

Quantitative Susceptibility Mapping Radiomics with Label Noise Compensation for Predicting Deep Brain Stimulation Outcomes in Parkinson's Disease

Alexandra G. Roberts, MS^{1,2}, Jinwei Zhang, PhD³, Ceren Tozlu, PhD², Dominick Romano, MS², Sema Akkus, MD⁴, Heejong Kim, PhD², Mert R. Sabuncu, PhD^{1,2}, Pascal Spincemaille, PhD², Jianqi Li, PhD⁵, Yi Wang, PhD^{1,2}, Xi Wu, MD⁶, Brian H. Kopell, MD⁴

¹ Electrical and Computer Engineering, Cornell University

² Department of Radiology, Weill Cornell Medicine, Cornell University

³ Electrical and Computer Engineering, Johns Hopkins University

⁴ Department of Neurosurgery, Mount Sinai Hospital

⁵ School of Physics and Electronic Science, East China Normal University

⁶ Department of Neurosurgery, Changhai Hospital

Abstract

Parkinson's disease patients with motor complications are often considered for deep brain stimulation (DBS) surgery. DBS candidate selection involves an assessment known as the levodopa challenge test (LCT). The LCT aims to predict DBS outcomes by measuring symptom improvement accompanying changes in levodopa dosage. While used in the patient selection process, inconsistent LCT predictions have been widely documented, verified here with Pearson's correlation $r = 0.12$. Estimating symptom improvement to separate DBS responders and non-responders remains an unmet need. Quantitative susceptibility mapping (QSM) is routinely acquired for pre-surgical planning and depicts the iron distribution in substantia nigra and subthalamic nuclei. Iron deposition in these nuclei has correlated with disease progression and motor symptom severity. A novel QSM radiomic approach is presented using a regression model and features extracted from the pre-surgical targeting acquisition. Noise compensation in training labels improves outcome prediction in regression and classification models. The model predicts improvement in the unified Parkinson's disease rating scale (UPDRS-III) ($r = 0.75$). Predictive feature maps in deep gray nuclei offer contrast between responders and non-responders. The QSM radiomic approach has potential to improve DBS candidate selection by accurately estimating symptom improvement, eliminating difficult medication manipulation, and avoiding time-consuming evaluations to reduce patient and clinician burden.

1. Introduction

Deep brain stimulation (DBS) is a surgery to recalibrate neural circuitry in movement and psychiatric disorders including essential tremor, dystonia, and obsessive-compulsive disorder¹. DBS in the subthalamic nucleus (STN) is a widespread treatment for advanced Parkinson's disease (PD) motor symptoms². Selection of candidates for DBS involves assessing responsiveness to levodopa through the levodopa challenge test (LCT). The LCT has shown limitations in estimating improvements, yielding results with large uncertainty^{3,4} and is burdensome to patients and clinicians as it requires multiple assessments and medication withdrawal. Developing a consistent and convenient method for predicting DBS outcomes in PD is an important unmet clinical need. The purpose of this work is to demonstrate the potential of presurgical quantitative susceptibility maps (QSM) to improve patient selection.

From the phase image of multi-echo gradient echo sequences (mGRE), reconstructed QSMs⁵ provide superior contrast in deep gray nuclei during pre-surgical planning⁶⁻⁸. QSM is also useful in monitoring disease progression⁹. Recently, deep gray nuclei QSM radiomic features have been shown to improve PD diagnosis^{10,11}, and correlation with both PD motor symptoms^{12,13} and DBS binary outcomes¹⁴.

Radiomic features represent quantitative information in an image beyond what is discernable to the eye¹⁵. Such features are generally defined as first-order histogram-based features describing voxel values, second-order features representing joint probability distributions of voxel values, and region of interest (ROI) shape features¹⁶. Features act as biomarkers across imaging modalities and encode phenotype differences in a variety of pathologies.

The goal of the QSM radiomics model is to preoperatively predict post-surgical UPDRS-III improvement, referred to as “ground truth” or “label” data. As the UPDRS-III metric is measured on an interval scale, a linear regression with the least absolute shrinkage operator (Lasso) penalization technique is proposed using QSM radiomic features to predict DBS outcomes. Specifically, the outcome of interest is the ratio of postoperative to preoperative UPDRS-III scores, known as the UPDRS-III improvement. The UPDRS-III rating contains variability despite good agreement, described in supervised learning as “noise” ubiquitous in human measurements¹⁷. “Label noise” degrades regression models and supervised learning generally and label noise compensation has been shown to improve model performance¹⁸. Accordingly, a QSM radiomics Lasso model with noise compensation for DBS outcome prediction is presented.

2. Methods

2.1 Patient Cohort

This retrospective study was approved by the institutional ethics committee. Informed consent was obtained for forty patients undergoing bilateral STN-DBS. Patients were selected by the department of neurosurgery based on 1) primary PD with good response to levodopa; 2) decreased drug efficacy or motor complications; 3) adverse drug reactions; 4) tremors uncontrolled by drugs. Exclusion criteria were: 1) psychiatric disorders; 2) dementia; 3) diffuse cerebral ischemic lesions on MRI; 4) comorbidities interfering with surgery. The patient demographics were age 63.3 ± 7.45 years, sex ratio of 22 males to 18 females, disease duration 8.45 ± 2.87 years, and levodopa equivalent daily dosage (LEDD) of 323.78 ± 253.58 mg/day. For this cohort, the off medication UPDRS-III score was 59.63 ± 20.01 , the on medication UPDRS-III score was 23.65 ± 17.76 and the on medication, off stimulation UPDRS-III score was 25.00 ± 13.46 . The LEDD percent reduction following surgery was $54\% \pm 31\%$. The absolute UPDRS-III post-operative improvement was 35 ± 17 .

