

RESEARCH ARTICLE

Automatic localization of the subthalamic nucleus on patient-specific clinical MRI by incorporating 7 T MRI and machine learning: Application in deep brain stimulation

Jinyoung Kim¹  | Yuval Duchin¹ | Reuben R. Shamir¹ | Remi Patriat² | Jerrold Vitek³ | Noam Harel^{1,2,4} | Guillermo Sapiro^{1,5,6,7,8}

¹Surgical Information Sciences, Inc., Minneapolis, Minnesota

²Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, Minnesota

³Department of Neurology, University of Minnesota, Minneapolis, Minnesota

⁴Department of Neurosurgery, University of Minnesota, Minneapolis, Minnesota

⁵Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina

⁶Department of Biomedical Engineering, Duke University, Durham, North Carolina

⁷Department of Computer Science, Duke University, Durham, North Carolina

⁸Department of Mathematics, Duke University, Durham, North Carolina

Correspondence

Jinyoung Kim, Surgical Information Sciences, Inc., 50 South 6th St., Suite 1310, Minneapolis, MN 55402.

Email: aeneas.kim@gmail.com

Funding information

University of Minnesota Udall center, Grant/Award Number: P50NS098573; National Institutes of Health, Grant/Award Numbers: P30 NS076408, P41 EB015894, R01-NS085188

Abstract

Deep brain stimulation (DBS) of the subthalamic nucleus (STN) has shown clinical potential for relieving the motor symptoms of advanced Parkinson's disease. While accurate localization of the STN is critical for consistent across-patients effective DBS, clear visualization of the STN under standard clinical MR protocols is still challenging. Therefore, intraoperative microelectrode recordings (MER) are incorporated to accurately localize the STN. However, MER require significant neurosurgical expertise and lengthen the surgery time. Recent advances in 7 T MR technology facilitate the ability to clearly visualize the STN. The vast majority of centers, however, still do not have 7 T MRI systems, and fewer have the ability to collect and analyze the data. This work introduces an automatic STN localization framework based on standard clinical MRIs without additional cost in the current DBS planning protocol. Our approach benefits from a large database of 7 T MRI and its clinical MRI pairs. We first model in the 7 T database, using efficient machine learning algorithms, the spatial and geometric dependency between the STN and its adjacent structures (predictors). Given a standard clinical MRI, our method automatically computes the predictors and uses the learned information to predict the patient-specific STN. We validate our proposed method on clinical T₂W MRI of 80 subjects, comparing with experts-segmented STNs from the corresponding 7 T MRI pairs. The experimental results show that our framework provides more accurate and robust patient-specific STN localization than using state-of-the-art atlases. We also demonstrate the clinical feasibility of the proposed technique assessing the post-operative electrode active contact locations.

KEYWORDS

7 T MR imaging, deep brain stimulation, machine learning, neuromodulation, patient-specific sub-region targeting, subthalamic nucleus

Abbreviations: 7T3T, 7T MRI subthalamic nucleus atlas for use with 3T MRI; ANOVA, analysis of variance; DBS, deep brain stimulation; DC, Dice coefficient; DISTAL, DBS intrinsic template atlas; FDA, U.S. Food and Drug Administration; FGATIR, fast gray matter acquisition T₁ inversion recovery; FLAIR, fluid attenuated inversion recovery; MER, microelectrode recordings; MNIPD25, population-averaged atlas that was made with 3T MRI of 25 Parkinson's disease patients; PD, Parkinson's disease; QSM, quantitative susceptibility mapping; SN, substantia nigra; STN, subthalamic nucleus; SWI, susceptibility-weighted image; T, Tesla; T₁W, T₁-weighted; T₂W, T₂-weighted; UHFA, ultrahigh-field atlas

1 | INTRODUCTION

Deep brain stimulation (DBS) is a neuromodulation intervention for relieving the motor symptoms of Parkinson's disease (PD), dystonia, and Essential tremor, among others (Dormont et al., 2010; Kim et al., 2010; Krack, Hariz, Baunez, Guridi, & Obeso, 2010; Limousin et al., 1998; Mallet et al., 2007; Patel, Khan, & Gills, 2008; The Deep-Brain Stimulation for Parkinson's Disease Study Group, 2001; Volkmann, 2007). In particular, DBS of the subthalamic nucleus (STN) has been shown to be an effective symptom's treatment for advanced PD

(Dormont et al., 2010; Limousin et al., 1998; The Deep-Brain Stimulation for Parkinson's Disease Study Group, 2001; Volkmann, 2007).

Accurate 3D positioning of the chronic electrode within the STN is critical for the success of the DBS surgery, as its efficacy and adverse effects are highly correlated with the electrode's location (Hamid et al., 2005; Kerl et al., 2012; Mallet et al., 2007; Patel et al., 2008; Starr et al., 2002). Thus, precise identification of the STN (lens-shaped) of individual patients facilitates the DBS planning (and post-op programming). However, such identification in clinical settings still remains challenging due to its small size (approximately $6 \times 4 \times 5$ mm) and the ambiguous border with neighboring regions (Abosch, Yacoub, Ugurbil, & Harel, 2010).

Both direct- and indirect-targeting is often incorporated to estimate the 3D STN location and shape. Direct identification of the STN and its surrounded regions (e.g., substantia nigra [SN] and zona incerta) is possible using various MRI modalities where these structures' appearance is hypo-intense such as T_2W , fluid attenuated inversion recovery (FLAIR), fast gray matter acquisition T_1 inversion recovery (FGATIR), and susceptibility weighted imaging (SWI). However, the distinction among the neighborhood structures is often unclear on the clinical MRI, and the STN appears in only one or two slices, thereby resulting in sub-optimal targeting within the 3D STN (positive effects), and relative to adjacent structures of the STN (potential negative/side effects). There are efforts underway to directly visualize the STN via MRI reconstruction methods such as susceptibility weighted phase imaging or quantitative susceptibility mapping (QSM) (Chandran, Bynevelt, & Lind, 2015; Liu et al., 2013; Rasouli et al., 2017). Moreover, recent studies have proposed automatic methods to segment the STN on these contrast enhanced MR sequences (Garzón, Sitnikov, Bäckman, & Kalpouzos, 2017; Milletari et al., 2017; Visser, Keuken, Forstmann, & Jenkinson, 2016). While early work on QSM shows promising results, uncertainty in susceptibility estimates under different acquisition protocols needs to be further investigated (Lauzon, McCreary, McLean, Salluzzi, & Frayne, 2016) and the overall performance needs to be validated on large-scale clinical data.

The indirect STN localization approach refers to the selection of targets based on nearby anatomical landmarks and then estimates consensus coordinates in relation to these landmarks. The most common method is to select the anterior- and posterior-commissure anatomical landmarks on the clinical MRI, define the midline and the mid commissural point, and use the consensus coordinates ± 12 mm lateral, 4 mm posterior, and 5 mm inferior to the mid commissural point (Starr et al., 2002). This is used as an initial estimation for the location of the STN that is later refined based on the neurosurgeon's expertise and preferences. However, this consensus method does not account for the obvious variability in the patients' anatomy (Daniluk, Davies, Elias, Novak, & Nazzaro, 2010; Kerl et al., 2012). Moreover, it was reported that there is significant inter-surgeon variability in the selection of anterior- and posterior-commissure points, which has a substantial effect on the localization of targets (Pallavaram et al., 2008). For these reasons, indirect targeting is almost always complemented by other refinement methods.

Atlas-based approaches have been proposed to improve targeting accuracy. Some studies show promising results for the visualization of the STN on 3 Tesla (T) MR datasets. Patch based label fusion methods were used for segmenting the STN and its adjacent structures using

multimodal 3 T MRIs (Haegelen et al., 2013; Xiao, Fonov, et al., 2014). Xiao, Jannin, et al. (2014) analyzed the morphometric variability of the STN obtained by majority voting of labels that were augmented on 3 T MRIs of advanced PD patients. D'Albis et al. (2015) provided a pipeline for DBS planning and post-operative validation and adopted an atlas-based segmentation in the surgical planning flow. Post-operative active contacts' clusters projected onto the CranialVault atlas were used for DBS target prediction in Pallavaram et al. (2015). A probabilistic approach to map DBS electrode coordinates onto the MNI space was presented in Horn, Kühn, et al. (2017). More recently, to achieve anatomical precision in the MNI space, a histological atlas (Chakravarty, Bertrand, Hodge, Sadikot, & Collins, 2006) was merged with subcortical atlases based on multiple contrast 3 T MR sequences from PD patients (Xiao et al., 2017) and high resolution multimodal MRIs and structural connectivity data (Ewert et al., 2017).

While encouraging initial results were obtained with these techniques, the targeting accuracy is oftentimes insufficient, due in part to the large per-patient STN variability, to fully ensure DBS treatment efficacy and safety. Based on patient-specific clinical data, further revisions of the approximated lead location are often required when using these atlas-based approaches.

Microelectrode recordings (MER) are often incorporated to define the precise location of the STN and correctly place the electrode lead within the targeted structure. These electrophysiological measurements require significant team expertise and extend the surgery time. MER are often complemented with patient's behavioral feedback while the subject is awake (Abosch et al., 2010). Note that brain shift has been shown to increase with the length of the procedure (e.g., resulting from extended MER), causing further challenges. Others have suggested that while the incidence of brain shift is infrequent, it is also unpredictable (Halpern, Danish, Baltuch, & Jaggi, 2008; Ivan et al., 2014; Petersen et al., 2010). Lack of standardization and variability in how individual centers use MER emphasizes the importance of developing additional standard and objective approaches for targeting the STN.

With recent advances in ultrahigh magnetic fields hardware and acquisition protocols, 7 T MR imaging techniques now allow the direct identification of small and complex anatomical structures, including the 3D STN, thanks to its superior contrast and high resolution (Abosch et al., 2010; Cho et al., 2011; Kerl et al., 2012). Furthermore, 7 T MRI has already facilitated the study of connectivity within the basal ganglia and thalamus and enabled the subdivision of the STN into motor, associative, and limbic sub-regions (Abosch et al., 2010; Lenglet et al., 2012; Plantinga et al., 2016). Keuken et al. (2013) investigated structural changes of the STN using atlases based on 7 T MRI in different age groups (healthy subjects). Probabilistic atlas maps obtained from multiple 7 T MR contrasts were used for analyzing anatomical variability on subcortical structures (Keuken et al., 2014). Wang et al. (2016) generated the ultrahigh-field atlas of the STN using the 7 T T_1 -weighted (T_1W) and T_2 -weighted (T_2W) MRI. For use with clinical data, Plassard et al. (2017) created atlases based on the 7 T MRI of healthy subjects and used them to segment subcortical structures on 3 T contrast enhanced MR sequences. Moreover, a high quality 7 T atlas obtained from elderly subjects was registered onto the 3 T MRI template averaged on PD patient's data (Milchenko et al., 2018). Automated methods to segment brain subcortical structures

TABLE 1 Demographic and clinical details of the used 80 subjects

Gender		Diagnosis			Age (mean $\pm$ standard deviation)
Male	Female	Essential Tremor	Parkinson's Disease	Normal Control	
61	19	12	56	12	60 $\pm$ 13.8

using 7 T MRI have been proposed, leveraging sufficient intensity information from multiple MR contrasts (Kim, Lenglet, Duchin, Sapiro, & Harel, 2014; Visser, Keuken, Douaud, et al., 2016; Visser, Keuken, Forstmann, et al., 2016). The clinical feasibility of 7 T MRI for localization of the STN has been previously demonstrated (Duchin, Abosch, Yacoub, Sapiro, & Harel, 2012). More recently, the U.S. Food and Drug Administration (FDA) cleared the first 7 T MRI system (The Magnetom Terra, Siemens Medical Solutions) for clinical use. However, 7 T MR machines are still rare in current clinical practice and are associated with significant infrastructure costs (Plantinga et al., 2014). Therefore, the STN still needs to be localized on the ubiquitous clinical platforms of 1.5 T or 3 T MRI.

In this work, we propose to incorporate a high-quality 7 T MR dataset for training machine learning methods to statistically model geometrical dependencies of the subcortical structures (the framework proposed here is named “7 T-ML”). We demonstrate that this approach facilitates the accurate prediction of the patient-specific STN location and shape on standard clinical MRI,¹ thereby enjoying the advantages of high-quality data without the need for a 7 T machine, or any additional hardware. The proposed 7 T-ML framework is extensively compared with ground truth obtained from 7 T MRI of the same subjects and with state-of-the-art atlas-based results (comprising histology, healthy population, Parkinson's population, and elderly, thereby introducing high variability in the testing and validation), showing the significantly improved performance, which is critical for accurate targeting, DBS lead localization, and clinical outcomes. The algorithm's relevance for DBS in particular and neuromodulation in general is demonstrated with the accurate visualization of active contacts from real DBS surgeries.

2 | METHODS

2.1 | Overview

We create an annotated dataset of multiple pairs (same subject) of clinical (1.5 T or 3 T) and 7 T MRIs along with the segmented (labeled) subcortical structures of interest. The 3D STN was segmented on the 7 T MRI, but not on the clinical image (where as discussed above, is not clearly visible). Subcortical structures in the vicinity of the STN were also segmented on the 7 T MRI (hereafter named “predictors”),

these are visible on the clinical image as well. This dataset was used for learning the geometric relationships between the STN and its predictors. The dataset containing the clinical and 7 T MRIs along with the segmented structures is referred hereafter as the “training set.” Given a new patient's clinical MRI (no 7 T images for this patient), the algorithm automatically detects the predictors and then computes the patient-specific, that is, patient's own, 3D STN location and shape. This is based on the learned information from the training set. To evaluate and validate the quality of the automatic STN localization, the results are compared with the STN segmented by experts on the 7 T MRI (the standard data split between training and testing set is used here), and further clinically validated with active contact positions obtained from the post-operative CT of the same patient. Comparisons with existing literature are provided as well, showing a significant improvement on accuracy and robustness, both critical for patient-specific STN DBS and neuromodulation targeting.

