Multimodal Neural and Behavioral Data Predict Response to Rehabilitation in Chronic Poststroke Aphasia

Anne Billot, MSc; Sha Lai, MSc; Maria Varkanitsa, PhD; Emily J. Braun, MSc; Brenda Rapp, PhD; Todd B. Parrish, PhD; James Higgins, BA; Ajay S. Kurani, PhD; David Caplan, MD, PhD; Cynthia K. Thompson, PhD; Prakash Ishwar, PhD; Margaret Betke, PhD; Swathi Kiran, PhD

BACKGROUND: Poststroke recovery depends on multiple factors and varies greatly across individuals. Using machine learning models, this study investigated the independent and complementary prognostic role of different patient-related factors in predicting response to language rehabilitation after a stroke.

METHODS: Fifty-five individuals with chronic poststroke aphasia underwent a battery of standardized assessments and structural and functional magnetic resonance imaging scans, and received 12 weeks of language treatment. Support vector machine and random forest models were constructed to predict responsiveness to treatment using pretreatment behavioral, demographic, and structural and functional neuroimaging data.

RESULTS: The best prediction performance was achieved by a support vector machine model trained on aphasia severity, demographics, measures of anatomic integrity and resting-state functional connectivity (F1=0.94). This model resulted in a significantly superior prediction performance compared with support vector machine models trained on all feature sets (F1=0.82, \( P<0.001 \)) or a single feature set (F1 range=0.68–0.84, \( P<0.001 \)). Across random forest models, training on resting-state functional magnetic resonance imaging connectivity data yielded the best F1 score (F1=0.87).

CONCLUSIONS: While behavioral, multimodal neuroimaging data and demographic information carry complementary information in predicting response to rehabilitation in chronic poststroke aphasia, functional connectivity of the brain at rest after stroke is a particularly important predictor of responsiveness to treatment, both alone and combined with other patient-related factors.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.
Nonstandard Abbreviations and Acronyms

| AQ     | aphasia quotient          |
|--------|---------------------------|
| FA     | fractional anisotropy     |
| LS     | lesion size               |
| MRI    | magnetic resonance imaging|
| RF     | random forest             |
| rs-fMRI or RS | resting-state functional magnetic resonance imaging |
| SVM    | support vector machine    |

at baseline assessment\textsuperscript{10,11} and demographic information\textsuperscript{5} may also play a role in the amount of language abilities recovered over time. However, the role of demographic data in therapy outcomes did not reach a consensus in the literature.\textsuperscript{12} Furthermore, better brain structural integrity\textsuperscript{8,13,14} and functional local activity and connectivity\textsuperscript{15–18} are positively related to the degree of spontaneous and treatment-related language recovery. Importantly, most of these studies investigated the value of individual variables or combined only a few of them. Therefore, it remains unclear (1) how each of the aforementioned factors comparatively predict natural language recovery and recovery after rehabilitation and (2) whether a combination of multiple factors is superior in prediction relative to a single type of factor. This question is important because clinicians need to know which types of data are necessary to provide an accurate prognosis to patients.

Recent advances in machine learning have allowed the application of multivariate analysis methods on multimodal neuroimaging data to predict language impairments at a single time point after brain damage\textsuperscript{19–22} or, in longitudinal studies, to predict natural language recovery over time after stroke.\textsuperscript{13,16,23,24} However, these studies present several limitations: the period of recovery investigated varied across participants,\textsuperscript{13,23} the amount of rehabilitation received by each individual was not controlled,\textsuperscript{13,16,23} and only one type of imaging data was included (ie, functional or structural).\textsuperscript{13,16,23,24}

In this study, we sought to identify the independent and cumulative importance of behavioral, demographic, and multimodal structural and functional imaging data to predict treatment-related language recovery in chronic poststroke aphasia. Building on our pilot work,\textsuperscript{25} we investigate the efficacy of 2 different machine learning models, support vector machine (SVM) and random forest (RF), to predict the improvement in language ability after 12 weeks of rehabilitation. We hypothesized that model accuracy will be improved by the combination of behavioral, demographic, structural and functional imaging variables compared with single modality models.

METHODS

This study has been conducted in adherence to the TRIPOD guidelines (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis; see the checklist in the Supplemental Material). Data can be shared upon request based on a formal data sharing agreement. Figure 1 presents the methodological framework of the study.