2.2 Data Acquisition

QSMs from mGRE data were reconstructed with MEDI-L₁¹⁹. Pre-surgical UPDRS-III scores u on and off-medication from LCT and post-surgical UPDRS-III scores on-stimulation and off-medication were collected²⁰ and the UPDRS-III improvement was computed:

$$y = \frac{u_{\text{presurgical,on-medication}} - u_{\text{postsurgical,on-stimulation,off-medication}}}{u_{\text{presurgical,on-medication}}} \quad (1)$$

The substantia nigra (SN) and STN were segmented with radiologist supervision. Collected sample size was 40, with 3 cases excluded by motion artifacts. The proposed pipeline is discussed in section 2.3 *Preprocessing* and illustrated in Figure 1.

2.3 Preprocessing

Figure 1 depicts the proposed pipeline with QSM and ROI inputs. Features were extracted from each ROI (joint entropy feature map shown) and passed to a Lasso model after normalization, label augmentation, and recursive feature selection. A total of 93 features were extracted over 17 image types available in PyRadiomics²¹. Additionally, 14 shape features were extracted on the original image for a total of 1595 features per ROI.

2.4 Model

A linear regression with Lasso penalization technique implemented with `SciKit-Learn` was performed to predict UPDRS-III improvement outcome using 1595 features per ROI²².

$$w^* = \operatorname{argmin} \frac{1}{2N} \|y - X_{\Phi} w\|_2^2 + \lambda \|w\|_1 \quad (2)$$

Here, UPDRS-III improvement outcomes y are predicted using optimal weights w^* and test features \hat{X}_{Φ} , λ is a regularization parameter, and $N = 37$ denotes the number of patients. Given the dataset size, a leave-one-out approach for subject-specific models²³ was used with each patient withheld and remaining $N - 1$ patients were used for training.

2.5 Noise Compensated Data Augmentation

The Lasso training dataset can be augmented using computed label noise statistics. In this work, a Gaussian distribution was fit to the $N - 1$ original UPDRS-III improvements as $y_t^0 \sim \mathcal{N}(\mu_{N-1}, \sigma_{N-1})$, with mean μ_{N-1} and standard deviation σ_{N-1} . For each model with original features X_t^0 , noise instances ϵ_q were drawn from the Gaussian distribution $\mathcal{N}(0, z\sigma_{N-1})$, with $z = 1.96$, and added to samples in the dataset y_t^0 :

$$\{X_t^Q, y_t^Q\} = \left\{ X_t^0, \dots, X_t^0, y_t^0, y_t^0 + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_1 \end{bmatrix}, y_t^0 + \begin{bmatrix} \epsilon_2 \\ \vdots \\ \epsilon_2 \end{bmatrix}, \dots, y_t^0 + \begin{bmatrix} \epsilon_Q \\ \vdots \\ \epsilon_Q \end{bmatrix} \right\} \quad (3)$$

This newly augmented dataset of size $(Q + 1)(N - 1)$ with labels $y_t^Q = Y + \epsilon = \begin{bmatrix} y_0 & y_0 & \dots & y_0 \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-1} & y_{N-1} & \dots & y_{N-1} \end{bmatrix} + \begin{bmatrix} 0 & \epsilon_1 & \dots & \epsilon_Q \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \epsilon_1 & \dots & \epsilon_Q \end{bmatrix}$. Notice the original labels are preserved in the first column.

Here, the use of this augmentation was referred to as “label noise compensation.” For comparison, Lasso was trained with other data augmentation strategies including bootstrap, wild bootstrap²⁴, and synthetic minority oversampling with Gaussian noise (SMOBN)²⁵.

2.6 Training

Nested cross-validation was used to train and test models for each subject. In the outer loop, the model was trained with $N - 1$ subjects and tested with one subject. For the inner loop, k -fold cross-validation was performed.

Recursive feature selection was used to reduce the optimal number of normalized features m^* from all ROI features $p = 6380$, with step size $\Delta m = 1000$ features, for input to Lasso.

Splits $k = \{2,3,4,5\}$, selected features m , and regularization λ were chosen by minimizing training error. The minority class ($y_{train} < 0.3$) was resampled for stratified cross-validation.

The dataset was augmented ($Q = 10$) with label noise. Least angle regression solver weights (machine epsilon regularization $\epsilon_c = 0.1$) were initialized at 0. The model was retrained over the entire ($N - 1$) dataset using m^* features, split k^* , and regularization λ^* from the inner loop. Training was repeated with standard and noise compensated classifiers.

For each patient, the prediction \hat{y} was a linear combination of features and weights, $\hat{y} = \hat{X}_\phi w^*$.

2.7 Evaluation

Performance of LCT and Lasso regressors was evaluated using linear regression (Pearson correlation r , slope \mathcal{M} , intercept b , significance p) between the observed UPDRS-III improvement and predicted UPDRS-III improvement. Model accuracy was measured by correlation between the observed and predicted outcome. To assess augmentations over different initial noise states, models were evaluated under 10 additional seeds.

Classifiers and regressors were evaluated with receiver operating characteristic (ROC) curves and analogous regression error characteristic (REC) curves, respectively. The frequency of features with nonzero weights were compared for models. The predictive power of a feature is measured by the number of nonzero weights w^* across N patient-specific models. Feature maps were compared in section 3.2 *Predictive Features* for non-responders ($y < 0.3$) and strong responders ($y > 0.85$), a threshold selected to balance with the number of non-responders.

3. Results

3.1 Model Performance

Figure 2 illustrates model UPDRS-III improvement predictions. LCT estimates failed to correlate ($r = 0.12, p = 0.48$, Figure 2a) with true outcome measured by Equation 1. QSM radiomic predictions correlated with improvements ($r_{max} = 0.75, p < 0.001$) for noise compensated Lasso, Lasso, bootstrap Lasso, wild bootstrap Lasso, and SMOGN (Figures 2b-2f). The noise compensated Lasso model correctly classified all outcomes: “response” with improvement $y \geq 0.3$ in the first quadrant or “non-response” in the third quadrant in Figure 1b.