2.2 | Database and preprocessing

The 7 T MRI and its corresponding (same subject) clinical MRI (1.5 T or 3 T) of 80 subjects were used in this study under approval of the Institutional Review Board at the University of Minnesota. Demographic and clinical details for the subjects are presented in Table 1. Table 2 presents the MRI modality and resolution of the 7 T and clinical data that were used in this work.

Anatomical experts in the team (including neurologists, neurosurgeons, and radiologists) manually segmented the STN and its predictors on the 80 pairs of 7 T T₂W and SWI MRIs, and carefully cross-validated segmentation results. This is the same protocol and ground truth used for FDA approval of the technology that resulted from this algorithm.² The 7 T data acquisition protocol, pre-processing, and manual segmentation are detailed in Lenglet et al. (2012) and Duchin et al. (2018), and omitted here for brevity. For illustration purposes, manual delineation of the STN and SN on a pair of 7 T T₂W and SWI MRIs is visualized in Figure 1. Duchin et al. (2018) also demonstrated that the 7 T MRI based STN segmentation highly agrees with MER data (see also Shamir et al., 2018). Hypo-intense regions that are spatially adjacent with the STN are considered as predictors. Such subcortical structures are well visualized both on the 7 T and clinical MRIs (Cho et al., 2011).

The 7 T T₂W MRIs were co-registered to their clinical T₂W MRI counterparts using the FSL FMRIB's Linear Image Registration Tool (Jenkinson & Smith, 2001), and the structures segmented on the 7 T images were transformed accordingly (hereafter “7 T priors”). Data acquisition, pre-processing and co-registration for 7 T and clinical MRI (1.5 T or 3 T) used in this study were performed following protocols described in Duchin et al. (2012). The MRI pairs and the 7 T priors were all stored in a database for retrieval upon the introduction of a

¹The framework and algorithm here described are components of the patented and FDA cleared patient-specific STN visualization tool developed by Surgical Information Sciences, Inc. (Harel & Sapiro, 2016; Sapiro, Harel, Duchin, & Kim, 2017). A preliminary work was presented at conferences (Kim, Duchin, Kim, et al., 2015; Kim, Duchin, Sapiro, Vitek, & Harel, 2015a, b). The scope of this study is the validation and analysis of the method on a large scale clinical data, and comparison of the proposed method with available state-of-the-art methods.

²https://www.accessdata.fda.gov/cdrh_docs/pdf16/K162830.pdf

TABLE 2 MRI modality and resolution of 7 T and clinical data used in our 7 T-ML

Magnetic field strength MRI modality	7 T		1.5 T	3 T
	T ₂ W	SWI	T ₂ W	T ₂ W
Resolution	0.39 × 0.39 × 1.0 mm	0.39 × 0.39 × 0.8 mm	0.5 × 0.5 × 2.0 mm	0.55 × 0.55 × 2.0 mm
	0.39 × 0.39 × 2.0 mm		0.72 × 0.72 × 2.0 mm	

new patient's clinical MRI on which the STN needs to be localized. About 46 datasets of 7 T and 1.5 T MRIs pairs of patient groups were used for training and 34 datasets of 7 T and corresponding 1.5 T or 3 T MRIs of patients and normal control subjects were added in our database later for further validation. A subset of 15 PD patients in the training set was selected to further study in detail various factors that affect the performance of our algorithm. Table 3 presents clinical details and magnetic field strength of the clinical images in the training and validation sets. Note that 3 T MRIs were not used at all to train the algorithm, only for validation.

Given a new clinical T₂W MRI, we first compute its affine registration with all 46 clinical T₂W MRIs in the training set within our database. Intensity similarity scores are measured by calculating the mutual information between mid-brain regions of the clinical query image and the registered images in the training set (Wells, Viola, Atsumi, Nakajima, & Kikinis, 1996). About 30% of training images sorted by the similarity scores are empirically selected as the most similar sets to reduce biases (Aljabar, Heckemann, Hammers, Hajnal, & Rueckert, 2009), and then they are non-linearly registered onto the clinical query image. The Advanced Normalization Tools (Avants, Tustison, Song, Cook, et al., 2011; Avants, Tustison, & Song, 2011; Avants & Gee, 2004) was used for global affine registration (with the mutual information cost metric) and local nonlinear registration of the basal ganglia region (with cross-correlation cost metric, SyN option, and bspline interpolation). Then, we resample the clinical query image to a 0.5 mm isotropic resolution and transform the selected datasets

into the query image coordinates system. As a result, the clinical query data and the 7 T priors in the training set are now in a common coordinate system.

2.3 | 7 T-machine learning based STN prediction

The geometric relationship between the 3D STN and its predictors on the training set was analyzed using a regression-based shape prediction approach (Baka et al., 2011; Blanc, Seiler, Székely, Nolte, & Reyes, 2012; Rao, Aljabar, & Rueckert, 2008).

We first automatically segment the predictors on a query clinical T₂W MRI, which are observable subcortical structures near the STN, later used to predict it. To this end, the training dataset is incorporated in a unified framework of active shape model and active appearance model algorithms (Cerrolaza, Villanueva, & Cabeza, 2012; Cerrolaza, Reyes, Summers, González-Ballester, & Linguraru, 2015; Cootes, Taylor, Cooper, & Graham, 1995; Frangi, Rueckert, Schnabel, & Niessen, 2002; Heimann & Meinzer, 2009; Matthews & Baker, 2004; Sung, Kanade, & Kim, 2007; Tzimiropoulos & Pantic, 2013). Then, we apply the regression-based shape prediction, incorporating the computed predictors on the query clinical image and the geometric relationship learned from the STN and the segmented predictors on the training dataset.

More specifically, 3D shapes of the predictors and the STN are represented as the coordinates of surface points (vertices in a mesh), in correspondence, across random subsets of the most similar training sets previously selected (registered onto the clinical query image). The poses of the structures are extracted using the generalized Procrustes analysis (Cootes, Taylor, Cooper, & Graham, 1992; Gower, 1975), and their shape variations are then modeled using kernel principal component analysis (Guo, 2010; Rathi, 2006). A bagging procedure (Breiman, 1996) is applied in the partial least squares regression technique (Abdi, 2010; Krishnan, Williams, McIntosh, & Abdi, 2011; Wold, 1982; Wold, Sjöström, & Eriksson, 2001) to learn the dependency between the STN and its corresponding predictors (in the pose and shape feature space) from these randomly selected subsets of the most similar training sets (Kim, Duchin, Sapiro, et al., 2015a). We also investigated a regression forest model (Breiman, 2001; Criminisi et al., 2013) for

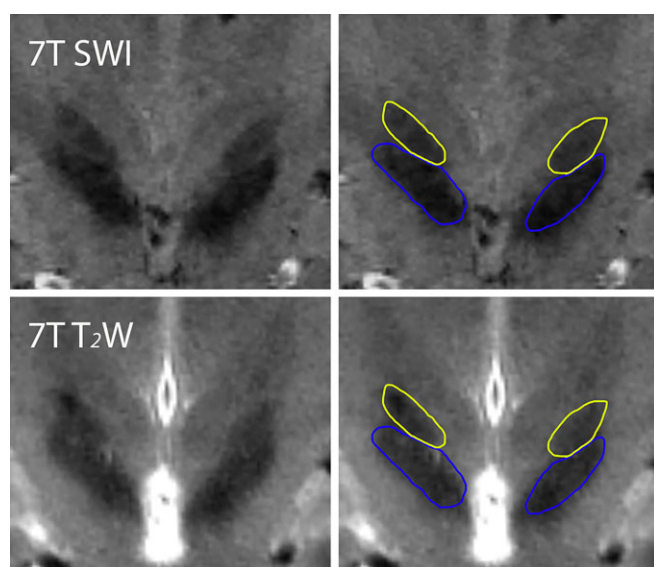


FIGURE 1 Direct visualization of SN and STN on a coronal 7 T SWI (top) and a coronal 7 T T₂W image registered onto the SWI (bottom). The blue and yellow represent contours of manually segmented SN and STN, respectively [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Clinical details and magnetic field strength of clinical images in the training and validation sets (80 subjects)

Dataset	Training set	Validation set	
Magnetic field strength	1.5 T	1.5 T	3 T
The number of patients	46 (15)	10	24
Essential tremor	10	–	2
Parkinson's disease	36 (15)	10	10
Normal control	–	–	12

Note. Numbers in parenthesis indicates the number of PD patients for in-depth study.

TABLE 4 Data acquisition and template T₂W images for tested state-of-the-art atlases

Atlases	7 T3 T (Milchenko et al., 2018)	UHFA (Wang et al., 2016)	DISTAL (Ewert et al., 2017)	MNIPD25 (Xiao et al., 2017)
Data on which the STN atlas was obtained	7 T T ₂ W MRI (elderly normal)	7 T T ₂ W MRI (normal)	3 T T ₁ W/T ₂ W/proton density MRI (normal), histology, and DWI (normal/PD patients)	3 T T ₂ *W MRI (PD patients)
Provided STN atlas template T ₂ W MRI	3 T T ₂ W MRI in ICBM MNI152 space (PD patients)	7 T T ₂ W MRI (normal)	3 T T ₂ W MRI in ICBM 2009b Asym MNI152 space (normal)	3 T T ₂ *W MRI in ICBM MNI152 space (PD patients)

finding the non-linear mapping between the STN and its predictors (Kim, Duchin, Sapiro, et al., 2015b). Given the computed (visible) predictors in a new patient clinical data, the 3D STN binary volume is predicted by exploiting this learnt spatial relationships. See Kim, Duchin, Sapiro, et al. (2015a) and (2015b) for additional technical details.

Such an ensemble approach with equal weights does not consider the influence of each training subset on the final confidence map of the predicted STN. If some training subsets are more influential than others with respect to the specific patient's STN prediction, the prediction accuracy can be further improved by increasing their weights/relevance. To estimate the contribution of each training subset to the prediction, we investigate non-linear relationships between pose-related features (Kim, Duchin, Kim, et al., 2015) and the prediction accuracy from various subjects and corresponding training subsets. This is done using a random forest model (Breiman, 2001). The global error score, namely a weighted sum of all the geometric measures, was used to determine the influence-weight of each training subset on the final prediction (Kim, Duchin, Kim, et al., 2015). Given new features from a query patient and each training subset, the error scores are estimated using the learned information. An ensemble of predictions from the random subsets with larger weighting (i.e., more influential sub-atlases for the prediction) yields lower error scores. The weighted ensemble also provides a final confidence map on the query patient. We refer to this procedure as "robust prediction" to differentiate it from "ensemble," that simply incorporates equal weights for all the subsets. See additional details and explicit parameters' values in Kim, Duchin, Kim, et al. (2015).

2.4 | Validation

To evaluate the performance of the proposed 7 T-ML method, we used clinical data (1.5 T or 3 T MRI) from 80 subjects (46 cases using a leave-one-out approach on the training set and additional 34 naïve cases for validation, not present at all in the training). One clinical MRI was selected as a query image at each iteration. When the query image was from the training set, all its associated data (segmented structures and other MRIs) was excluded from the 46 training sets. The STN was then automatically computed on the query image using the proposed framework. This was done with automatically segmented (on the clinical data) predictor structures, bagged partial least squares regression, and uniform weights of 100 training subsets. The current implementation utilizes Amazon Web Services to facilitate the large scale parallel processing.

Our 7 T-ML method is compared with the following publicly available STN atlases in the literature to validate its reliability for patient-specific clinical MRI based STN-DBS targeting: (1) 7 T MRI STN atlas for use with 3 T MRI (7 T3 T) (Milchenko et al., 2018); (2) ultrahigh-

field atlas (UHFA) (Wang et al., 2016); (3) DBS intrinsic template atlas (DISTAL) (Ewert et al., 2017; Horn, Neumann, Degen, Schneider, & Kühn, 2017); and (4) population-averaged atlas that was made with 3 T MRI of 25 PD patients (MNIPD25) (Xiao et al., 2017). Data acquisition and template images for these atlases are summarized in Table 4. This is a very comprehensive comparison since these atlases include 7 T data, histology, healthy subjects, PD patients, and elderly. To this end, we computed the transformation between the atlas T₂W MRI templates and the clinical T₂W MRIs across 80 subjects, and standard STN atlases were accordingly transformed to the clinical data. For a fair comparison, we applied the same approach as the inter-patient registration between database clinical T₂W images and a clinical query T₂W image in our proposed framework. It is a multi-step registration (global affine registration and local nonlinear registration on the basal ganglia region) that was adjusted based on a long trial and error process.

Additionally, we provide the STN segmentation results obtained with a conventional approach based on intensity information on the image - the active shape model and active appearance model framework (also used for predictor segmentation in our 7 T-ML) (Sung et al., 2007; Tzimiropoulos & Pantic, 2013). This also motivates a novel approach in challenging situations, for example, when the STN borders are not visible.

The STN manually segmented on the 7 T MRI and transformed onto the clinical MRI pair of the same subject was used as "ground truth." The following measures were first calculated to compare our 7 T-ML and standard atlases with the 7 T manual ground truth STN: (1) distance between the centers of mass; (2) mean Euclidean distance of surface points in correspondence; (3) dice coefficient (DC), which is a normalized overlap measure between two co-aligned binary datasets; and (4) volume of the STN (added here for completeness). As suggested in Shamir, Joskowicz, Spektor, and Shoshan (2009), 2 mm accuracy is considered an acceptable threshold for neurosurgical and neuromodulation applications. Therefore, we counted the number of cases that are less than 2 mm centers of mass distance as a clinically relevant prediction accuracy measure. Our 7 T-ML STN confidence map and standard atlases-based STN were normalized to [0, 1] and binarized with threshold values between 0.3 and 0.4 to eliminate resampling artifacts, thereby avoiding a bias in the measures induced by unexpectedly large volumes. A one-way analysis of variance (ANOVA) was calculated for each measure and a multiple comparison correction was performed to estimate the methods. Then, post hoc tests with Tukey's honest significant difference were executed.