Participants

Participants were 55 individuals (37 male) with chronic poststroke aphasia due to a single left-hemisphere stroke (mean age=58.8 years, mean time poststroke onset=59.0 months, mean education=15.8 years). For individual demographic data and aphasia severity (Table S1). Participants were selected from a larger sample (n=81) who were enrolled in a multi-site study between 2015 and 2018 examining neurobiological features of aphasia recovery (https://cnlr.northwestern.edu/). Of the full sample, 55 participants were included based on data availability of all input features of interest for this investigation (see flowchart in Figure S1). The participants included in this study were recruited at Boston University (N=30), Johns Hopkins University (N=16), and Northwestern University (N=9). Figure 2 presents the lesion distribution of all participants. Exclusion criteria included premorbid neurological disease, history of multiple left-hemisphere strokes, and contraindications for magnetic resonance imaging (MRI). All participants provided written informed consent before study participation. The study protocol was approved by the institutional review boards at Boston University, Massachusetts General Hospital, Northwestern University, and Johns Hopkins University.

Behavioral Assessment and Treatment

Participants completed a battery of standardized assessments at baseline. The Western Aphasia Battery—Revised\textsuperscript{26} was used to assess aphasia severity per the aphasia quotient (AQ). The Doors and People test\textsuperscript{27} the Wechsler Adult Intelligence Scale Digit Span,\textsuperscript{28} the Raven’s Coloured Progressive Matrices,\textsuperscript{29} the Corsi block-tapping test,\textsuperscript{30} and the Serial Reaction Time Task\textsuperscript{30} were used to assess overall cognitive function. Participants at 3 sites received different types of language treatments (see Methods in the Supplemental Material). The treatment protocols and the successful results of these treatments have been reported elsewhere.\textsuperscript{11,15,31–34} For the purposes of this study, data for these patients are collapsed as we examine responsiveness to overall language rehabilitation (and not to any treatment type in particular).

MRI Data Acquisition

MRI was completed on a Siemens 3T Skyra with a 20-channel head/neck coil at the Martinos Center in Charlestown, MA, for Boston University; on a Siemens TIM Trio with a 32-channel head coil or a Siemens Prisma with a 64-channel head/neck coil at the Center for Translational Imaging in Chicago, IL, for Northwestern University; and on a Philips Intera with a 32-channel head coil at Johns Hopkins University. Imaging protocols were harmonized across sites to ensure similar quality and timing and these protocols have been reported in previous papers.\textsuperscript{35–37} Structural imaging included a T1-weighted sagittal sequence (voxel size=1x1x1 mm\textsuperscript{3}), and a high-resolution whole-brain cardiac-gated diffusion-weighted imaging sequence (voxel...
size=1.983×1.983×2.000 mm³, 72 interleaved slices with 60 gradient directions and 10 nondiffusion weighted (b=0) volumes, b value=1500 s/mm². Whole-brain functional images were collected using a gradient-echo T2*-weighted sequence (voxel size=1.72×1.72×3 mm³). Complete imaging sequence parameters are provided in the Supplemental Material.

MRI Data Preprocessing

Several brain structural and functional measures were calculated including (1) lesion volume extracted through in-house MATLAB scripts from lesion masks manually drawn using MRICron software and normalized to MNI space, (2) the integrity of gray and white matter regions calculated by computing the percentage of spared tissue in 69 left-hemisphere gray matter regions of the Automated Anatomical Labeling atlas, and 36 white matter tracts of the BCBToolKit probabilistic atlas, respectively, (3) average fractional anisotropy (FA) in 12 tracts of interest, computed from diffusion tensor imaging data preprocessed using the Advanced Diffusion Preprocessing Pipeline and the Northwestern University Neuroimaging Data Archive, and converted to tracts using the Automated Fiber Quantification software (deterministic tractography), and (4) resting-state functional MRI (rs-fMRI) connectivity, involving the rs-fMRI data first preprocessed in fMRIPrep version 1.4.1, a Nipype-based tool, and then through the CONN toolbox to extract bivariate Fisher-transformed Pearson correlations between 50 bilateral anatomic regions of interest specified using the Automated Anatomical Labeling atlas, resulting in 625 pairwise correlations per individual.

Methodological details on the MRI preprocessing are available in the Supplemental Material.

