroImage **238**, 118231 (2021)
16. Bertino, S., et al.: Ventral intermediate nucleus structural connectivity-derived segmentation: anatomical reliability and variability. Neuroimage **243**, 118519 (2021)
17. Sotiropoulos, S.N., et al.: Advances in diffusion MRI acquisition and processing in the Human Connectome Project. Neuroimage **80**, 125–143 (2013)
18. Van Essen, D.C., et al.: The WU-Minn human connectome project: an overview. Neuroimage **80**, 62–79 (2013)
19. Glasser, M.F., et al.: The minimal preprocessing pipelines for the human connectome project. Neuroimage **80**, 105–124 (2013)
20. Andersson, J.L.R., Skare, S., Ashburner, J.: How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage **20**(2), 870–888 (2003)
21. Andersson, J.L.R., Sotiropoulos, S.N.: An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage **125**, 1063–1078 (2016)
22. Jenkinson, M., et al.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage **17**(2), 825–841 (2002)
23. Jenkinson, M., et al.: FSL. Neuroimage **62**(2), 782–790 (2012)
24. Miller, K.L., et al.: Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci. **19**(11), 1523–1536 (2016)
25. Alfaro-Almagro, F., et al.: Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. Neuroimage **166**, 400–424 (2018)
26. Destrieux, C., et al.: Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage **53**(1), 1–15 (2010)
27. Warrington, S., et al.: XTRACT-standardised protocols for automated tractography in the human and macaque brain. Neuroimage **217**, 116923 (2020)
28. Tang, Y., et al.: A probabilistic atlas of human brainstem pathways based on connectome imaging data. Neuroimage **169**, 227–239 (2018)
29. Zheng, S., et al.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision (2015)
30. LNCS Homepage. http://www.springer.com/lncs. Accessed 4 Oct 2017