To demonstrate the relevance of the high accuracy of the proposed 7 T-ML approach for DBS and neuromodulation, we also computed the distance between the DBS chronic electrode's active contacts and the predicted STN's centers of mass. The center of mass

TABLE 5 Summary of patient data used for the post-operative analysis in the training and validation sets

Dataset	Training set	Validation set
The number of patients	46	34
The number of patients for post-operative analysis	15	15
Unilateral STN-DBS	12 (11/12)	10 (9/10)
Bilateral STN-DBS	3 (5/6)	5 (6/10)

Note. Numbers in parenthesis represents the number of available active contacts out of total electrode reconstruction image.

is used here for reference, providing intrinsic distances for proper and consistent comparison (also across subjects); it should be noted that the STN's center of mass was not considered as the optimal target point. Pallavaram et al. (2015) measured the distance between active contacts and the estimated target position (using different targeting protocols, for example, stereotactic coordinates corresponding to the center of the STN's motor territory). For this, pre-operative clinical MRI and the STNs computed from each method are co-registered onto the post-operative CT of the same patient (where the implanted electrode is detected). We extensively examined the registration results to ensure the correct transformation of the STN.

Furthermore, to compare across different approaches the active contacts location of individual patients, we used a best-fit ellipsoid representation of the individual's STN. Then, we divided the space into eight regions (in posterior/anterior, medial/lateral, and superior/inferior axes) and counted the number of active contacts in each region (for all patients for which we have available data). We defined an active contact as "in" a specific region of the STN if its center of mass was in that region. Note that in contrast to common averaging techniques where very different shapes are (often significantly) deformed into a common coordinate system, this is a patient-specific intrinsic measurement, each individual STN is split into its own eight regions, and then votes are accumulated. This study demonstrates the sub-region accuracy of the proposed visualization/targeting approach that is critical both to localize the active DBS contacts and for future studies on the clinical targeting value of different STN sub-regions. The clinical data used for the post-operative analysis in the training and validation sets is summarized in Table 5. We used 38 electrode's lead images (with information available on 31 active contacts, the others had missing clinical data) reconstructed from the post-operative CT of 30 PD patients out of the total 80 subjects. These represent all the subjects from which we could retrieve from the available clinical records both post-operative data and outcome information. This study was repeated also for the location of non-active contacts for completeness.

Lastly, we selected from the training set 15 PD patients whose clinical 1.5 T T₂ W MRIs contain a whole head image in order to investigate the different factors that affect the STN prediction in our proposed 7 T-ML framework. For this in-depth analysis, we computed the STN under various setups: (1) *Predictors*: manual or automatic segmentation; (2) *Regression methods*: bagged partial least squares regression or random forest; (3) *Ensemble sizes*: 10, 100, and 200; and (4) *Weighting methods*: uniform or non-uniform weights on training subsets based on their contribution to the prediction accuracy.

3 | RESULTS

Our proposed 7 T-ML and standard atlases-based STNs are compared with the 7 T manual ground truth STN on 80 (160 STNs) subjects' clinical data, Table 6 and Figure 2. The 7 T-ML STNs were significantly closer to the 7 T manual ground truth STNs in comparison to the atlas-based results ($p < .0001$; ANOVA). Our results demonstrate the high accuracy and consistency of the proposed 7 T-ML. Compare the 1.25 ± 0.60 mm, 2.37 ± 1.74 mm, 2.94 ± 1.49 mm, 3.50 ± 3.57 mm, and 4.25 ± 3.33 average centers of mass distance between the STN that was computed from the 7 T-ML, 7 T3 T, UHFA, DISTAL, and MNIPD25 atlases, respectively, and the ground truth STN ($p < .0001$; ANOVA). Note that the average error in the 7 T-ML STN is close to half of the clinical data slice thickness (2 mm), which is roughly a lower bound to the possible segmentation accuracy. Furthermore, 89.4% (143/160), 53.8% (86/160), 28.1% (45/160), 51.9% (83/160), and 24.4% (39/160) of the STNs computed from the 7 T-ML, 7 T3 T, UHFA, DISTAL, and MNIPD25 atlases, respectively, were less than 2 mm from the ground truth center of mass (acceptable maximal error).

The 7 T-ML also yields significantly better average mean distance of surface points and DC values (0.57 ± 0.18 mm and $64 \pm 12\%$) than 7 T3 T (1.33 ± 1.03 mm and $44 \pm 21\%$), UHFA (1.76 ± 1.01 mm and $24 \pm 23\%$), DISTAL (2.16 ± 2.97 mm and $39 \pm 27\%$), and MNIPD25 (2.60 ± 2.73 mm and $26 \pm 24\%$) atlases, respectively, in comparison with the ground truth ($p < .0001$; ANOVA).

Volumes of the 7 T-ML based predicted STN (135.3 ± 25 mm³) and 7 T3 T atlas (129.4 ± 19 mm³) are comparable to those of the 7 T manual ground truth STN (134.1 ± 29 mm³) on average ($p > .05$; ANOVA and post hoc test), while the STNs based on other atlases are significantly different ($p < .05$; ANOVA and post hoc test). DISTAL atlas-based STN is the largest (142.5 ± 27 mm³), but UHFA and MNIPD25 based STNs are smaller on average (106.4 ± 23 mm³ and 116.0 ± 21 mm³, respectively).

Figure 3a presents histograms of the centers of mass distance measured on 160 STNs computed by the various discussed methods. Each zone represents bins that include median or average centers of mass distance (bin size: 0.5 mm). Median and average centers of mass distance of our proposed 7 T-ML STN are in zone (i), while median and average centers of mass distance for the STNs computed based on standard atlases are distributed in zones (ii) and (iii). Our proposed 7 T-ML demonstrates both higher accuracy and consistency in the centers of mass distance than standard atlases (i.e., smaller average and narrower distribution). Figure 3b and c visualize, for the compared methods and example subjects, the STNs in different average and median centers of mass distance zones, exemplifying the superior performance of the proposed 7 T-ML.

TABLE 6 Quantitative comparison of the STNs obtained using the proposed 7 T-ML and the 7 T3 T, UHFA, DISTAL, and MNIPD25 atlases

	7 T3 T						UHFA			DISTAL			MNIPD25		
	7 T-ML			(Milchenko et al., 2018)			(Wang et al., 2016)			(Ewert et al., 2017)			(Xiao et al., 2017)		
	CMD (mm)	MSD (mm)	DC	Volume (mm ³)	CMD (mm)	MSD (mm)	DC	Volume (mm ³)	CMD (mm)	MSD (mm)	DC	Volume (mm ³)	CMD (mm)	MSD (mm)	DC
Average	1.25	0.57	0.64	135.3	2.37	1.33	0.44	129.4	2.94	1.76	0.24	106.4	3.50	2.16	0.39
Standard Deviation	0.60	0.18	0.12	24.85	1.74	1.03	0.21	18.56	1.49	1.01	0.23	16.62	3.57	2.97	0.27
CMD < 2 mm (%)	89.4 (143/160)				53.8 (86/160)				28.1 (45/160)				51.9 (83/160)		
														24.4 (39/160)	

The various methods were compared with the 7 T manual ground truth STNs of 80 patients (160 STNs). A one-way ANOVA and post hoc test (with Tukey's method) showed that the proposed 7 T-ML is significantly better than atlas-based methods in each measure ($p < .0001$). CMD = centers of mass distance; MSD = mean distance of surface points.

The STN segmentation results based on the active shape model and active appearance model framework (with 7 T priors) showed 1.51 ± 0.78 mm (centers of mass distance), 0.61 ± 0.20 mm (mean distance of surface points), and $52.3 \pm 12\%$ (DC) on average, comparing to the 7 T manual ground truth. 79.4% (127/160) of the segmented STNs were less than 2 mm in centers of mass distance. Our 7 T-ML was still significantly better in average centers of mass distance and DC ($p < .001$; ANOVA). The volume of the segmented STN was also significantly different from our 7 T-ML and the 7 T manual ground truth ($p < .001$; ANOVA).

We also observed that the distance between active contacts and the 7 T-ML STN's centers of mass (again, here used to provide per-patient intrinsic coordinates) is much closer to that of the 7 T manual ground truth STN in comparison to standard atlas-based results (Table 7; compare 2.4 mm for our 7 T-ML and ground truth and 3.2–5.6 mm for standard atlases; see also Supporting Information Table S1 in the Supporting Information material for data on all the contacts). Once again, the STN's center of mass is not clinically considered the optimal target position, and is here used simply as a reference for intrinsic coordinates. The target region (preferred by implanting team) can be defined/visualized within our 7 T-ML STN. Nevertheless, the results here reported for our tested clinical data are comparable to those in the literature (Pallavaram et al., 2015).

Furthermore, as shown in Table 8 and Figure 4, the active contacts were frequently populated at the posterior lateral parts (i.e., regions 2 and 4) of the STN as obtained from the 7 T manual ground truth STN. The proposed 7 T-ML was consistent with this observation, while the spatial distribution of active contacts was markedly different in the atlas-based results. This is repeated for all the contacts in Supporting Information Table S2 and Supporting Information Figures S1–S4.

Figure 5 presents an STN visualization based on the 7 T manual ground truth, our proposed 7 T-ML, 7 T3 T, UHFA, DISTAL, and MNIPD25 atlases from zone (i) and (iii) in Figure 3a, representing average centers of mass distance, along with the post-operative electrode contacts of a specific PD patient. According to the post-operative monopolar reviews, the activation of the contact 1 was associated with the best motor improvement (62%), while activation of other contacts had resulted in lower motor improvements (15–46%). Contact 1 was entirely inside the dorsal STN according to our proposed 7 T-ML, confirmed with the 7 T manual ground truth STN. Active contacts associated with motor improvements were also placed in the dorsal area of the STN based on 7 T3 T, DISTAL, and MNIPD25. However, the atlas-based STNs were associated with large errors in shape and location with respect to the ground truth (unreliable STN trajectory). Note that a system operating/deciding based on our proposed 7 T-ML would have achieved optimal results (with the predicted STN close to the ground truth and the current lead location), while the same system operating based on those atlases would have misled the electrode placement, thereby resulting in adverse effects. Therefore, atlas-based STN segmentation would require further revision before being utilized into a clinical setup. While a full investigation of these aspects and consequences is beyond the scope of this study, these results confirm that our proposed sub-region accuracy 7 T-ML provides a reliable guide for

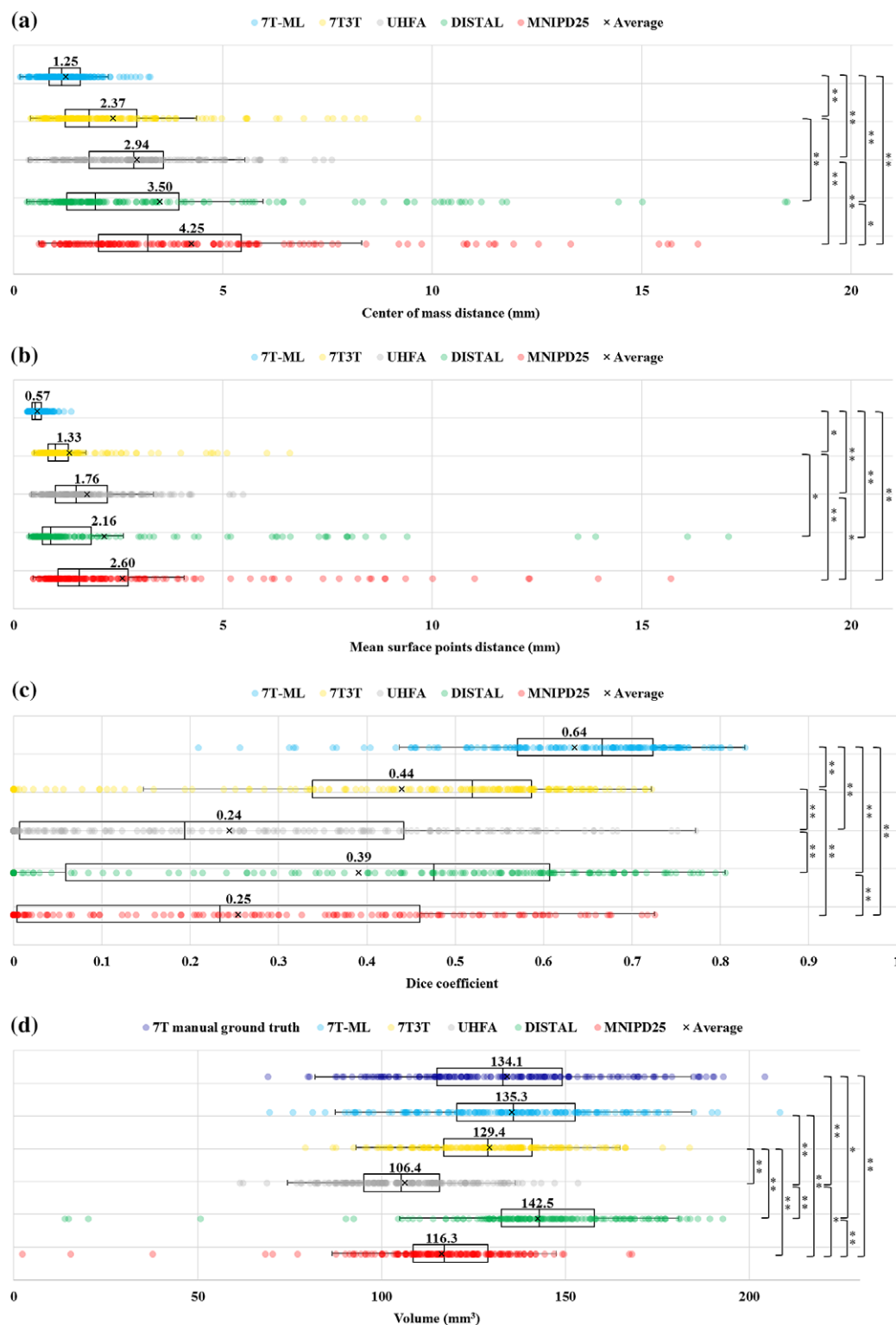


FIGURE 2 Comparison of (a) centers of mass distance, (b) mean distance of surface points (c) DC, and (d) volume for our proposed 7 T-ML based STN prediction (using the bagged partial least squares regression, uniform weights, 100 ensemble size, and automatically segmented predictors), 7 T3 T, UHFA, DISTAL, and MNIPD25 atlases, in comparison with 7 T manual ground truth across 80 patients. A one-way ANOVA and post hoc test (with Tukey's method) is performed for multiple comparisons. The significance level is denoted by asterisks (* for $p < .05$ and ** for $p < .001$; ANOVA and post hoc test). Note that the volume difference between the 7 T manual ground truth and our 7 T-ML based STN and 7 T3 T, respectively, are not significant ($p > .05$; ANOVA and post hoc test) [Color figure can be viewed at wileyonlinelibrary.com]

the STN (sub-region) targeting for DBS surgery and treatment based on patient-specific standard clinical MR data.