Model Development and Evaluation

The construction and comparison of classification models predicting treatment response included several steps detailed in the Supplemental Material. First, responsiveness to rehabilitation was determined by calculating the change in accuracy percentage on the treatment probes. Participants were classified into responders (n=33) and nonresponders (n=22) based on a cutoff of 25 percentage points change in accuracy (Figure 3 and Supplemental Material). Second, given the large number of variables in the feature sets (Table), dimensionality...
reduction methods were used. The total scores of the 6 neuropsychological tests assessing overall cognitive function were entered in a varimax-rotated principal component analysis, performed in R, version 3.4.3.47 Two components (hereafter, cognitive composite scores), representing visuospatial processing and verbal working memory, were selected and explained 63% of the variance of the data (Table S2 and Supplemental Material). Additionally, some neuroimaging feature sets, namely percentage spared in white matter regions, percentage spared in gray matter regions, FA, and resting-state fMRI (RS) data, had dimensions much larger than any of the rest: 36, 69, 12, and 625, respectively. Therefore, supervised feature selection48 was performed on these variables and Pearson correlation coefficients between all selected features were computed (Figure 4 and Tables S3 and S4 and Figure S2).

SVM and RF models were trained, tuned, and tested using a leave-one-out cross-validation procedure on all possible combinations of feature sets, including individual feature sets, to predict treatment response labels. Training and testing steps, hyperparameter values and details of fine-tuning are provided in the Supplemental Material and Tables S5 and S6. Prediction performance was evaluated using 4 metrics: accuracy, F1 score, precision, and recall, capturing different types of prediction errors (Supplemental Material). The F1 score was selected as the primary metric because it is less affected by an imbalanced class distribution (as is the case in this data set) than other metrics, and allows researchers to evaluate models based on a balance between precision and recall. Finally, distributions of F1 scores were computed for each of these models using a leave-one-out approach by iteratively removing one sample and computing the F1 scores based on performance on the 54 other samples. These scores were used in 2-tailed Wilcoxon signed-rank tests49 (function wilcox.test in R version 3.4.3) to compare the prediction performance of SVM/RF models based on a single feature set with (1) the optimal SVM/RF model, that is, the SVM/RF model trained on the feature set combination that resulted in the highest average F1 score and (2) the SVM/RF model trained on all feature sets combined. Statistical comparison was considered significant at an alpha level of 0.05.

**RESULTS**

Figure 5A and 5B present all evaluation metrics for each individual-feature-set model, the all-feature-sets model and the optimal model (ie, best F1 score). Figure 5C and 5D show the distribution of F1 scores for each of these 3
model types. In addition, Tables S7 and S8 show evaluation metrics for (1) the top 20 models trained on the best combinations of feature sets (ranked by F1 score), (2) the model trained on all feature sets, and (3) individual-feature-set models, for both SVM and RF, respectively. Results of statistical comparisons between these 3 types of models are presented in Table S9.

Across all models and all evaluation metrics, responsiveness to rehabilitation at the chronic stage was best predicted by SVM models including multiple feature sets, with maximum accuracy (0.927), F1 (0.941), precision (0.914), and recall values (0.970). The optimal SVM model included aphasia severity, demographics, FA, percentage of spared tissue in gray matter regions and resting-state fMRI connectivity, and performed significantly better than any of the models trained on an individual feature set (\(P<0.001\)). Notably, among SVM models trained on an individual feature set, aphasia severity (ie, AQ) and rs-fMRI connectivity (ie, RS) independently provided the best prediction performance scores, with an accuracy of 0.782 and 0.764 and an F1 score of 0.842 and 0.812, respectively.

Compared with SVM models, the top 10 RF models resulted in lower prediction performance. Three equally optimal RF models (accuracy=0.836, F1=0.873, precision=0.816, and recall=0.939) included RS alone, LS+RS and AQ+demographics+FA+RS. These top models correctly classified 46/55 participants as responders and nonresponders. Most of the top 10 RF models (9/10) included information on functional connectivity (ie, RS), and some models included information about the structural integrity of the brain (ie, LS, percentage spared in gray matter regions, percentage of spared tissue in white matter regions, or FA), the behavioral performance (ie, AQ or cognitive composite scores), and demographics. Similar to SVM models, the combination of all feature sets (accuracy=0.709, F1=0.784, precision=0.707, and recall=0.879) did not improve the prediction performance compared with the

Figure 4. Resting-state functional connectivity features.