Noise compensation also increases classifier ROC-AUC from 0.86 to 0.99 (Figure 3a). The regressor REC measures accuracy (vertical axis) of a regression estimator at allowable error

bound (horizontal axis) described by the absolute deviation $|y - X_{\phi}w|$. Regressor REC-AUC was 0.82 for radiomics models and 0.73 for LCT (Figure 3b). Note the radiomics models outperform the mean reference estimator, unlike LCT.

When noise compensated, wild bootstrap, and SMOGN results were fit over the additional noise states, the median result over all fits was found and the regression correlation coefficient approached approximately $r \approx 0.70$ for each augmentation strategy. The multi-noise state model performances were ($r = 0.73, \mathcal{M} = 1.15, b = -0.1, accuracy = 0.97$ for noise compensated Lasso, ($r = 0.70, \mathcal{M} = 1.12, b = -0.07, accuracy = 0.97$) for SMOGN Lasso, and ($r = 0.68, \mathcal{M} = 0.94, b = 0.03, accuracy = 0.95$) for wild bootstrap Lasso. All correlations were highly significant ($p < 0.0001$).

3.2 Predictive features

The most predictive features for the considered ROIs were found on the QSM Coiflet wavelet decomposition of the STN and the gray level dependence matrix (GLDM) from the SN QSM (Figure 4). Specific features were SN large dependence high gray level emphasis (a measure of coarseness) and STN gray level size zone matrix (GLSZM) entropy shown in Figure 5 for strong responders ($y > 0.85$) and non-responders ($y < 0.3$). SN coarseness was higher in non-responders (0.92) versus responders (0.65). Responder STN wavelet entropy was higher (0.61) versus non-responders (0.52). However, neither difference was significant ($\alpha = 0.05$).

4. Discussion

4.1 Summary

All models that used QSM radiomics predict DBS UPDRS-III improvements better than LCT. Particularly, label noise compensated models using QSM radiomics performed the highest prediction accuracy. This QSM radiomic prediction is consistent with previous findings of ROI analyses including textual features on PD symptom burden^{13,26}. This study is significant in that a numerical prediction of UPDRS-III improvement is offered from routinely acquired QSM used for presurgical targeting.

4.2 Prior Works

There are large uncertainties in LCT predictions^{20,27}. LCT fails to predict levodopa-resistant symptoms such as tremor improvement²⁸ and certain dyskinesias²⁹. Additional inaccuracies in LCT result from dose dependence³⁰ and disease duration³ while inducing burdensome withdrawal symptoms²⁸.

Magnetoencephalography³¹ and transcranial direct current stimulation (tDCS)³² both demonstrated correlation with DBS improvement, yet neither is routinely acquired in pre-surgical planning. MRI-based brain morphology and functional connectomes have been used to predict DBS motor outcome with limited accuracy³³⁻³⁷. QSM from routinely acquired MRI provides predictive features from deep gray nuclei iron distributions. SN texture features reflect degeneration of nigral neurons, reducing dopamine receptor stimulation where iron is a cofactor in dopamine synthesis³⁸. Loss of GABAergic inhibition in the STN results in iron deposition encoded by texture features. Dysregulation of SN and STN iron disrupts locomotor activation and inhibition³⁹, and subsequent motor symptoms measured by UPDRS-III.

Of the compared augmentations, wild bootstrap is the most similar method to noise compensation. Though it improves non-responder estimates, modeling label variability in noise compensation is more effective. The Lasso model provides a significant and accurate prediction of DBS improvements using routinely acquired QSM under typical levodopa regimen.

4.3 Model Interpretation

The prediction is a weighted sum of relevant radiomic features in UPDRS-III improvements. Frequency of nonzero weights across patient-specific models indicate features important in outcome prediction. Comparison of responder and non-responder feature maps demonstrate distinction in both the STN and SN. As iron deposition in the SN is associated with

PD progression⁴⁰, predictive feature SN coarseness, or prevalence of high susceptibility regions, suggests elevated iron, advanced tissue damage, or less viable circuitry. This finding is supported by DBS patient selection guidance that considers “very late stages” of PD to be an impediment to symptom improvement⁴¹. Given the reports of SN iron accumulation and disease progression in PD⁴², the association between the SN coarseness feature and non-responders may support this guidance. Also, loss of QSM STN gradient, captured in predictive wavelet entropy features²¹, correlates with loss of motor function²⁶, which is consistent with the wavelet decomposition features identified in previous work¹⁴. These findings affirm both the observation that inhomogeneity of the STN iron distribution and elevated susceptibility in the SN iron distribution may be predictive of motor symptom severity and patient responsiveness to DBS. Yet, these features alone are insignificant between responders and non-responders, underscoring the need for multivariable models such as Lasso for patient selection.

4.4 Limitations

This work has several limitations. As a feasibility study, further validation is required. DBS eligibility requires LCT improvement, generating bias in the dataset against patients without LCT improvement who benefit from DBS and precluding healthy subjects entirely. Additionally, the sample size is small compared to the number of features. Finally, the Gaussian distribution is chosen for label noise modeling but may oversimplify the UPDRS-III interval scale.

5. Conclusion

In summary, QSM can provide predictive radiomic features to improve DBS patient selection and better outcome prediction over LCT, in addition to assisting in presurgical targeting. Alongside motor symptoms, non-motor predictions may be further explored (apathy scales, linked to microstructure⁴³ and depression inventories). The patient population in this work featured STN-DBS, but the ventral intermediate nucleus and globus pallidus pars interna are also well-visualized on QSM for pre-surgical targeting and should be considered in future studies⁴⁴. Beyond deep gray nuclei, whole brain ROIs may be included⁴⁵. Label noise distributions should be investigated to reflect multiple raters present in the clinical setting.