Next, the effect of the main components in our proposed 7 T-ML framework is analyzed as follows:

(1) *Predictors*: the average centers of mass distance between the predictors (non-STN subcortical structures) automatically segmented on the selected clinical MRI and their manual 7 T MRI counterparts was 0.92 ± 0.45 mm. The 7 T-ML STN using manually segmented

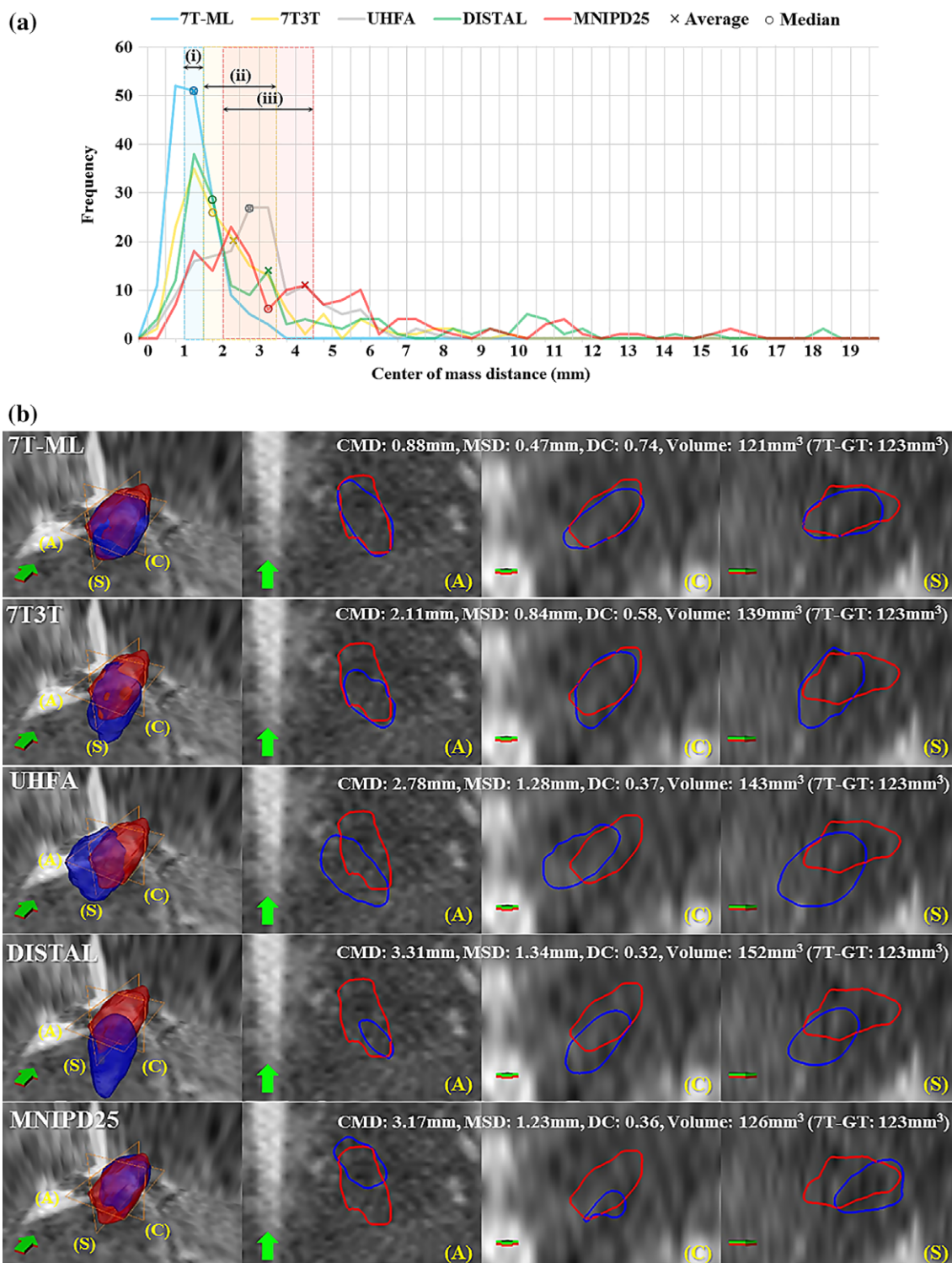


FIGURE 3 Continued [Color figure can be viewed at wileyonlinelibrary.com]

predictors showed comparable results to these observed using automatically segmented predictors (Figure 6; $p > .05$; ANOVA).

(2) *Regression methods*: The 7 T-ML using the bagged partial least squares regression produced significantly better results than using the regression forest method (Figures 6 and 7; $p < .05$; ANOVA).

(3) *Ensemble sizes*: The number of training subsets (10, 100, and 200) had insignificant effect on the STN prediction accuracy ($p > .05$; ANOVA).

(4) *Weighting methods*: Incorporating non-uniform weights based on estimated error scores produced comparable results to these observed using uniform weights (Figure 7; $p > .05$; ANOVA) for the bagged partial least squares regression and regression forest methods. However, incorporating the actual error scores for determining the weights of the training sets resulted in significantly better prediction with the regression forest method ($p < .01$; ANOVA), but insignificant with the bagged partial least squares regression. The mean squared

(c)

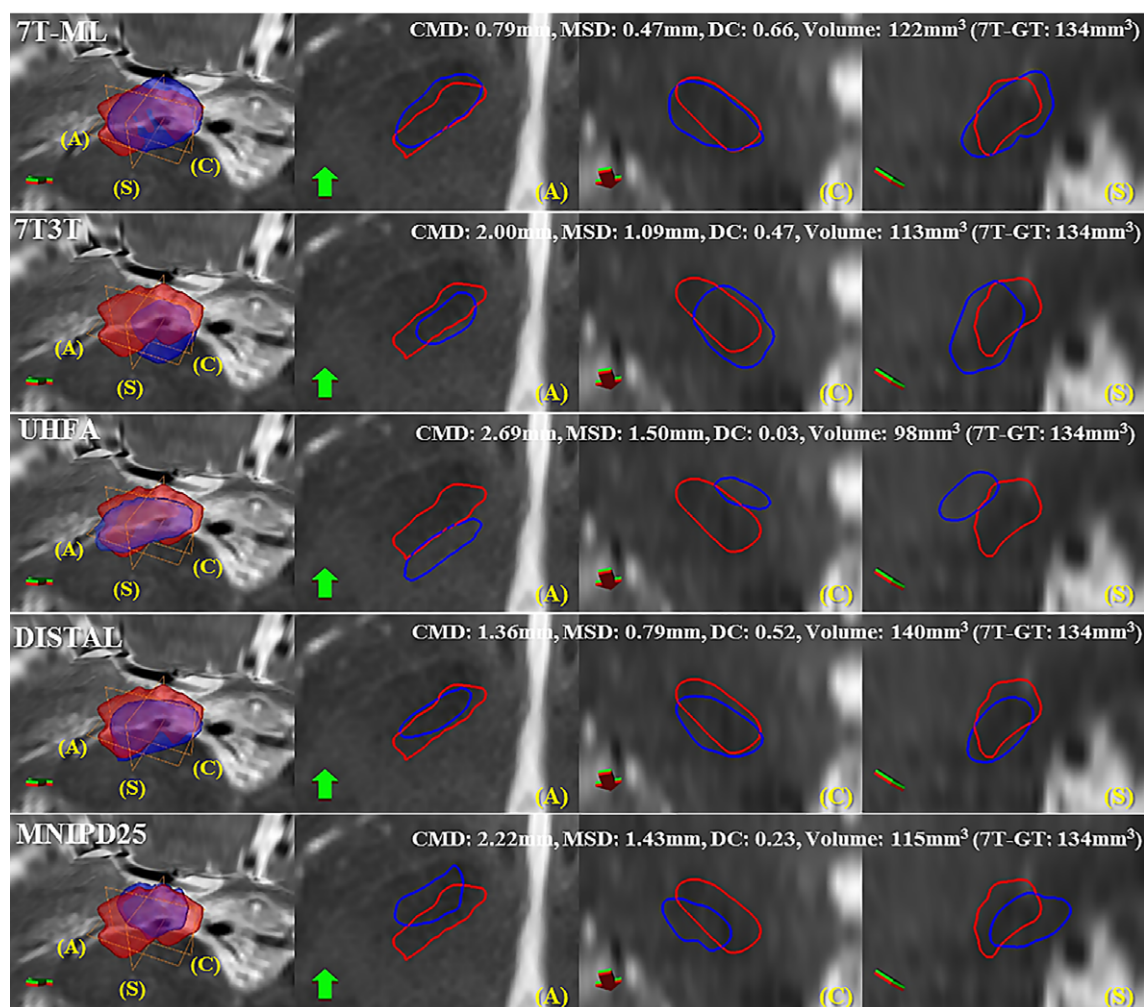


FIGURE 3 Centers of mass distances' histogram and visual examples of the proposed 7 T-ML and standard atlases (blue) overlaid with 7 T manual ground truth (red) from specific subjects. (a) Comparison of histograms of centers of mass distance measured on 160 STNs obtained from the different methods. (b) Visualization of the STNs from zone (i) and (iii) in (a), representing average centers of mass distance of the 7 T-ML and standard atlases, respectively on the 1.5 T MRI of an example PD patient (age: 51). (c) Visualization of the STNs from zone (i) and (ii) in (a), representing median centers of mass distance on the 3 T MRI of an example PD patient (age: 79). From left to right: 3D surface and contours in (A) axial, (C) coronal, and (S) sagittal planes along with arrows indicating the anterior direction. CMD: Centers of mass distance. MSD: Mean distance of surface points [Color figure can be viewed at wileyonlinelibrary.com]

error computed between the estimated error scores and the actual ones was lower for the regression forest method in comparison to the bagged partial least squares regression (Figure 8; $p < .05$ ANOVA). The smaller the number of training subsets, the higher the obtained mean squared error. This observation suggests that the robust framework was more effective when using the regression forest method with over 100 subsets.

4 | DISCUSSION

In this work, we proposed a computational framework to automatically localize and visualize the STN based on the standard clinical MRI, where it is not clearly identifiable, by taking advantage of our 7 T MRI database and machine learning. For validation, we used the STN ground truth that was defined on the 7 T MRI, with careful manual annotation and cross-validation. Particularly, the 7 T manual

segmentation was shown to be accurate and consistent with neurophysiological data (Duchin et al., 2018; Shamir et al., 2018). Moreover, the most-effective contacts were located in the dorso-lateral area of the 7 T manual STN segmentation, which matches other clinical reports (Garcia-Garcia et al., 2016; Herzog et al., 2004). These studies show that the 7 T manual segmentation data is highly consistent with biological and clinical measures. Once segmented accurately, the 7 T manually segmented STN can be transformed to a lower-resolution clinical image as done here to train the proposed 7 T-ML method, and still accurately represent the STN, although it may be hard to observe it on the standard clinical image. The proposed 7 T-ML based STN is highly consistent with the 7 T manual ground truth in center of mass, mean surface points, DC, and volume. Moreover, the population and location of the active contacts at the different sub-regions within the proposed 7 T-ML based predicted STN highly agrees with that of the 7 T manual ground truth. We believe that the consistency of our proposed method with neurological measures and clinical outcomes

TABLE 7 Distance between active contacts and STN's centers of mass computed from the 7 T manual ground truth, proposed 7 T-ML, 7 T3 T, UHFA, DISTAL, and MNIPD25 atlases

	Ground truth	7 T-ML	7 T3 T (Milchenko et al., 2018)	UHFA (Wang et al., 2016)	DISTAL (Ewert et al., 2017)	MNIPD25 (Xiao et al., 2017)
Average (mm)	2.39	2.43	3.56	3.24	5.14	5.61
Standard deviation	1.06	1.02	1.98	1.13	4.86	4.10

supports its use to facilitate the guidance of the DBS electrode based on direct targeting using standard clinical MRI of individual patients.

As summarized in Table 9, a variety of automatic STN segmentation methods based on MRI have been proposed. While most methods are not publicly available, state-of-the-art atlases and corresponding templates were freely downloadable, and thus we have tested them on our own clinical datasets. While some automatic approaches exploit the sufficient intensity information from contrast enhanced MR sequence or even higher field MRI (e.g., QSM, FGATIR, or 7 T), it remains to be investigated if these methods provide accurate segmentation of the STN on standard clinical T₂W MRI.

For this reason, we compared the performance of the proposed approach with that of state-of-the-art atlas-based methods: (1) 7 T3 T (Milchenko et al., 2018), (2) UHFA (Wang et al., 2016), (3) DISTAL (Ewert et al., 2017; Horn, Neumann, et al., 2017), and (4) MNIPD25 (Xiao et al., 2017). While these atlases reasonably localize and visualize the STN on subject-specific data they represent well, their accuracy and consistency on our tested clinical data were much lower than the proposed 7 T-ML method. A one-way ANOVA and post hoc test with Tukey's method showed that our 7 T-ML is significantly better than any other popular method we tested against. Moreover, the atlas-based STNs missed a large portion of post-operative active contacts, potentially resulting in sub-optimal planning, programming, or outcomes.