Red dot and lines represent functional regions and connections selected after feature selection and included in the machine learning models. Blue dots represent regions excluded from the analyses after feature selection. ACC indicates anterior cingulate cortex; AG, angular gyrus; FUS, fusiform gyrus; IFG, inferior frontal gyrus; INS, insula; IPG, inferior parietal gyrus; ITG, inferior temporal gyrus; L, left; midTP, middle temporal pole; MTG, middle temporal gyrus; orb, pars orbitalis; PCC, posterior cingulate gyrus; PCG, precentral gyrus; pre, pregenual; R, right; ROI, region of interest; SFG, superior frontal gyrus; STG, superior temporal gyrus; supTP, superior temporal pole; SMA, supplementary motor area; sub, subgenual; SMG, supramarginal; SOG, superior occipital gyrus; SPG, superior parietal gyrus; and tri, pars triangularis.
optimal models combining a subset of the feature sets (P<0.001).

Across the top combinations of feature sets ranked by F1 scores, the cumulative occurrence of each feature set shows that RS is predominant in both SVM and RF models, followed by percentage spared in gray matter regions in SVM models (Figure S3).

DISCUSSION

Previous studies from our group on a subset of these data showed that brain function data,\(^{15,36}\) brain structure information,\(^{9,14}\) and language and cognitive performance\(^{11,36}\) independently predict treatment-related outcomes. This study is the first, to our knowledge, to investigate the cumulative importance of patient-related information using a comprehensive data set including behavioral, demographic, and multimodal neuroimaging information to predict rehabilitation-induced language recovery in individuals with chronic poststroke aphasia. Our results show that models that combine a subset of multimodal neuroimaging, behavioral, and demographic data outperform models trained on a single type of information and, more importantly, all the available types of information. Three types of patient-related data were consistently important in predicting responsiveness to language treatment: functional connectivity at rest (ie, rs-fMRI), the anatomical integrity, and the aphasia severity. Previous studies that have examined predictors of natural language recovery over time also demonstrated that combining information from language tests and structural MRI\(^{50,51}\) or task-based fMRI\(^{16}\) improved prognosis accuracy. In this study, we found an additional benefit of combining both functional and structural MRI information with behavioral abilities to predict responsiveness to language rehabilitation. Interestingly, the present study also demonstrates that combining all data types may not provide the best estimates for treatment outcomes.

Figure 5. Predictive performance for support vector machine (SVM) and random forest (RF) models trained on a single feature set, all feature sets or the optimal combination of feature sets (aphasia quotient [AQ], demographics [DM], fractional anisotropy [FA], percentage spared in gray matter regions [PSg], resting-state [RS] for SVM and RS for RF).

F1 scores of models trained on a single feature set were significantly lower than the models trained on the optimal combination of feature sets (P<0.001): (A) Performance of SVM models on all 55 samples; (B) performance of RF models on all 55 samples; (C) distribution (kernel density estimation in R with automatic bandwidth selection) of SVM F1 scores computed from all 55 subsets of 54 samples each; and (D) distribution (kernel density estimation in R with automatic bandwidth selection) of RF F1 scores computed from all 55 subsets of 54 samples each.
Instead, the results show that a combination of features that provide unique and salient information are sufficient in the optimal model.

Importantly, not all relevant behavioral, demographic and multimodal neuroimaging variables were equally important for the prognosis of language recovery after rehabilitation. In addition to confirming the importance of aphasia severity in predicting treatment-related language recovery, the present study showed that the status of neural connectivity at rest, even in the chronic stage, strongly predicts response to language treatment. Indeed, resting-state connectivity data contributed to all top performing models, regardless of the algorithm used. While structural neuroimaging data, such as lesion size or percentage of spared tissue in brain regions, informed the model on the extent and location of the initial damage, rs-fMRI connectivity between several undamaged left-hemisphere and a few right-hemisphere regions (Figure 4 and Table S4) provided information on the status of brain functional (re)organization and potential for relearning. Interestingly, all pairs of functional regions of interest selected in the models consisted of at least one region of interest from the language network, and all white matter tracts retained from the diffusion imaging data set corresponded to ventral or dorsal streams involved in language processing suggesting the importance of the left hemisphere in predicting recovery. In contrast, the regions where the degree of spared tissue was highly correlated with treatment response and selected as part of the final percentage of spared tissue feature sets, were not circumscribed to the left language network. Future studies are needed to draw specific conclusions on the relative importance of individual brain regions.