6. Disclosure

Pascal Spincemaille and Yi Wang are co-inventors on QSM-related patents owned by Cornell University and have ownership shares in MedImageMetric LLC. The remaining authors have no conflicts of interest to report.

7. Data Availability Statement

Code and data will be available upon publication at:

<https://github.com/agr78/RadDBS-QSM>.

8. Author Contributions

AGR: Conceptualization; Formal analysis; Software; Validation; Visualization; Original draft. **JZ:** Conceptualization; Formal analysis; Software; Validation; Visualization; Review & editing. **CT:** Formal analysis; Validation; Visualization; Review & editing. **DR:** Conceptualization; Review & editing. **SA:** Data curation; Review & editing. **HK:** Formal analysis; Validation; Visualization, Review & editing. **MRS:** Resources; Supervision; Review & editing. **PS:** Conceptualization; Formal analysis; Supervision; Review & editing. **JL:** Funding acquisition; Resources; Supervision; Review & editing. **YW:** Conceptualization; Funding acquisition; Resources; Supervision; Review & editing. **XW:** Data acquisition; Review & editing. **BHK:** Investigation; Methodology; Resources; Supervision; Writing - review & editing.

9. Funding Information

This research was supported by grants from the National Institutes of Health R01 NS095562 and National Multiple Sclerosis Society TA-2204-39428 and FG-2008-36976.

10. Figures

Figure 1. Proposed pipeline with inputs QSM and ROIs. Features were extracted from the input QSM χ over each ROI (joint entropy feature map shown) and passed to a Lasso regression model after normalization, label noise compensation augmentation, and recursive feature selection. A total of 93 features were extracted over 17 image types available in PyRadiomics. Additionally, 14 shape features were extracted on the original image for a total of 1595 features per ROI.

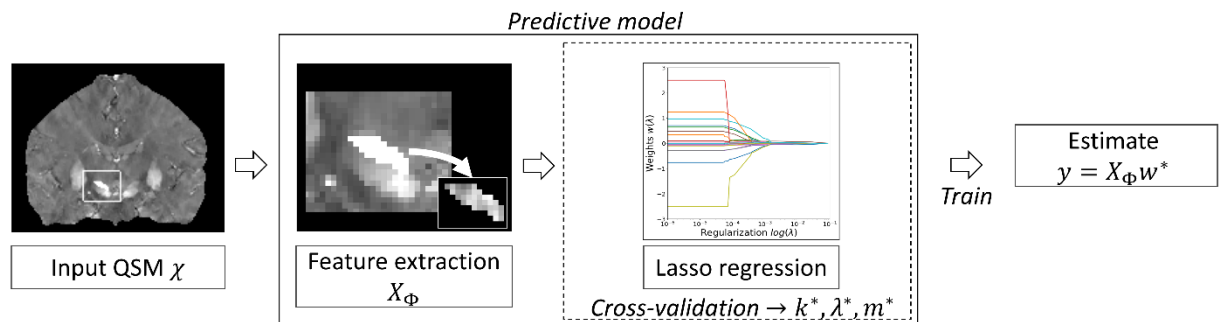


Figure 2. Predicted (horizontal axis) UPDRS-III improvement and true outcomes (vertical axis) using predictors: a) LCT, b) Noise Compensated Lasso (proposed), c) Standard Lasso, d) Bootstrap Lasso, e) Wild Bootstrap Lasso, f) SMOGN Lasso. Blue points indicate a responsive surgical outcome (responder) while red points indicate a non-responsive surgical outcome (non-responder). Note the noise compensated Lasso model correctly classifies all cases.