Standard atlases that are well defined in a normalized space could be of great value for retrospective population studies and are highly appreciated in the field (Horn, Kühn, et al., 2017; Horn, Neumann, et al., 2017; Horn, Reich, et al., 2017; Keuken et al., 2014). However, uncertainty in registration needs to be addressed when using such atlases for patient-specific STN-DBS targeting. Single atlas-based methods heavily rely on the registration quality between the atlas template and the clinical MRI from individual patients. Since the registration error may be larger on standard clinical images where the STN is not clearly visible or when the atlas template does not represent subject-specific data, atlas-based methods usually require further

revision of the segmentation. All the significant errors from standard atlases are not simple biases, which would be easy to correct, the errors are unpredictable and have large variance as well. Large variability of atlas-based results on our tested clinical data might explain this issue. Furthermore, an inaccurate definition of the STN in the atlas template may produce an additional error in patient-specific targeting (Ewert et al., 2017).

We provided comprehensive results from a variety of state-of-the-art atlases to discuss uncertainty in registration that might be induced by (1) morphological variability (normal vs. patient), (2) different contrast (magnetic field or modality) between atlas template and clinical data, and (3) its sub-optimization. This also confirms the benefits of our proposed 7 T-ML that combines accurate STN models and machine learning for prediction from clinical data.

More specifically, the DISTAL atlas (Ewert et al., 2017) and UHFA (Wang et al., 2016) were defined on the atlas template from normal subjects, and segmentation results on data from normal subjects were provided. The results on the PD-specific data might be deteriorated by morphological variability between the atlas template and the clinical data. Moreover, UHFA (Wang et al., 2016) utilized the 7 T T₂ W MRI atlas template, and thus the contrast discrepancy between the 7 T MRI and standard clinical data also might have affected the registration (this might also explain the often smaller STN volume found when computed with this atlas compared with other atlases-based results).

Although the MNIPD25 atlas (Xiao et al., 2017) is PD-specific, the MR modality of the STN atlas templates (T₂*W) is different from that of our standard clinical T₂W data. This might have caused an error in registration between the atlas template and the clinical data; this atlas was associated with worse performance and smaller volume than using other atlases.

Milchenko et al. (2018) created a 7 T atlas from elderly subjects and registered it onto the average template representing 3 T T₂ W MRIs from PD patients. While promising results on PD patient-specific data was reported, segmentation results on our own clinical

TABLE 8 Numbers of active contacts placed in each region within the per-patient ellipsoid representation of the STN computed based on the 7 T manual ground truth, proposed 7 T-ML, 7 T3 T, UHFA, DISTAL, and MNIPD25 atlases (see Figure 4 for the localization of each region)

Region	Ground truth	7 T-ML	7 T3 T (Milchenko et al., 2018)	UHFA (Wang et al., 2016)	DISTAL (Ewert et al., 2017)	MNIPD25 (Xiao et al., 2017)
1	4	5	2	2	3	1
2	9	9	5	3	7	1
3	5	6	3	2	2	5
4	10	7	5	9	2	3
5	0	0	2	0	0	0
6	0	0	3	0	2	0
7	0	0	2	0	0	0
8	0	1	1	0	0	0
9	3	3	8	15	15	21

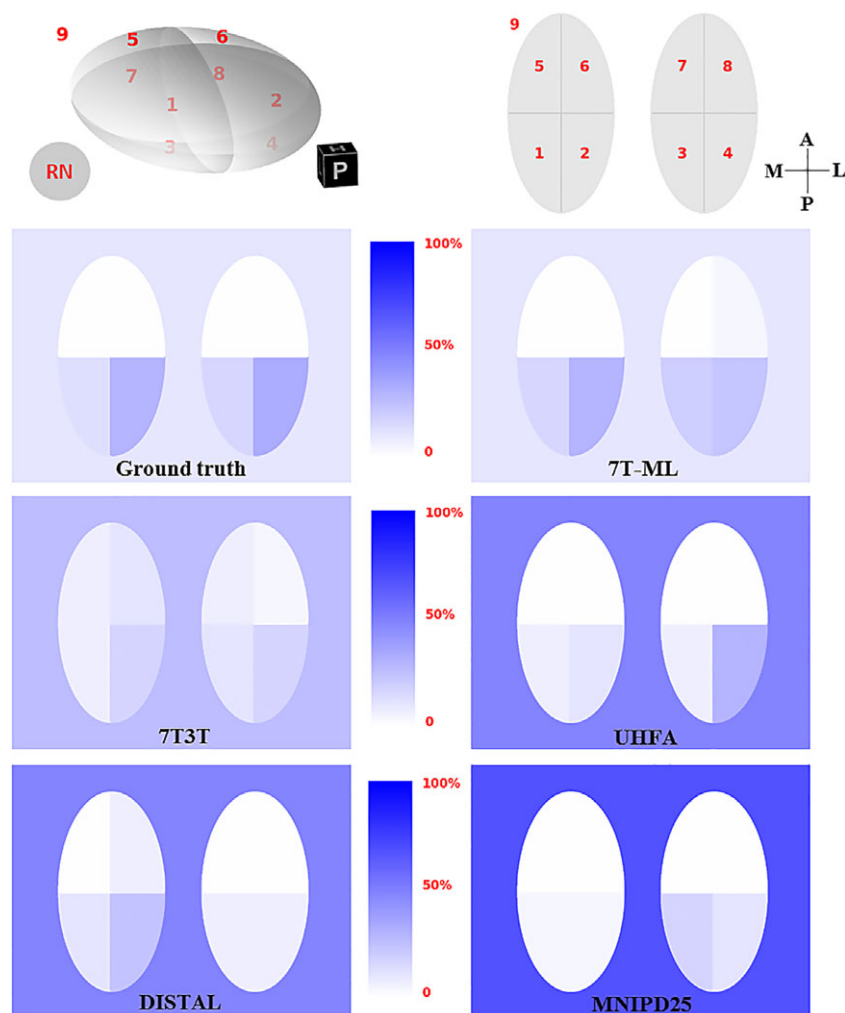


FIGURE 4 Density maps of the active contact population in different regions within the per-patient ellipsoid representation of the STN based on 7 T manual ground truth, proposed 7 T-ML, 7 T3 T, UHFA, DISTAL, and MNIPD25. Top left represents the region numbers in 3D and the right represents the corresponding numbers in the 2D slices along with posterior(P)/anterior(A) and medial(M)/lateral(L) axes (the example shows the right STN. Left STN is mirrored into M-L direction for the population analysis.) [Color figure can be viewed at wileyonlinelibrary.com]

data (mostly PD patients) were insufficient for clinical utilization (although it showed better performance than using other atlases). Sub-optimization in a single warp might have affected the results.

Table 10 presents quantitative comparison for centers of mass distance and DC of the STNs obtained using each method, clustered according to magnetic field strength. Specifically, 7 T3 T and DISTAL using a 3 T T₂ W MRI template, and UHFA using a 7 T T₂ W MRI template, showed much better performance on 3 T MRIs, that have closer appearance to those templates, than on 1.5 T MRIs. This illustrates that atlas-based methods require a template image closer to a given clinical MRI to improve their accuracy. Note that our 7 T-ML shows comparable accuracy regardless of the quality (and strength) of the clinical MRIs, although the algorithm was not trained on 3 T MRIs. In a complementary study, we have computed with the exact same method here introduced the STN on 3 T MRIs obtained in another center and observed that it is very similar (~1 mm) to the STN that was defined based on MER and blindly compared with our method (Shamir et al., 2018). These complementary studies show the high accuracy and consistency of the proposed computational method, regardless of the proxy used for representing the ground truth (7 T or MER).

Registration steps used in this work oftentimes caused large errors in atlas-based results that could lead to misplacement of the stimulating electrode and ineffective DBS treatment as shown in Figure 5. This might be attributed to the protocol, specifically tuned to clinical data, used in this study. An optimization of the registration processes between atlas templates and the used clinical data may improve the fitting of the atlas and the STN segmentation accuracy since atlas-based studies showed promising results using different registration protocols. Also, additional modalities can be incorporated to further improve the registration quality (Ewert et al., 2017). Such multi-spectral images, however, are not always available in clinical scenarios (e.g., only T₂W MRIs are given in our case). Moreover, assuring an optimization in the single registration mode still remains a challenging task even though the field has progressed (Viergever et al., 2016). More importantly, generalizing/scaling an optimized registration into cases in large-scale population is highly ill-posed. The full investigation of the registration performance is beyond the scope of this work, but is important for automatic targeting as here illustrated.

Our 7 T-ML minimizes the effort to find the optimal single warp in a robust and fully automatic way. More specifically, we register

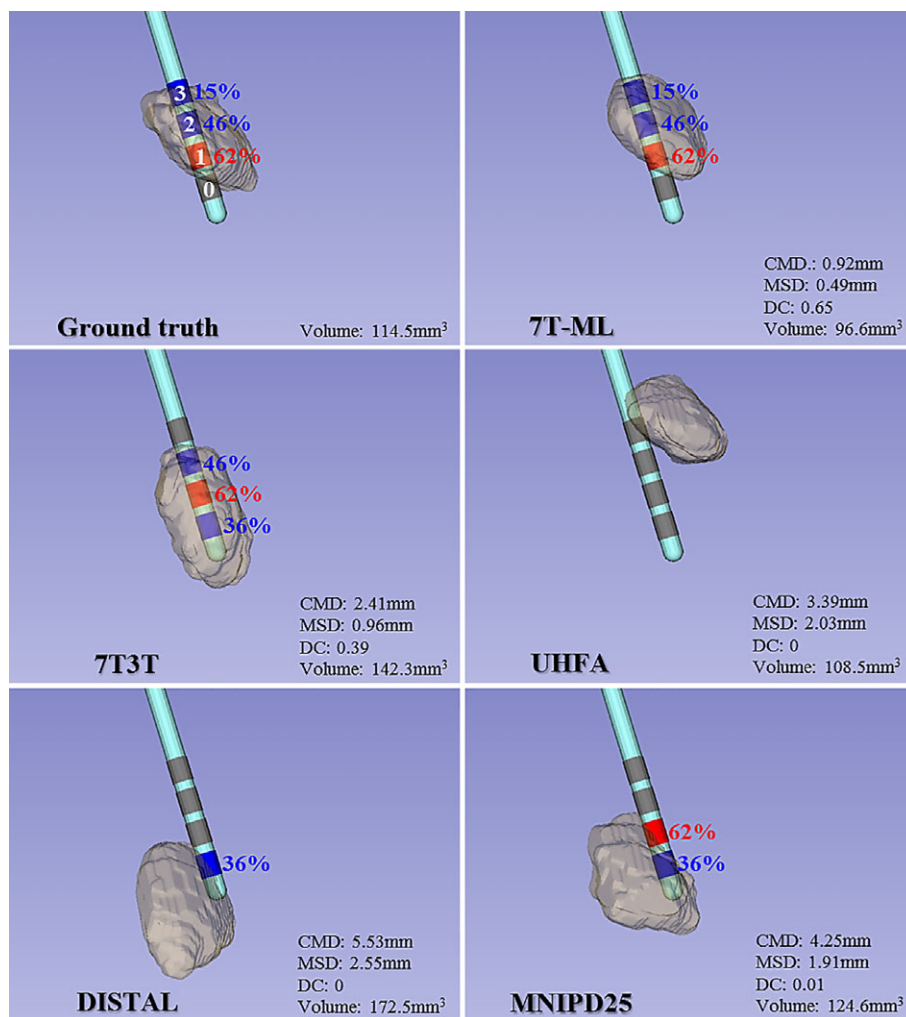


FIGURE 5 Comparison of the STN computed based on the 7 T manual ground truth, our proposed 7 T-ML, 7 T3 T, UHFA, DISTAL, and MNIPD25 atlases, overlaid with the electrode contacts for a specific PD patient (age: 48). The visualization is an example from zone (i) and (iii) in Figure 3a, representing average centers of mass distance of our 7 T-ML and standard atlases-based STNs, respectively. Electrode contacts were placed into the dorsal zone according to the ground truth. Similarly to the ground truth, our 7 T-ML STN completely includes the active contact (red, contact 1) associated with the best motor improvement (62%) and the active contacts (blue, contact 2 and 3) with lower motor improvement (15–46%). Dorsal STNs based on 7 T3 T, DISTAL, and MNIPD25 also would have been activated by the contacts associated with motor improvements, but they showed much larger errors in shape and location than 7 T-ML. This means that it might lead to misleading placement. None of contacts are placed in the dorsal STN based on UHFA. CMD: Centers of mass distance. MSD: Mean distance of surface points [Color figure can be viewed at wileyonlinelibrary.com]

clinical training images in our database directly to the clinical query image and select the most similar sets to reduce random error and bias in the single warp (Aljabar et al., 2009). Moreover, we refine the STN location and shape by learning 7 T knowledge. It should be noted that although our 7 T-ML framework uses the same registration steps that affected inaccurate atlas-based segmentation, thanks to these important additions the obtained results are significantly better than those obtained using the standard atlases. Large population studies on the shape and position of the STN and its sub-regions would obviously benefit from single atlases that are accurately defined on the MNI template (Horn, Neumann, et al., 2017). However, in this article, we stress that such high quality atlases on the template could be readily biased due to uncertainty in registration when using them for patient-specific targeting, and it is not trivial to find the optimal registration protocol, thereby resulting in tedious manual intervention. Our 7 T-ML overcomes the challenge in a fully automatic way.