This study also shows that the choice of machine learning algorithm can influence the prediction performance. On our data set, SVM models seemed to better leverage complementary information from the patient-related data than RF models, resulting in higher prediction performance. These differences, however, may be related to not only the characteristics of each algorithm but also to the small sample size. Thus, differences between models should be interpreted with caution and replication studies are needed to generalize these findings.

Importantly, this study demonstrates that language recovery after rehabilitation is multifactorial. In particular, language abilities at baseline, brain structural integrity and functional connectivity comprise unique and complementary information that can improve language treatment prognosis estimations. Machine learning models presented in this study are a first step towards a more personalized treatment approach for individuals with poststroke aphasia. By leveraging behavioral and multimodal neuroimaging data, future models trained on data from a larger sample size and different treatment types could assist clinicians with targeting the best treatment approach for individuals with poststroke aphasia. As a result of this advanced personalization, stronger evidence of the effectiveness of language treatment at the chronic stage could also be used to advocate for more insurance coverage beyond a few months after stroke.

**Limitations**

Although the sample size of this study was limited due to the availability of the multimodal MRI data, it was still representative of the heterogeneity among individuals with poststroke aphasia in terms of aphasia severity and treatment response (Table S1). Previous studies demonstrated that small sample sizes can lead to less accurate results when using machine learning models on neuroimaging data to predict behavior. Therefore, we performed feature selection on all eligible feature sets to improve learning accuracy and model stability. As part of this multi-site project, all participants received an impairment-based treatment of the same intensity, yet targeting a different language impairment at each site (ie, naming, syntax, and spelling). Importantly, the goal of this study was to investigate patient-specific factors that would predict recovery after language treatment and not to investigate treatment-related factors (eg, intensity, duration, and language target). Although differences in treatments may add noise to the results of this study, the accuracy-maximizing prediction based on site-information alone is to predict all participants as responders, irrespective of the site, as Figure 3 shows. This is so since at each site, the number of responders either equals or exceeds the number of nonresponders. Thus, site as a feature may have only a limited influence on performance.

Ideally, we should use k-fold cross-validation in all computer experiments, but we used leave-one-out cross-validation for both feature selection and model training and validation steps. Furthermore, the samples used in both steps should be different, but we used the same set of samples in both steps. Both decisions were motivated by the limited total number of samples in the data set. This approach is commonly used and is particularly appropriate for small sample sizes. Thus, testing models on a larger and independent data set will be needed in future studies to improve the generalizability of these results.

**ARTICLE INFORMATION**

Received July 27, 2021; final revision received October 27, 2021; accepted November 24, 2021.

**Affiliations**

Sargent College of Health and Rehabilitation Sciences (A.B., M.V., E.J.B., S.K.), School of Medicine (A.B.), and Department of Computer Science (S.L., P.I., M.B.), Boston University, MA. Department of Cognitive Science, Johns Hopkins University, Baltimore, MD (B.R.). Department of Radiology (T.B.P., J.H.) and Department of Neurology (A.S.K.), Feinberg School of Medicine and Department of Communication Sciences and Disorders (C.K.T.), Northwestern University, Chicago, IL. Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston (D.C.).
REFERENCES

1. Adamson J, Beswick A, Ebrahim S. Is stroke the most common cause of disability? J Stroke Cerebrovasc. 2004;13:171–177. doi: 10.1016/j.jstrokecerebrovasc.2004.06.003

2. Daniel K, Wolfe CD, Busch MA, McKevitt C. What are the social consequences of stroke for working-aged adults? A systematic review. Stroke. 2009;40:e431–e440. doi: 10.1161/STROKEAHA.108.534487

3. Kauhanen ML, Korpelainen JT, Hiltunen P, Määttä R, Mononen H, Brusin E, Sotaniemi KA. Myofacial pain, Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. Cerebrovasc Dis. 2000;10:455–461. doi: 10.1159/000106107

4. Wilson SM, Eriksson DK, Brandt TH, Schneck SM, Lucanie JM, Burchfield M. Patient-related and stroke-related factors. J Eval Clin Pract. 2015;21:352–358. doi: 10.1111/jep.12257

5. Plowman E, Hentz B, Ellis C Jr. Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. J Eval Clin Pract. 2012;18:689–694. doi: 10.1111/j.1365-2753.2011.1650x0

6. Lazar RM, Speizer AE, Festo JR, Krakauer JW, Marshall RS. Variability in language recovery after first-time stroke. J Neurol Neurosurg Psychiatry. 2008;79:530–534. doi: 10.1136/jnnp.2007.122497