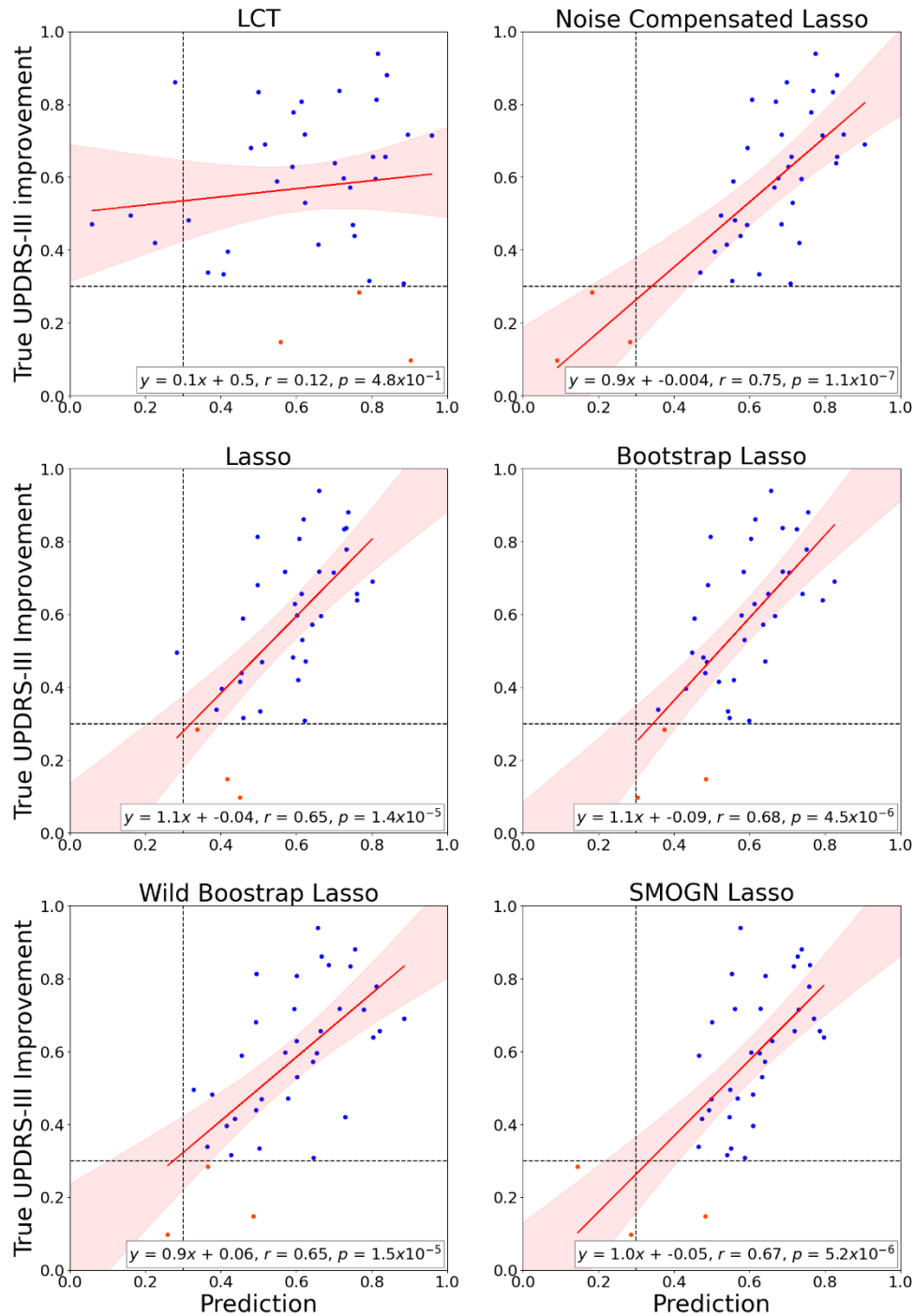


Figure 3. Comparison of classifiers with and without noise compensation (a) and regression estimators with various data augmentation during training (b). The ROC shows an improvement of $\Delta AUC = 0.13$. The REC measures accuracy (vertical axis) of a regression estimator at allowable error bound (horizontal axis) described by the absolute deviation $|y - X_{\phi}w|$. Note the radiomics models outperform the mean reference estimator, unlike LCT.

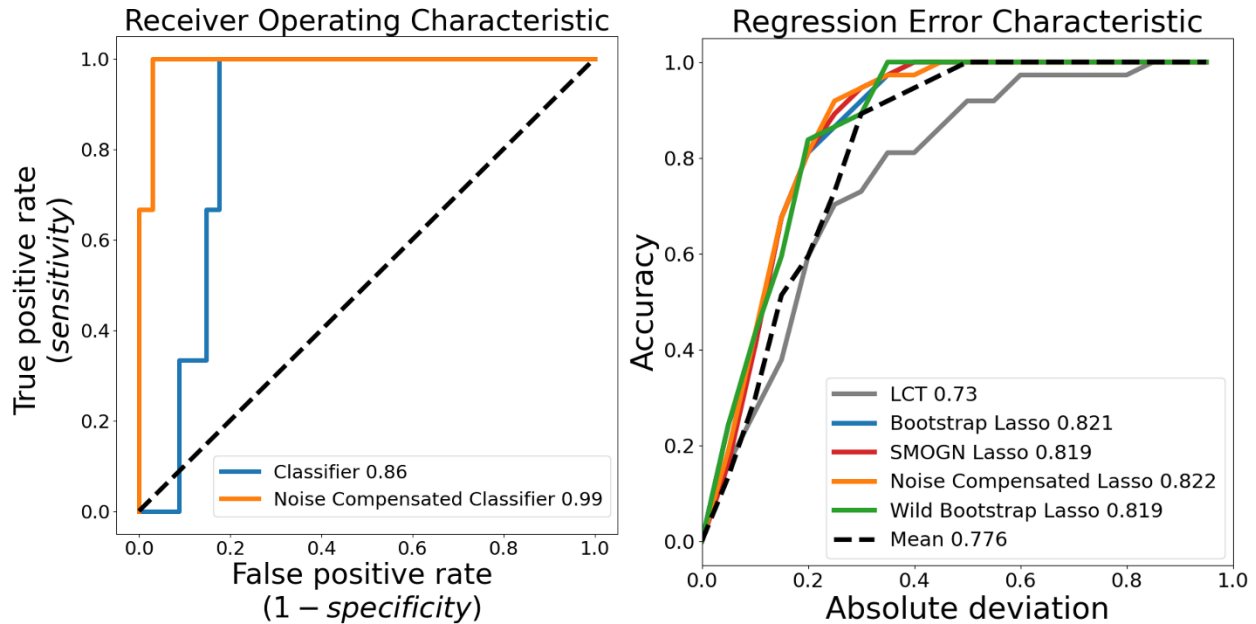


Figure 4. Predictive radiomic features sorted by frequency in $N = 37$ models. All models use the first feature, GLDM large dependence high gray level emphasis, which measures the coarseness of the substantia nigra. In the subthalamic nucleus, the wavelet GLZSM entropy is most predictive, which measures the heterogeneity of the susceptibility distribution.