Recent multiple atlases-based approaches showed promising results for the localization of the STN (Haegelen et al., 2013; Xiao, Fonov, et al., 2014; Xiao, Jannin, et al., 2014). In these studies, manual labeling was performed, based on the appearance of 3 T MRI, and automatic segmentation of the subcortical structures on the query patient was done by registration between the 3 T MRIs atlas templates and patient. While the automatic segmentation closely matches its manual STN on the patient images, taking advantage of multiple atlas templates with similar appearances (reducing the registration uncertainty), it is unclear if the STN that appeared as hypo-intense on the 3 T MRI reflects an accurate geometrical representation of the STN of individual subjects, especially in the dorso-lateral part that is critical for DBS targeting (Cho et al., 2011; Plantinga et al., 2014) (see also difference between volumes of 7 T manual ground truth and MNIPD25 atlas-based STN in Figure 2d). Moreover, employing such methods in standard clinical scenarios based on lower quality images

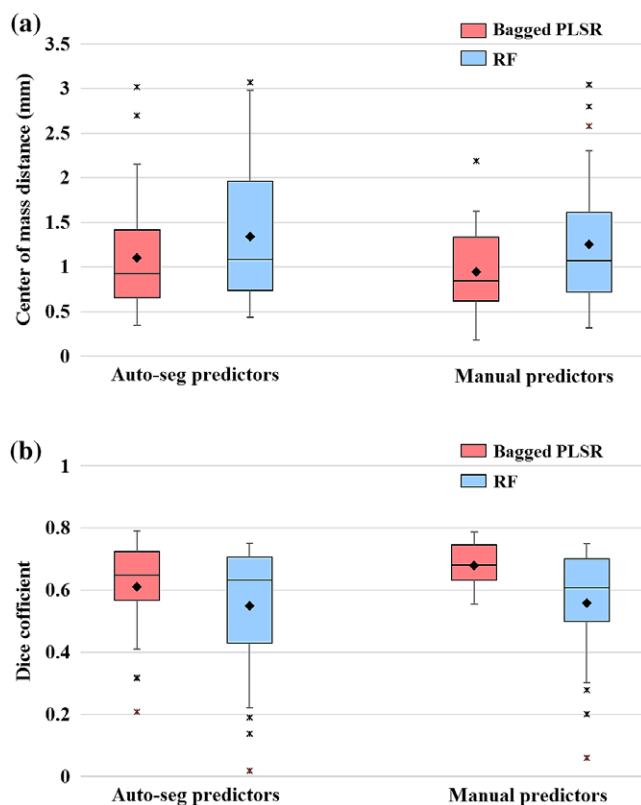


FIGURE 6 Comparison of (a) centers of mass distance and (b) DC values for predicted STNs with automatically and manually segmented predictors on the 1.5 T MRI of 15 PD patients [Color figure can be viewed at wileyonlinelibrary.com]

(e.g., 1.5 T MRI) may result in larger errors since the registration accuracy is expected to be lessened (Avants, Tustison, & Song, 2011; Ou, Akbari, Bilello, Da, & Davatzikos, 2014).

Using current standard clinical imaging protocols, it is not feasible to differentiate the STN from the SN in clinical 1.5 T or 3 T MRI (Abosch et al., 2010). As such, it remains unclear how approaches based on intensity and texture information on the target image (Bernard, Gemmar, Husch, & Hertel, 2012; Li, Jiang, Li, Zhang, & Meng, 2016) handle leakage around the border between the STN and SN. To explore this issue, we automatically segmented the STN using the active shape model and active appearance model framework (the same method that was used for segmentation of predictor structures in our 7 T-ML framework). While it reasonably localized the STN with the initialization from the 7 T priors, our 7 T-ML was still significantly better in average centers of mass distance and DC. Particularly, the size of the segmented STN was much smaller than the 7 T manual ground truth. This might be attributed to fuzzy boundary of the STN on the image.

Some approaches use pre-processed sequences or high quality data to make the STN more discernible or segment visible various mid-brain structures. For example, Garzón et al. (2017), Visser, Keuken, Forstmann, et al. (2016), Milletari et al. (2017), and Plassard et al. (2017) utilized contrast enhanced MR sequences (QSM or FGATIR) to visualize subcortical structures, including the STN. Garzón et al. (2017) mentioned that the algorithm showed lower accuracy on R_2^* and T_2W FLAIR images, indicating that it was specialized to high contrast data. Visser, Keuken, Forstmann, et al. (2016) automatically segmented the

STN on the 7 T multimodal MRI. Also, Visser, Keuken, Douaud, et al. (2016) segmented the striatum and globus pallidus—that are fairly visible on 1.5 T T_1 W MRI. Recently, several state-of-the-art methods using deep neural networks to segment brain structures are of interest (Bao & Chung, 2015; De Brébisson & Montana, 2015; Dolz, Desrosiers, & Ben Ayed, 2018; Shakeri et al., 2016). Similarly, they focused on segmentation of brain regions (e.g., Thalamus, Caudate, Putamen, Pallidum, etc.) that are discernable on the image. Low quality clinical images, where even manual segmentation of the STN is not possible, might lead to challenges when using deep learning for the task here considered (Zhou, Liu, & Huang, 2018). We could potentially apply deep learning architectures to instead segment the predictor structures (fairly visible on the clinical image) that are used to predict the STN.

The proposed 7 T-ML leverages our 7 T MRI database and machine learning to predict the STN that is not normally visible on the clinical MRI. It learns anatomical knowledge encoded from our 7 T training data, which are independent of image intensity values. The 7 T-ML, thereby, achieves comparable results to 7 T manual

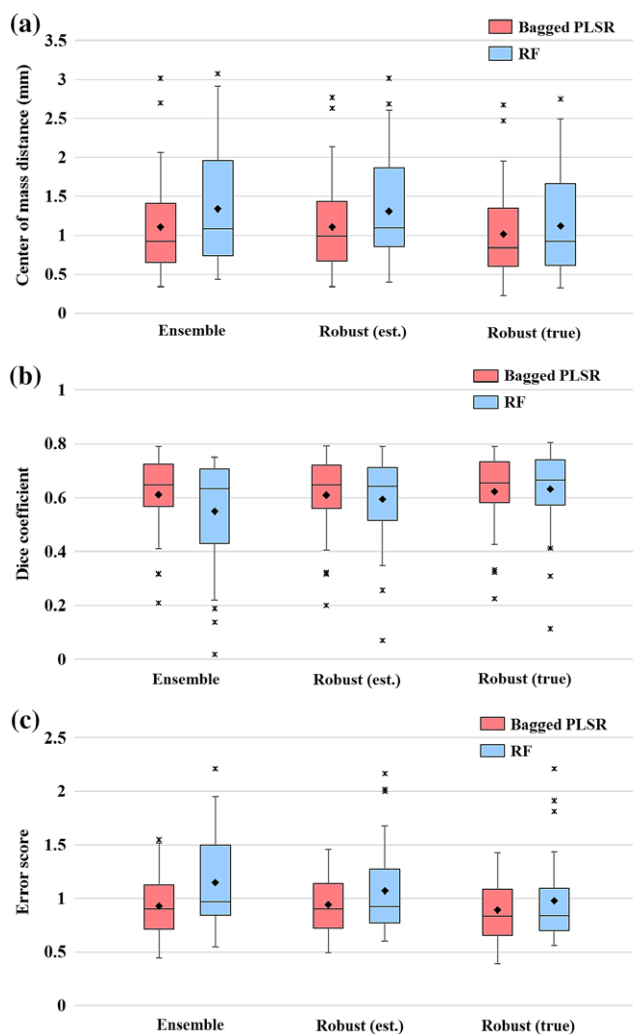


FIGURE 7 Comparison of (a) centers of mass distance, (b) DC values, and (c) error scores for ensemble STN prediction and robust prediction (for estimated error scores and actual ones) by bagged partial least squares regression and regression forest learning, with automatically segmented predictors and an ensemble size of 100 [Color figure can be viewed at wileyonlinelibrary.com]

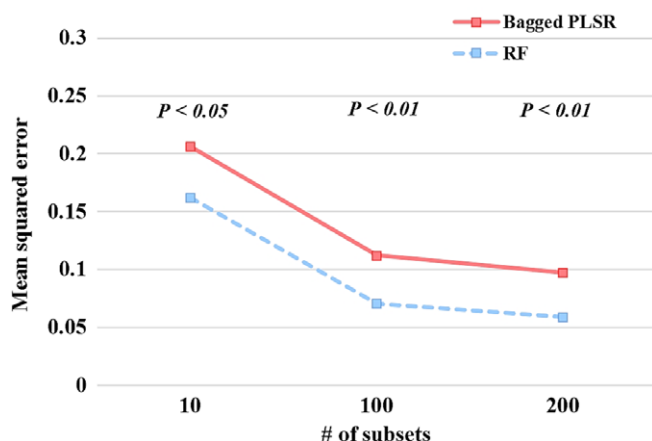


FIGURE 8 Average mean squared error between estimated error scores and true ones (computed from the 7 T manual ground truth STN) for predicted STN using the bagged partial least squares regression and the regression forest according to different number of training subsets within the robust framework [Color figure can be viewed at wileyonlinelibrary.com]

segmentation on the clinical image. We observed an average of 1.1 ± 0.6 mm in centers of mass distance (93% of the cases were better than 2 mm accuracy) from even lower quality data (1.5 T MRI) of selected 15 PD patient's data used for the in-depth study. This is consistent with the reported results for 160 STNs from 80 subjects and also validates that the 7 T-ML is robust on clinical 1.5 T MRI as well as 3 T MRI.

Similarly to the 7 T manual ground truth, the 7 T-ML based STN accurately localized and predicted the active contacts, especially in the posterior lateral region that is considered a motor territory often targeted for Parkinson's DBS (Plantinga et al., 2016). We should stress that we analyzed the spatial relationship between contact location and the STN using the ellipsoid representation within the STN in an individual patient data space. This is important since the transformation of the patient data into the common space and vice versa entails biases in the relationship. This also validates that our proposed 7 T-ML approach facilitates precise localization of the electrode's leads within patient-specific STN's sub-regions. A few active contacts were found to be outside the posterior part of our 7 T-ML and the 7 T manual ground truth STN. However, all of them were placed closely to the boundary, possibly having an overlap between the volume of tissue activated and the STN sub-region. Small resampling or registration errors may also explain this slight mismatch. Micro-lesion effect may bias contact localization as well (Granziera et al., 2008). Overall, our results show that while standard atlases do not achieve the accuracy and consistency needed for sub-region STN targeting, the proposed 7 T-ML is successful in this major DBS challenge: consistently providing accurate target localization. Importantly, the clinical feasibility of our 7 T-ML approach is further demonstrated, comparing to MER mapping in Shamir et al. (2018).

The 7 T-ML STN resulted in what could be considered at first glance as relatively low DC values, although the geometric measurements are within the tolerance level of the stereotactic frame used for the surgery (Shamir et al., 2009). The partial volume effect for small structures, such as the STN, under clinical imaging resolution affects the DC values (Hoffman, Huang, & Phelps, 1979). It has been also

reported that the size of objects affects the DC, where small structures are associated with smaller DC values (Rohlfing, 2012; Zou et al., 2004). Shamir, Duchin, Kim, Sapiro, and Harel (2016) provided numerical analysis of DC by modeling the clinical MRI resolution and center of mass error distribution in the STN manually segmented on the 7 T MRI, and an average upper bound DC was estimated at 64%. This indicates that the proposed 7 T-ML, with average 63.5% DC value, achieves near optimal accuracy. Recent automatic segmentation of the STN on the 7 T data also resulted in comparable DC values (Visser, Keuken, Forstmann, et al., 2016). We should note that our 7 T-ML targeted the STN on only standard clinical T₂W MRI data where its borders are not visible, while the automatic segmentations in the above studies were obtained from the 7 T multi-modal MRI with clear texture and boundary information.

The geometric distortion on the 7 T MRI and inaccurate co-registration between the 7 T MRI and clinical data may potentially affect the quality of the 7 T priors and thus cause errors in our 7 T-ML framework. Therefore, minimizing such biases was critical to increase the reliability. Experts in the team manually segmented the STN and it predictor subcortical structures by leveraging superior contrast and anatomical details on both 7 T T₂W and SWI and carefully cross-validated. Furthermore, Duchin et al. (2012) demonstrated clinical feasibility of the 7 T MRI by evaluating the distortion based on the co-registration quality between 7 T and 1.5 T MRI. Following the proposed protocols we performed the co-registration between 7 T and clinical MRI. Recently, we further validated that the 7 T manual STN segmentation is highly consistent with the MER data (Duchin et al., 2018; Shamir et al., 2018).

We also examined the effect of multiple factors in our proposed 7 T-ML approach. Generally, and as expected, the accuracy of predictors' segmentation highly affects the accuracy of the resulting STN prediction. We observed comparable STN prediction results using the manual and automatic predictors' (non-STN) segmentation. This indicates that an error level in predictors' segmentation on the clinical data was not influential in the STN prediction, and the automatic segmentation was near optimal. Comparable STN prediction accuracy was also observed for 10, 100, and 200 randomly selected subsets from the training set, but the estimation of STN prediction error was more accurate as the subset size increased. Weighting the training set based on the estimation of its contribution to the prediction accuracy reduced the variance observed with regression forest, but not bagged partial least squares regression. Therefore, the larger subset size results in more accurate estimation of STN prediction error that, in turn, helps to improve the STN prediction (an ensemble size of 100 is considered sufficient and used for validation of our 7 T-ML approach). Adding more subjects to our database may result in a significant correlation between the subset size and the STN prediction accuracy, which was not yet observed with training sets.