7. Bengheman S, Rosso C, Arbizu C, Moulton E, Dormont D, Leger A, Pires C, Samson Y. Aphasia outcome: the intersections between initial severity, lesion size and location. J Neurol 2019;266:1303–1309. doi: 10.1007/s00415-019-09259-3

8. Meier EL, Johnson JP, Pan Y, Kiran S. The utility of lesion classification in predicting language and treatment outcomes in chronic stroke-induced aphasia. Brain Imaging Behav. 2019;13:1510–1525. doi: 10.1007/s11862-019-01183-3

9. Lambon Ralph MA, Snell C, Fillingham JK, Conroy P, Sage K. Predicting the outcome of anoma therapy for people with aphasia post-CVA: both language and cognitive status are key predictors. Neurorehabil Neural Repair. 2010;24:209–235. doi: 10.1177/1545968309358337

10. Defries J. Ethnic differences in the prevalence of aphasia: a reassessment of the literature. J Speech Lang Hear Res. 1989;32:611–630. doi: 10.1044/jslhr.L-19-0254

11. Wilson SM, Eriksson DK, Brandt TH, Schneck SM, Lucanie JM. Patient-related and stroke-related factors. J Eval Clin Pract. 2015;21:352–358. doi: 10.1111/jep.12257

12. Hope TMH, Jeffs AP, Price CJ. Predicting language outcomes after stroke: is structural disconnection a useful predictor? Neurorehabil Neural Repair. 2012;18:22–29. doi: 10.1177/1545968312451642

13. Kessels RP, van Zandvoort MJ, Postma A, de Haan EH, The Corsi Block-Tapping Task: standardization and normative data. Appl Neuropsychol. 2000;7:259–264. doi: 10.1080/09602011.2000.1086515

14. Defries J. Ethnic differences in the prevalence of aphasia: a reassessment of the literature. J Speech Lang Hear Res. 1989;32:611–630. doi: 10.1044/jslhr.L-19-0254

15. Lai D, Dubey S, Anitua D, Ray BK. Determinants of aphasia recovery: exploratory decision tree analysis. Lang Cogn Neurosci. 2020;1–8. doi: 10.1038/s41385-019-0511-0

16. Lambon Ralph MA, Snell C, Fillingham JK, Conroy P, Sage K. Predicting the outcome of anoma therapy for people with aphasia post-CVA: both language and cognitive status are key predictors. Neurorehabil Neural Repair. 2010;24:209–235. doi: 10.1177/1545968309358337

17. Swischuk DL, Weschler AD. Western Aphasia Battery-Revised. PsychCorp; 2007.

18. Baddeley AD, Hitch G. Working memory: a critical review of the evidence. J Exp Psychol Hum Percept Perform. 1974;1:259–300. doi: 10.1037/h0038330

19. Lai D, Dubey S, Anitua D, Ray BK. Determinants of aphasia recovery: exploratory decision tree analysis. Lang Cogn Neurosci. 2020;1–8. doi: 10.1038/s41385-019-0511-0

20. Lai D, Dubey S, Anitua D, Ray BK. Determinants of aphasia recovery: exploratory decision tree analysis. Lang Cogn Neurosci. 2020;1–8. doi: 10.1038/s41385-019-0511-0

21. Hope TMH, Jeffs AP, Price CJ. Predicting language outcomes after stroke: is structural disconnection a useful predictor? Neurorehabil Neural Repair. 2012;18:22–29. doi: 10.1177/1545968312451642

22. Kessels RP, van Zandvoort MJ, Postma A, de Haan EH, The Corsi Block-Tapping Task: standardization and normative data. Appl Neuropsychol. 2000;7:259–264. doi: 10.1080/09602011.2000.1086515

23. Kessels RP, van Zandvoort MJ, Postma A, Kappelle LJ, de Haan E, The Corsi Block-Tapping Task: standardization and normative data. Appl Neuropsychol. 2000;7:252–258. doi: 10.1007/S1371268240721098

24. Lahiri D, Dubey S, Ardila A, Sanyal D, Ray BK. Determinants of aphasia recovery: exploratory decision tree analysis. Lang Cogn Neurosci. 2020;1–8. doi: 10.1038/s41385-019-0511-0

25. Lai D, Dubey S, Anitua D, Ray BK. Determinants of aphasia recovery: exploratory decision tree analysis. Lang Cogn Neurosci. 2020;1–8. doi: 10.1038/s41385-019-0511-0