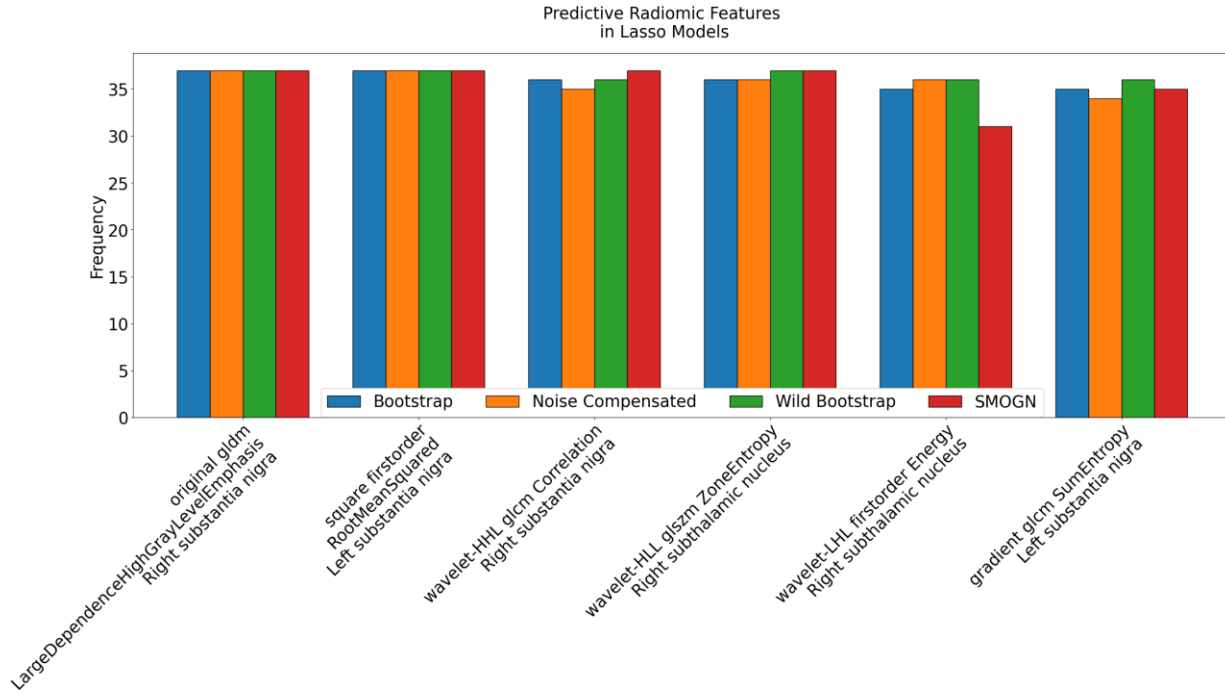
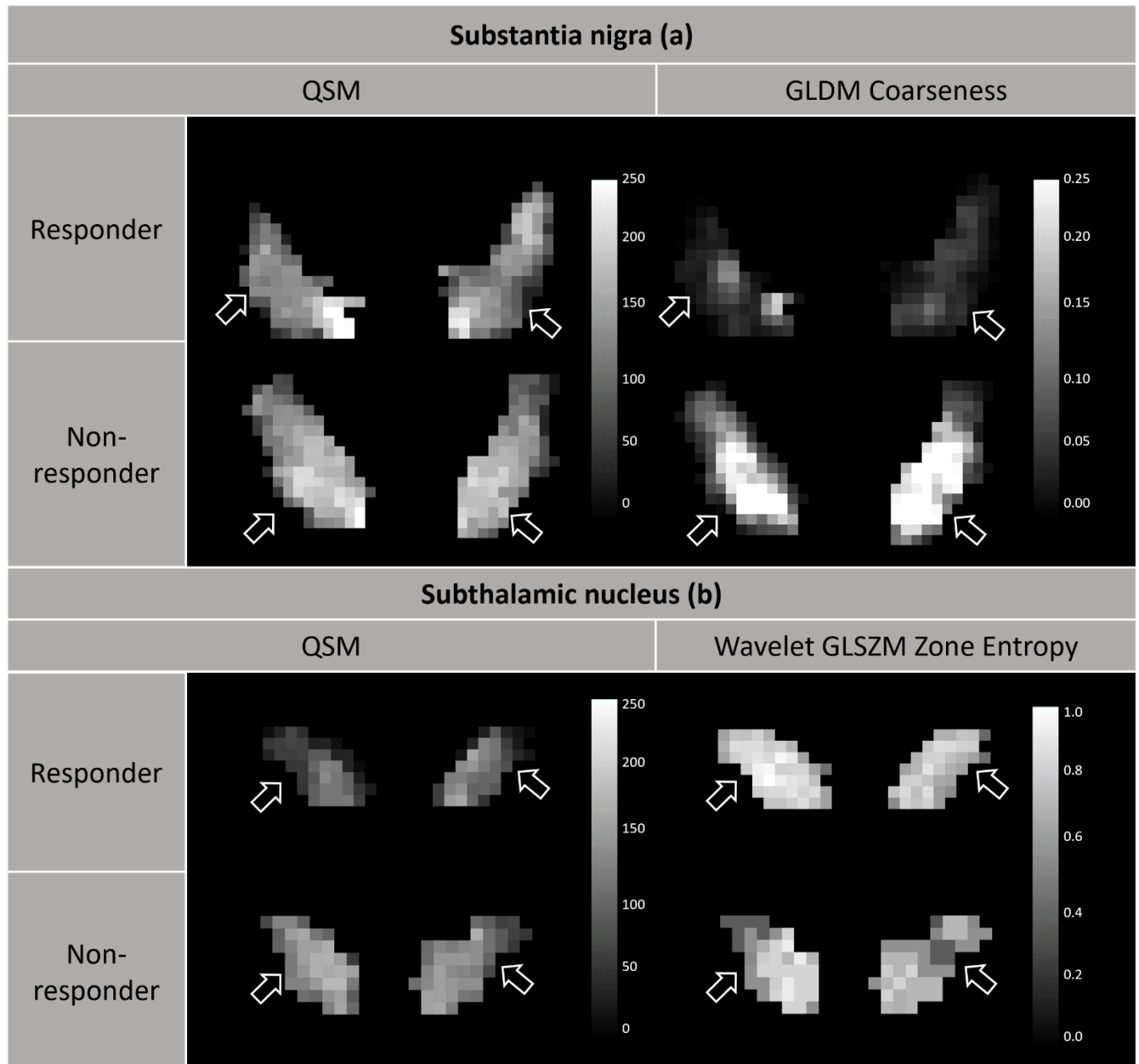


Figure 5. QSM and normalized predictive feature maps or responders (first row) and non-responders (second row) in the substantia nigra (left, a, GLDM large dependence high gray level emphasis, or coarseness) and subthalamic nucleus (right, b, wavelet GLZSM zone entropy). The increased gray level emphasis (a) and the decreased zone entropy (b) in non-responders versus responders were consistent with the ROI feature trends in the *Predictive features* section of the *Results*.