5 | FUTURE WORK AND CONCLUSIONS

We are currently investigating the shape refinement of the 7 T-ML STN on a standard clinical MRI. If the ventral border of the STN that is adjacent to the SN can be identified in an automatic way, the STN prediction can be further refined to facilitate an even more reliable

TABLE 9 Summary of automatic STN segmentation methods based on MRI

Author (year)	Method	Test data (subject-specific)				Implementation availability
		MRI modality	Magnetic field strength	Dataset source	The number of datasets (clinical details)	
Ewert et al. (2017)	Single atlas propagation	T ₂	1.5 T/3 T	IXI dataset	22 (normal)	3 T T ₂ template (MNI)/atlas
Wang et al. (2016)	Single atlas (7 T) propagation	T ₂	1.5 T	Own dataset	Unknown (normal)	7 T T ₂ template/atlas
Xiao et al. (2017)	Single atlas propagation	-	-	-	-	3 T T ₂ * template (MNI)/atlas
Milchenko et al. (2018)	Single atlas (7 T) propagation	T ₂	3 T	Own dataset	56 (PD)	3 T T ₂ template (MNI)/atlas
Haegelen et al., 2013	Multiple atlases propagation (patch based label fusion)	T ₁ /T ₂	3 T	Own dataset	10 (PD)	No
Xiao, Fonov, et al. (2014)	Multiple atlases propagation (patch based label fusion)	T ₁ -T ₂ * fusion	3 T	Own dataset	10 (PD)	No
Xiao, Jannin, et al. (2014)	Multiple atlases propagation (majority voting)	T ₁ /T ₂	3 T	Own dataset	10 (PD)	No
Plassard et al. (2017)	Multiple atlases (7 T) propagation	T ₁ /inversion recovery scan (O-IR)	3 T	Own dataset	9 (Normal)	No
Bernard et al. (2012)	Active shape model	SWI	3 T	Own dataset	24 (unknown)	No
Kim et al. (2014)	Active contour with multi-contrast edge maps and shape priors and non-overlapping constraints	SWI/T ₂	7 T	Own dataset	6 (Normal)	Available upon request
Li et al. (2016)	Level set (chan and vese with MLE initial contour)	T ₂	3 T	Own dataset	10 (PD)	No
Garzón et al. (2017)	Spatial priors based GMM of image intensities	QSM	3 T	Own dataset	40 (Normal)	Yes
Visser, Keuken, Forstmann, et al. (2016)	MRF shape prior model and intensity model	T ₂ *(QSM)	7 T	Own dataset	53 (Normal)	No (will be available in upcoming version of FSL)
Milletari et al. (2017)	Hough-voting to acquire mapping from CNN features	QSM	Unknown	Own dataset	55 (unknown)	No
Proposed 7 T-ML	Statistical shape and pose relationship learning from 7 T priors	T ₂	1.5 T/3 T	Own dataset	80 (PD/ET/Normal)	Available soon

TABLE 10 Centers of mass distance and DC value between the STNs obtained using the proposed 7 T-ML, 7 T3 T, UHFA, DISTAL, MNIPD25 atlases, and the 7 T manual ground truth STNs, grouped according to magnetic field strength in the training and validation sets

Magnetic field strength		Training set 1.5 T	Validation set	
			1.5 T	3 T
7 T-ML	CMD(mm)	1.28 ± 0.65	1.16 ± 0.48	1.22 ± 0.55
	DC	0.62 ± 0.13	0.66 ± 0.11	0.66 ± 0.08
7 T3 T (Milchenko et al., 2018)	CMD (mm)	2.63 ± 2.09	2.26 ± 1.16	1.92 ± 0.93
	DC	0.40 ± 0.24	0.41 ± 0.21	0.52 ± 0.11
UHFA (Wang et al., 2016)	CMD (mm)	3.28 ± 1.60	3.16 ± 1.42	2.21 ± 0.96
	DC	0.21 ± 0.22	0.17 ± 0.21	0.33 ± 0.23
DISTAL (Ewert et al., 2017)	CMD (mm)	4.35 ± 4.15	3.59 ± 3.41	1.83 ± 0.84
	DC	0.31 ± 0.27	0.39 ± 0.32	0.54 ± 0.17
MNIPD25 (Xiao et al., 2017)	CMD (mm)	4.41 ± 3.68	3.85 ± 2.65	4.11 ± 2.89
	DC	0.28 ± 0.26	0.18 ± 0.21	0.24 ± 0.18

Note. A one-way ANOVA and post hoc test (with Tukey's method) showed that the proposed 7 T-ML is significantly better than atlas-based methods in each measure ($p < .0001$).

targeting. However, it remains questionable if the intensities around the STN boundary are consistent across clinical MR datasets from a large population of patients and centers.

While we focused on localization of the STN in this manuscript, which is the most popular DBS target for Parkinson's disease, the approach presented here can be exploited for segmenting other structures such as the internal globus pallidus and Vim. With a greater number of centers beginning to target the internal globus pallidus for PD and given that it is the predominant target for dystonia, precise localization of the internal globus pallidus and its sub-regions may also prove valuable for physicians targeting this structure.

The identification of the STN on a standard clinical MRI is challenging. Therefore, more than one targeting method is incorporated today in DBS practice, often involving a more time-consuming and potentially extended-risk approach using intraoperative validation of electrode location with microelectrode recordings. Given that MER requires a level of expertise not typically found in most surgical centers, its utility and the ability of surgical sites to localize the STN and its sub-regions using this technique is highly variable across centers. To address these problems we introduced a patient-specific automatic software-only method for the visualization of the 3D STN location and shape from standard clinical MRI. The method incorporates a database of high-field 7 T MRI and a novel set of machine learning algorithms. The experimental results validated that our proposed 7 T-ML approach can automatically and accurately localize the STN and its sub-regions on standard clinical MRI. This work provides neurosurgeons and neurologists with accurate means for automatic patient-specific targeting of the STN and its sub-regions, potentially reducing the need for other approaches that may lengthen the procedure and/or be associated with a higher risk of side effects. Surgical Information Sciences, Inc. plans to make the 7 T-ML based STN segmentation tool available for its clinical use in the near future.

ACKNOWLEDGMENTS

This study was partially supported by the National Institutes of Health R01-NS085188, P41 EB015894, P30 NS076408, the University of Minnesota Udall center P50NS098573.

CONFLICT OF INTEREST

The authors are shareholders of Surgical Information Sciences, Inc. N.H. and G.S. are inventors of a patent related to high-resolution brain image system (U.S. Patent 9,412,076). G.S., N.H., Y.D., and J.K. are inventors of a patent related to brain region prediction (U.S. Patent 9,600,778).

ORCID

Jinyoung Kim  <https://orcid.org/0000-0002-5862-0363>

REFERENCES

- Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 97–106.
- Abosch, A., Yacoub, E., Ugurbil, K., & Harel, N. (2010). An assessment of current brain targets for deep brain stimulation surgery with susceptibility-weighted imaging at 7 tesla. *Neurosurgery*, 67, 1745–1756.
- Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., & Rueckert, D. (2009). Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage*, 46, 726–738.
- Avants, B., & Gee, J. C. (2004). Geodesic estimation for large deformation anatomical shape averaging and interpolation. *Neuroimage*, 23, S139–S150.
- Avants, B. B., Tustison, N., & Song, G. (2011). Advanced normalization tools (ANTS). *Insight Journal*, 2, 1–35.
- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*, 54, 2033–2044.
- Baka, N., Metz, C., Schaap, M., Lelieveldt, B., Niessen, W., & De Bruijne, M. (2011). Comparison of shape regression methods under landmark position uncertainty. In G. Fichtinger, A. Martel, & T. Peters (Eds.), *Lecture Notes in Computer Science 6893: Medical Image Computing & Computer-Assisted Intervention* (pp. 434–441). Berlin, Germany: Springer.
- Bao, S., Chung, A. C. S. (2015). *Multi-scale structured CNN with label consistency for brain MR image segmentation*. In The First Workshop on Deep Learning in Medical Image Analysis, Munich, Germany, 9 October 2015.
- Bernard, F., Gemmar, P., Husch, A., & Hertel, F. (2012). Improvements on the feasibility of active shape model-based subthalamic nucleus segmentation. *Biomed Engineering*, 57, 42–45.
- Blanc, R., Seiler, C., Székely, G., Nolte, L. P., & Reyes, M. (2012). Statistical model based shape prediction from a combination of direct

- observations and various surrogates: Application to orthopaedic research. *Medical Image Analysis*, 16, 1156–1166.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Cerrolaza, J. J., Reyes, M., Summers, R. M., González-Ballester, M. Á., & Linguraru, M. G. (2015). Automatic multi-resolution shape modeling of multi-organ structures. *Medical Image Analysis*, 25, 11–21.
- Cerrolaza, J. J., Villanueva, A., & Cabeza, R. (2012). Hierarchical statistical shape models of multiobject anatomical structures: Application to brain MRI. *IEEE Transactions on Medical Imaging*, 31, 713–724.
- Chakravarty, M. M., Bertrand, G., Hodge, C. P., Sadikot, A. F., & Collins, D. L. (2006). The creation of a brain atlas for image guided neurosurgery using serial histological data. *Neuroimage*, 30, 359–376.
- Chandran, A. S., Bynevelt, M., & Lind, C. R. P. (2015). Magnetic resonance imaging of the subthalamic nucleus for deep brain stimulation. *Journal of Neurosurgery*, 124, 96–105.
- Cho, Z.-H., Min, H.-K., Oh, S.-H., Han, J.-Y., Park, C.-W., Chi, J.-G., ... Lee, K. H. (2011). Direct visualization of deep brain stimulation targets in Parkinson disease with the use of 7-tesla magnetic resonance imaging. *Journal of Neurosurgery*, 113, 639–647.
- Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding*, 61, 38–59.
- Cootes, T. F., Taylor, C. J., Cooper, D. H., Graham, J. (1992): Training models of shape from sets of examples. In *British Machine Vision Conference* (pp. 9–18). Berlin: Springer-Verlag.
- Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., & Siddiqui, K. (2013). Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical Image Analysis*, 17, 1293–1303.
- D'Albis, T., Haegelen, C., Essert, C., Fernández-Vidal, S., Lalys, F., & Jannin, P. (2015). PyDBS: An automated image processing workflow for deep brain stimulation surgery. *The International Journal of Computer Assisted Radiology and Surgery*, 10, 117–128.
- Daniluk, S., Davies, K. G., Elias, S. A., Novak, P., & Nazzaro, J. M. (2010). Assessment of the variability in the anatomical position and size of the subthalamic nucleus among patients with advanced Parkinson's disease using magnetic resonance imaging. *Acta Neurochirurgica*, 152, 201–210.
- De Brébisson, A., Montana, G. (2015): Deep neural networks for anatomical brain segmentation. arXiv:1502.02445.
- Dolz, J., Desrosiers, C., & Ben Ayed, I. (2018). 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *Neuroimage*, 170, 456–470. <https://doi.org/10.1016/j.neuroimage.2017.04.039>
- Dormont, D., Seidenwurm, D., Galanaud, D., Cornu, P., Yelnik, J., & Bardinet, E. (2010). Neuroimaging and deep brain stimulation. *American Journal of Neuroradiology*, 31, 15–23.
- Duchin, Y., Abosch, A., Yacoub, E., Sapiro, G., & Harel, N. (2012). Feasibility of using ultra-high field (7 T) MRI for clinical surgical targeting. *PLoS One*, 7, e37328.
- Duchin, Y., Shamir, R. R., Patriat, R., Kim, J., Vitek, J. L., Sapiro, G., & Harel, N. (2018). Patient-specific anatomical model for deep brain stimulation based on 7 tesla MRI. *PLoS One*, 13, e0201469.
- Ewert, S., Plettig, P., Li, N., Chakravarty, M. M., Collins, L., Herrington, T. M., ... Horn, A. (2017). Toward defining deep brain stimulation targets in MNI space: A subcortical atlas based on multimodal MRI, histology and structural connectivity. *Neuroimage*, 170, 271–282.
- Frangi, A. F., Rueckert, D., Schnabel, J. A., & Niessen, W. J. (2002). Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *IEEE Transactions on Medical Imaging*, 21, 1151–1166.
- García-García, D., Guridi, J., Toledo, J. B., Alegre, M., Obeso, J. A., & Rodríguez-Oroz, M. C. (2016). Stimulation sites in the subthalamic nucleus and clinical improvement in Parkinson's disease: A new approach for active contact localization. *Journal of Neurosurgery*, 125, 1–12.
- Garzón, B., Sitnikov, R., Bäckman, L., & Kalpouzos, G. (2017). Automated segmentation of midbrain structures with high iron content. *Neuroimage*, 170, 199–209.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40, 33–51.
- Granziera, C., Pollo, C., Russmann, H., Staedler, C., Ghika, J., Villemure, J. G., ... Vingerhoets, F. J. G. (2008). Sub-acute delayed failure of subthalamic DBS in Parkinson's disease: The role of micro-lesion effect. *Parkinsonism & Related Disorders*, 14, 109–113.
- Guo, Y. (2010). A weighted cluster kernel PCA prediction model for multi-subject brain imaging data. *Statistics and Its Interface*, 3, 103–112.
- Haegelen, C., Coupé, P., Fonov, V., Guizard, N., Jannin, P., Morandi, X., & Collins, D. L. (2013). Automated segmentation of basal ganglia and deep brain structures in MRI of Parkinson's disease. *The International Journal of Computer Assisted Radiology and Surgery*, 8, 99–110.
- Halpern, C. H., Danish, S. F., Baltuch, G. H., & Jaggi, J. L. (2008). Brain shift during deep brain stimulation surgery for Parkinson's disease. *Stereotactic and Functional Neurosurgery*, 86, 37–43.
- Hamid, N. A., Mitchell, R. D., Mocroft, P., Westby, G. W. M., Milner, J., & Pall, H. (2005). Targeting the subthalamic nucleus for deep brain stimulation: Technical approach and fusion of pre- and postoperative MR images to define accuracy of lead placement. *The Journal of Neurology, Neurosurgery, and Psychiatry*, 76, 409–414.
- Harel, N., Sapiro, G. (2016). U.S. Patent 9,412,076, Minneapolis, MN: Surgical Information Sciences, Inc.
- Heimann, T., & Meinzer, H. P. (2009). Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13, 543–563.
- Herzog, J., Fietzek, U., Hamel, W., Morsnowski, A., Steigerwald, F., Schrader, B., ... Volkmann, J. (2004). Most effective stimulation site in subthalamic deep brain stimulation for Parkinson's disease. *Movement Disorders*, 19, 1050–1054.
- Hoffman, E. J., Huang, S. C., & Phelps, M. E. (1979). Quantitation in positron emission computed tomography: 1. Effect of object size. *The Journal of Computer Assisted Tomography*, 3, 299–308.
- Horn, A., Kühn, A. A., Merkl, A., Shih, L., Alterman, R., & Fox, M. (2017). Probabilistic conversion of neurosurgical DBS electrode coordinates into MNI space. *Neuroimage*, 150, 395–404.
- Horn, A., Neumann, W.-J., Degen, K., Schneider, G.-H., & Kühn, A. A. (2017). Toward an electrophysiological "sweet spot" for deep brain stimulation in the subthalamic nucleus. *Human Brain Mapping*, 38, 3377–3390.
- Horn, A., Reich, M., Vorwerk, J., Li, N., Wenzel, G., Fang, Q., ... Fox, M. D. (2017). Connectivity predicts deep brain stimulation outcome in Parkinson's disease. *Annals of Neurology*, 82, 67–78.
- Ivan, M. E., Yarlagadda, J., Saxena, A. P., Martin, A. J., Starr, P. A., Sootsman, W. K., & Larson, P. S. (2014). Brain shift during bur hole-based procedures using interventional MRI. *Journal of Neurosurgery*, 121, 149–160.
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5, 143–156.
- Kerl, H. U., Gerigk, L., Pechlivanis, I., Al-Zghloul, M., Groden, C., & Nölte, I. S. (2012). The subthalamic nucleus at 7.0 tesla: Evaluation of sequence and orientation for deep-brain stimulation. *Acta Neurochirurgica (Wien)*, 154, 2051–2062.
- Keuken, M. C., Bazin, P.-L., Schafer, A., Neumann, J., Turner, R., & Forstmann, B. U. (2013). Ultra-high 7T MRI of structural age-related changes of the subthalamic nucleus. *Journal of Neuroscience*, 33, 4896–4900.
- Keuken, M. C., Bazin, P. L., Crown, L., Hootsmans, J., Laufer, A., Müller-Axt, C., ... Forstmann, B. U. (2014). Quantifying inter-individual anatomical variability in the subcortex using 7T structural MRI. *Neuroimage*, 94, 40–46.
- Kim, H.-J., Jeon, B. S., Paek, S. H., Lee, J.-Y., Kim, H. J., Kim, C. K., & Kim, D. G. (2010). Bilateral subthalamic deep brain stimulation in Parkinson disease patients with severe tremor. *Neurosurgery*, 67, 626–632.
- Kim, J., Duchin, Y., Kim, H., Vitek, J., Harel, N., Sapiro, G. (2015): Robust prediction of clinical deep brain stimulation target structures via the estimation of influential high-field MR atlases. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 587–594).