11. References

1. Lozano AM, Lipsman N. Probing and regulating dysfunctional circuits using deep brain stimulation. *Neuron*. Feb 6 2013;77(3):406-24. doi:10.1016/j.neuron.2013.01.020
2. Bronstein JM, Tagliati M, Alterman RL, et al. Deep Brain Stimulation for Parkinson Disease: An Expert Consensus and Review of Key Issues. *Archives of Neurology*. 2011;68(2):165-165. doi:10.1001/archneurol.2010.260
3. Lachenmayer L, Mürset M, Antih N, et al. Subthalamic and pallidal deep brain stimulation for Parkinson's disease-meta-analysis of outcomes. *npj Parkinson's Disease*. 09/06 2021;7doi:10.1038/s41531-021-00223-5
4. Wolke R, Becktepe JS, Paschen S, et al. The Role of Levodopa Challenge in Predicting the Outcome of Subthalamic Deep Brain Stimulation. *Movement Disorders Clinical Practice*. 2023;10(8):1181-1191. doi:<https://doi.org/10.1002/mdc3.13825>
5. Wang Y, Liu T. Quantitative susceptibility mapping (QSM): Decoding MRI data for a tissue magnetic biomarker. *Magnetic Resonance in Medicine*. 2015;73(1):82-101. doi:10.1002/mrm.25358
6. Dimov AV, Gupta A, Kopell BH, Wang Y. High-resolution QSM for functional and structural depiction of subthalamic nuclei in DBS presurgical mapping. *Journal of Neurosurgery*. 2019;131(2):360-367. doi:10.3171/2018.3.jns172145
7. Liu T, Eskreis-Winkler S, Schweitzer AD, et al. Improved subthalamic nucleus depiction with quantitative susceptibility mapping. *Radiology*. Oct 2013;269(1):216-23. doi:10.1148/radiol.13121991
8. Rasouli J, Ramdhani R, Panov FE, et al. Utilization of Quantitative Susceptibility Mapping for Direct Targeting of the Subthalamic Nucleus During Deep Brain Stimulation Surgery. *Operative neurosurgery (Hagerstown, Md)*. Apr 1 2018;14(4):412-419. doi:10.1093/ons/opx131
9. Guan X, Lancione M, Ayton S, Dusek P, Langkammer C, Zhang M. Neuroimaging of Parkinson's disease by quantitative susceptibility mapping. *NeuroImage*. 2024/04/01/2024;289:120547. doi:<https://doi.org/10.1016/j.neuroimage.2024.120547>
10. Li G, Zhai G, Zhao X, et al. 3D texture analyses within the substantia nigra of Parkinson's disease patients on quantitative susceptibility maps and R2(*) maps. *NeuroImage*. Mar 2019;188:465-472. doi:10.1016/j.neuroimage.2018.12.041
11. Kang JJ, Chen Y, Xu GD, et al. Combining quantitative susceptibility mapping to radiomics in diagnosing Parkinson's disease and assessing cognitive impairment. *Eur Radiol*. Oct 2022;32(10):6992-7003. doi:10.1007/s00330-022-08790-8
12. Roberts A, Zhang J, Tozlu C, et al. Radiomic prediction of Parkinson's disease deep brain stimulation surgery outcomes using quantitative susceptibility mapping and label noise compensation. *neuromodectJ*. 2024/10/1 2024;doi:10.31641/nmj-ogms6387
13. Zhao W, Yang C, Tong R, et al. Relationship Between Iron Distribution in Deep Gray Matter Nuclei Measured by Quantitative Susceptibility Mapping and Motor Outcome After Deep Brain Stimulation in Patients With Parkinson's Disease. *J Magn Reson Imaging*. Aug 2023;58(2):581-590. doi:10.1002/jmri.28574
14. Liu Y, Xiao B, Zhang C, et al. Predicting Motor Outcome of Subthalamic Nucleus Deep Brain Stimulation for Parkinson's Disease Using Quantitative Susceptibility Mapping and Radiomics: A Pilot Study. *Front Neurosci*. 2021;15:731109. doi:10.3389/fnins.2021.731109
15. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. *Journal of Nuclear Medicine*. 2020;61(4):488-495. doi:10.2967/jnumed.118.222893
16. Materka A. Texture analysis methodologies for magnetic resonance imaging. *Dialogues in Clinical Neuroscience*. 2004;6(2):243-250. doi:10.31887/dcons.2004.6.2/amaterka

17. Frenay B, Verleysen M. Classification in the Presence of Label Noise: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*. 2014;25(5):845-869. doi:10.1109/tnnls.2013.2292894
18. Natarajan N, Dhillon IS, Ravikumar PK, Tewari A. Learning with Noisy Labels. 2013; https://proceedings.neurips.cc/paper_files/paper/2013/file/3871bd64012152bfb53fdf04b401193f-Paper.pdf
19. Liu J, Liu T, De Rochefort L, et al. Morphology enabled dipole inversion for quantitative susceptibility mapping using structural consistency between the magnitude image and the susceptibility map. *NeuroImage*. 2012;59(3):2560-2568. doi:10.1016/j.neuroimage.2011.08.082
20. Lin Z, Zhang X, Wang L, et al. Revisiting the L-Dopa Response as a Predictor of Motor Outcomes After Deep Brain Stimulation in Parkinson's Disease. *Frontiers in Human Neuroscience*. 2021;15doi:10.3389/fnhum.2021.604433
21. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*. 2017;77(21):e104-e107. doi:10.1158/0008-5472.can-17-0339
22. Roberts AG, Romano D, Zhang J, et al. QRadAR: An Open-source Toolbox for Quantitative Magnetic Resonance Radiomics Analysis and Reproducibility. presented at: International Society of Magnetic Resonance in Medicine (Submitted); 2023; Singapore.
23. Li X, Xiong Y, Liu S, et al. Predicting the Post-therapy Severity Level (UPDRS-III) of Patients With Parkinson's Disease After Drug Therapy by Using the Dynamic Connectivity Efficiency of fMRI. *Frontiers in Neurology*. 2019;10doi:10.3389/fneur.2019.00668
24. Wu CFJ. Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*. 1986;14(4):1261-1295. doi:10.1214/aos/1176350142
25. Branco P, Torgo L, Ribeiro RP. SMOGN: a pre-processing approach for imbalanced regression. PMLR; 2017:36-50.
26. Huang W, Ogbuji R, Zhou L, Guo L, Wang Y, Kopell BH. Motoric impairment versus iron deposition gradient in the subthalamic nucleus in Parkinson's disease. *J Neurosurg*. Jul 1 2021;135(1):284-290. doi:10.3171/2020.5.JNS201163
27. Wolke R, Becktepe JS, Paschen S, et al. The Role of Levodopa Challenge in Predicting the Outcome of Subthalamic Deep Brain Stimulation. *Movement Disorders Clinical Practice*. 2023;10(8):1181-1191. doi:10.1002/mdc3.13825
28. Saranza G, Lang AE. Levodopa challenge test: indications, protocol, and guide. *Journal of Neurology*. 2020;doi:10.1007/s00415-020-09810-7
29. Albanese A, Bonuccelli U, Brefel C, et al. Consensus statement on the role of acute dopaminergic challenge in Parkinson's disease. *Movement Disorders*. 2001;16(2):197-201. doi:10.1002/mds.1069
30. Kleiner-Fisman G, Herzog J, Fisman DN, et al. Subthalamic nucleus deep brain stimulation: Summary and meta-analysis of outcomes. *Movement Disorders*. 2006;21(S14):S290-S304. doi:10.1002/mds.20962
31. Hirschmann J, Steina A, Vesper J, Florin E, Schnitzler A. Neuronal oscillations predict deep brain stimulation outcome in Parkinson's disease. *Brain Stimulation*. 2022;15(3):792-802. doi:10.1016/j.brs.2022.05.008
32. Goede LL, Oxenford S, Kroneberg D, et al. Linking Invasive and Noninvasive Brain Stimulation in Parkinson's Disease: A Randomized Trial. *Movement Disorders*. 2024;doi:10.1002/mds.29940
33. Chen Y, Zhu G, Liu Y, et al. Predict initial subthalamic nucleus stimulation outcome in Parkinson's disease with brain morphology. *CNS neuroscience & therapeutics*. May 2022;28(5):667-676. doi:10.1111/cns.13797
34. Shang R, He L, Ma X, Ma Y, Li X. Connectome-Based Model Predicts Deep Brain Stimulation Outcome in Parkinson's Disease. *Frontiers in computational neuroscience*. 2020;14:571527. doi:10.3389/fncom.2020.571527

35. Horn A, Reich M, Vorwerk J, et al. Connectivity Predicts deep brain stimulation outcome in Parkinson disease. *Ann Neurol*. Jul 2017;82(1):67-78. doi:10.1002/ana.24974
36. Wang J, Shang R, He L, et al. Prediction of Deep Brain Stimulation Outcome in Parkinson's Disease With Connectome Based on Hemispheric Asymmetry. *Front Neurosci*. 2021;15:620750. doi:10.3389/fnins.2021.620750
37. Lai Y, He N, Wei H, et al. Value of functional connectivity in outcome prediction for pallidal stimulation in Parkinson disease. *J Neurosurg*. Jan 1 2023;138(1):27-37. doi:10.3171/2022.3.JNS212732
38. Ma L, Gholam Azad M, Dharmasivam M, et al. Parkinson's disease: Alterations in iron and redox biology as a key to unlock therapeutic strategies. *Redox Biology*. 2021/05/01/2021;41:101896. doi:<https://doi.org/10.1016/j.redox.2021.101896>
39. Kosta P, Argyropoulou MI, Markoula S, Konitsiotis S. MRI evaluation of the basal ganglia size and iron content in patients with Parkinson's disease. *Journal of Neurology*. 2006;253(1):26-32. doi:10.1007/s00415-005-0914-9
40. Shahmaei V, Faeghi F, Mohammadbeigi A, Hashemi H, Ashrafi F. Evaluation of iron deposition in brain basal ganglia of patients with Parkinson's disease using quantitative susceptibility mapping. *European Journal of Radiology Open*. 2019;6:169-174. doi:10.1016/j.ejro.2019.04.005
41. Hartmann CJ, Fliegen S, Groiss SJ, Wojtecki L, Schnitzler A. An update on best practice of deep brain stimulation in Parkinson's disease. *Therapeutic Advances in Neurological Disorders*. 2019;12:175628641983809. doi:10.1177/1756286419838096
42. Fu X, Deng W, Cui X, et al. Time-Specific Pattern of Iron Deposition in Different Regions in Parkinson's Disease Measured by Quantitative Susceptibility Mapping. *Frontiers in Neurology*. 2021;12doi:10.3389/fneur.2021.631210
43. Loehrer PA, Bopp MHA, Dafsari HS, et al. Microstructure predicts non-motor outcomes following deep brain stimulation in Parkinson's disease. *npj Parkinson's Disease*. 2024;10(1)doi:10.1038/s41531-024-00717-y
44. Roberts AG, Romano D, Zhang J, et al. χ -DBS: An Open-Source Susceptibility Atlas Pipeline for Deep Brain Stimulation Target Segmentation and Improved Visualization'. presented at: International Society of Magnetic Resonance in Medicine (Submitted); 2024; Hawaii.
45. Roberts AG, Romano DJ, Şişman M, et al. Maximum spherical mean value filtering for whole-brain QSM. *Magnetic Resonance in Medicine*. 2024;91(4):1586-1597. doi:10.1002/mrm.29963