- Kim, J., Duchin, Y., Sapiro, G., Vitek, J., Harel, N. (2015a): Clinical subthalamic nucleus prediction from high-field brain MRI. In 2015 I.E. 12th International Symposium on Biomedical Imaging (ISBI). (pp. 1264–1267).
- Kim, J., Duchin, Y., Sapiro, G., Vitek, J., Harel, N. (2015b): Clinical deep brain stimulation region prediction using regression forests from high-field MRI. In 2015 I.E. International Conference on Image Processing (ICIP). (pp 2480–2484).
- Kim, J., Lenglet, C., Duchin, Y., Sapiro, G., & Harel, N. (2014). Semiautomatic segmentation of brain subcortical structures from high-field MRI. *IEEE Journal of Biomedical and Health Informatics*, 18, 1678–1695.
- Krack, P., Hariz, M. I., Baunez, C., Guridi, J., & Obeso, J. A. (2010). Deep brain stimulation: From neurology to psychiatry? *Trends in Neuroscience*, 33, 474–484.
- Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial least squares (PLS) methods for neuroimaging: A tutorial and review. *Neuroimage*, 56, 455–475.
- Lauzon, M. L., McCreary, C. R., McLean, D. A., Salluzzi, M., & Frayne, R. (2016). Quantitative susceptibility mapping at 3 T: Comparison of acquisition methodologies. *NMR in Biomedicine*, 30, 3492.
- Lenglet, C., Abosch, A., Yacoub, E., De Martino, F., Sapiro, G., & Harel, N. (2012). Comprehensive in vivo mapping of the human basal ganglia and thalamic connectome in individuals using 7T MRI. Ed. Krystof Bankiewicz. *PLoS One*, 7, e29153.
- Li, B., Jiang, C., Li, L., Zhang, J., & Meng, D. (2016). Automated segmentation and reconstruction of the subthalamic nucleus in Parkinson's disease patients. *Neuromodulation: Technology at the Neural Interface*, 19, 13–19.
- Limousin, P., Krack, P., Pollak, P., Benazzouz, A., Ardouin, C., Hoffmann, D., & Benabid, A.-L. (1998). Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease. *The New England Journal of Medicine*, 339, 1105–1111.
- Liu, T., Eskreis-Winkler, S., Schweitzer, A. A. D., Chen, W., Kaplitt, M. G., Tsiouris, J., & Wang, Y. (2013). Improved subthalamic nucleus depiction with quantitative susceptibility mapping. *Neuroradiology*, 269, 216–223.
- Mallet, L., Schupbach, M., Diaye, K., Remy, P., Bardinet, E., Czernecki, V., ... Yelnik, J. (2007). Stimulation of subterritories of the subthalamic nucleus reveals its role in the integration of the emotional and motor aspects of behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 10661–10666.
- Matthews, I., & Baker, S. (2004). Active appearance models revisited. *IJCV*, 60, 135–164.
- Milchenko, M., Norris, S. A., Poston, K., Campbell, M. C., Ushe, M., Perlmutter, J. S., & Snyder, A. Z. (2018). 7T MRI subthalamic nucleus atlas for use with 3T MRI. *Journal of Medical Imaging and Radiation Oncology*, 5, 1.
- Milletari, F., Ahmadi, S.-A., Kroll, C., Plate, A., Rozanski, V., Maiostre, J., ... Navab, N. (2017). Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound. *Computer Vision and Image Understanding*, 164, 92–102.
- Ou, Y., Akbari, H., Bilello, M., Da, X., & Davatzikos, C. (2014). Comparative evaluation of registration algorithms in different brain databases with varying difficulty: Results and insights. *IEEE Transactions on Medical Imaging*, 33, 2039–2065.
- Pallavaram, S., Haese, P. F., Lake, W., Konrad, P. E., Dawant, B. M., & Neimat, J. S. (2015). Fully automated targeting using nonrigid image registration matches accuracy and exceeds precision of best manual approaches to subthalamic deep brain stimulation targeting in parkinson disease. *Neurosurgery*, 76, 756–763.
- Pallavaram, S., Yu, H., Spooner, J., Haese, P. F., Bodenheimer, B., Konrad, P. E., & Dawant, B. M. (2008). Intersurgeon variability in the selection of anterior and posterior commissures and its potential effects on target localization. *Stereotactic and Functional Neurosurgery*, 86, 113–119.
- Patel, N. K., Khan, S., & Gill, S. S. (2008). Comparison of atlas- and magnetic-resonance-imaging-based stereotactic targeting of the subthalamic nucleus in the surgical treatment of Parkinson's disease. *Stereotactic and Functional Neurosurgery*, 86, 153–161.
- Petersen, E. A., Holl, E. M., Martinez-Torres, I., Foltynie, T., Limousin, P., Hariz, M. I., & Zrinzo, L. (2010). Minimizing brain shift in stereotactic functional neurosurgery. *Operative Neurosurgery*, 67, ons213–ons221.
- Plantinga, B. R., Temel, Y., Duchin, Y., Uludag, K., Patriat, R., Roebroek, A., ... Harel, N. (2016). Individualized parcellation of the subthalamic nucleus in patients with Parkinson's disease with 7T MRI. *Neuroimage*, 168, 403–411.
- Plantinga, B. R., Temel, Y., Roebroek, A., Uludag, K., Ivanov, D., Kuijff, M. L., & ter Haar Romenij, B. M. (2014). Ultra-high field magnetic resonance imaging of the basal ganglia and related structures. *Frontiers in Human Neuroscience*, 8, 1–22.
- Plassard, A. J., D 'haese, P. F., Pallavaram, S., Newton, A. T., Claassen, D. O., Dawant, B. M., Landman, B. A. (2017): Multi-modal and targeted imaging improves automated mid- brain segmentation. In Proc SPIE Int Soc Opt Eng; 10133.
- Rao, A., Aljabar, P., & Rueckert, D. (2008). Hierarchical statistical shape analysis and prediction of sub-cortical brain structures. *Medical Image Analysis*, 12, 55–68.
- Rasouli, J., Ramdhani, R., Panov, F. E., Dimov, A., Zhang, Y., Cho, C., ... Kopell, B. H. (2017). Utilization of quantitative susceptibility mapping for direct targeting of the subthalamic nucleus during deep brain stimulation surgery. *Operative Neurosurgery*, 0, 1–8.
- Rathi, Y. (2006): Statistical shape analysis using kernel PCA. In Proc SPIE 6064:60641B-60641B-8.
- Rohlfing, T. (2012). Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. *IEEE Transactions on Medical Imaging*, 31, 153–163.
- Sapiro, G., Harel, N., Duchin, Y., Kim, J. (2017): U.S. Patent 9,600,778. Minneapolis, MN: Surgical Information Sciences, Inc.
- Shakeri, M., Tsogkas, S., Ferrante, E., Lippe, S., Kadoury, S., Paragios, N., Kokkinos, I. (2016): Sub-cortical brain structure segmentation using F-CNNs. Computer Vision and Pattern Recognition (cs.CV). arXiv: 1602.02130.
- Shamir, R. R., Duchin, Y., Kim, J., Marmor, O., Bergman, H., Vitek, J. L., ... Harel, N. (2018). Microelectrode recordings validate the visualization of subthalamic-nucleus based on 7T MRI and machine-learning for deep brain stimulation surgery. *Neurosurgery*, nyy212, <https://doi.org/10.1093/neuros/nyy212>.
- Shamir, R. R., Duchin, Y., Kim, J., Sapiro, G., & Harel, N. (2016). Segmentation overlap measures are biased to structure's size but correctable. *The International Journal for Computer Assisted Radiology and Surgery*, 11, S44–S45.
- Shamir, R. R., Joskowicz, L., Spektor, S., & Shoshan, Y. (2009). Localization and registration accuracy in image guided neurosurgery: A clinical study. *The International Journal of Computer Assisted Radiology and Surgery*, 4, 45–52.
- Starr, P. A., Christine, C. W., Theodosopoulos, P. V., Lindsey, N., Byrd, D., Mosley, A., & Marks, W. J. (2002). Implantation of deep brain stimulators into subthalamic nucleus: Technical approach and magnetic imaging—Verified electrode locations. *Journal of Neurosurgery*, 97, 370–387.
- Sung, J., Kanade, T., & Kim, D. (2007). A unified gradient-based approach for combining ASM into AAM. *The International Journal of Computer Vision*, 75, 297–309.
- The Deep-Brain Stimulation for Parkinson's Disease Study Group. (2001). Deep-brain stimulation of the subthalamic nucleus or the pars Interna of the Globus pallidus in Parkinson's disease. *The New England Journal of Medicine*, 345, 956–963.
- Tzimiropoulos, G., Pantic, M. (2013): Optimization problems for fast AAM fitting in-the-wild. In Proc IEEE Int Conf Comput Vis, pp. 593–600.
- Viergever, M. A., Maintz, J. B. A., Klein, S., Murphy, K., Staring, M., & Pluim, J. P. W. (2016). A survey of medical image registration – Under review. *Medical Image Analysis*, 33, 140–144.
- Visser, E., Keuken, M. C., Douaud, G., Gaura, V., Bachoud-Levi, A. C., Remy, P., ... Jenkinson, M. (2016). Automatic segmentation of the striatum and globus pallidus using MIST: Multimodal image segmentation tool. *Neuroimage*, 125, 479–497.
- Visser, E., Keuken, M. C., Forstmann, B. U., & Jenkinson, M. (2016). Automated segmentation of the substantia nigra, subthalamic nucleus and red nucleus in 7T data at young and old age. *Neuroimage*, 139, 324–336.

- Volkman, J. (2007). Update on surgery for Parkinson's disease. *Current Opinion in Internal Medicine*, 6, 511–515.
- Wang, B. T., Poirier, S., Guo, T., Parrent, A. G., Peters, T. M., & Khan, A. R. (2016). Generation and evaluation of an ultra-high-field atlas with applications in DBS planning. Ed. Martin A. Styner, Elsa D. Angelini. *SPIE Medical Imaging*, 9784, 97840H.
- Wells, W. M., III, Viola, P., Atsumi, H., Nakajima, S., & Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1, 35–51.
- Wold, H. (1982). Soft modeling : the basic design and some extensions. Syst under Indirect Obs causality, Struct Predict 2.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130.
- Xiao, Y., Fonov, V. S., Beriault, S., Gerard, I., Sadikot, A. F., Pike, G. B., & Collins, D. L. (2014). Patch-based label fusion segmentation of brainstem structures with dual-contrast MRI for Parkinson's disease. *The International Journal of Computer Assisted Radiology and Surgery*, 10, 1029–1041.
- Xiao, Y., Fonov, V., Chakravarty, M. M., Beriault, S., Al Subaie, F., Sadikot, A., ... Collins, D. L. (2017). A dataset of multi-contrast population-averaged brain MRI atlases of a Parkinson's disease cohort. *Data in Brief*, 12, 370–379.
- Xiao, Y., Jannin, P., Albis, T., Guizard, N., Haegelen, C., Lalys, F., ... Collins, D. L. (2014). Investigation of morphometric variability of subthalamic nucleus, red nucleus, and substantia nigra in advanced Parkinson's disease patients using automatic segmentation and PCA-based analysis. *Human Brain Mapping*, 35, 4330–4344.
- Zhou, Y., Liu, D., Huang, T. (2018): Survey of face detection on low-quality images. arXiv.
- Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M. C., Kaus, M. R., Haker, S. J., ... Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index. *Academic Radiology*, 11, 178–189.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Kim J, Duchin Y, Shamir RR, et al. Automatic localization of the subthalamic nucleus on patient-specific clinical MRI by incorporating 7 T MRI and machine learning: Application in deep brain stimulation. *Hum Brain Mapp*. 2019;40:679–698. <https://doi.org/10.1002/hbm.24404>