26. Defries J. Ethnic differences in the prevalence of aphasia: a reassessment of the literature. J Speech Lang Hear Res. 1989;32:611–630. doi: 10.1044/jslhr.L-19-0254

27. Baddeley AD, Hitch G. Working memory: a critical review of the evidence. J Exp Psychol Hum Percept Perform. 1974;1:259–300. doi: 10.1037/h0038330

28. Weschler AD. Western Aphasia Battery-Revised. PsychCorp; 2007.
37. Lukic S, Thompson CK, Barbieri E, Chiappetta B, Bonakdarpour B, Kiran S, Rapp B, Parris TB, Caplan D, Common and distinct neural substrates of sentence production and comprehension. Neuroimage. 2021;224:117374. doi: 10.1016/j.neuroimage.2020.117374

38. Billot et al Predicting Poststroke Treatment Response

39. Rorden C, Brett M. Stereotaxic display of brain lesions. Behav Neurol. 2000;12:191–201. doi: 10.1155/2000/421719

40. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI single-subject brain. Neuroimage. 2002;15:273–289. doi: 10.1006/nimg.2001.0978

41. Foulon C, Cerfani L, Kinkingnéhun S, Levy R, Rosso C, Urbanski M, Volle E, Thiebaut de Schotten M. Advanced lesion symptom mapping analyses and implementation as BCBtoolkit. Gigascienc. 2018;7:1–17. doi: 10.1053/gigascience/gsy004

42. Alpert K, Kogan A, Parrish T, Marcus D, Wang L. The Northwestern University Neuroimaging Data Archive (NUNDA). Neuroimage. 2016;124(Pt B):1191–1196. doi: 10.1016/j.neuroimage.2015.06.060

43. Yeatman JD, Dougherty RF, Mylläri NJ, Wandell BA, Feldman HM. Tract profiles of white matter properties: automating fiber-tract quantification. PLoS One. 2012;7:e49790. doi: 10.1371/journal.pone.0049790

44. Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, et al. fMRIprep: a robust preprocessing pipeline for functional MRI. Nat Methods. 2019;16:111–116. doi: 10.1038/s41592-018-0235-4

45. Whitfield-Gabrieli S, Nieto-Castanon A, Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect. 2012;2:125–141. doi: 10.1089/brain.2012.0073

46. R Core Team. R: A language and environment for statistical computing. 2018;74. doi: 10.1007/s41592-018-0235-4

47. Davatzikos C. Machine learning in neuroimaging: progress and challenges. Neuroimage. 2017;197:652–656. doi: 10.1016/j.neuroimage.2018.10.003

48. Lai S, Billot A, Varkanitsa M, Braun EJ, Rapp B, Parrish TB, Higgins J, Caplan D. One: a pipeline for clinical diffusion MRI single-subject brain. Neuroimage. 2020;214:117374. doi: 10.1016/j.neuroimage.2020.117374

49. Alpert K, Kogan A, Parrish T, Marcus D, Wang L. The Northwestern University Neuroimaging Data Archive (NUNDA). Neuroimage. 2016;124(Pt B):1191–1196. doi: 10.1016/j.neuroimage.2015.06.060

50. R Core Team. R: A language and environment for statistical computing. 2018;74. doi: 10.1007/s41592-018-0235-4

51. Tábuas-Pereira M, Beato-Coelho J, Ribeiro J, Nogueira AR, Cruz L, Delcroix N, Mazoyer B, Joliot M, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI single-subject brain. Neuroimage. 2002;15:273–289. doi: 10.1006/nimg.2001.0978

52. Foulon C, Cerfani L, Kinkingnéhun S, Levy R, Rosso C, Urbanski M, Volle E, Thiebaut de Schotten M. Advanced lesion symptom mapping analyses and implementation as BCBtoolkit. Gigascience. 2018;7:1–17. doi: 10.1053/gigascience/gsy004

53. Alpert K, Kogan A, Parrish T, Marcus D, Wang L. The Northwestern University Neuroimaging Data Archive (NUNDA). Neuroimage. 2016;124(Pt B):1191–1196. doi: 10.1016/j.neuroimage.2015.06.060

54. Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage. 2019;197:652–656. doi: 10.1016/j.neuroimage.2018.10.003

55. Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognit. 2015;48:2839–2846.

56. Zhao Y, Halai AD, Lambon Ralph MA. Evaluating the granularity and statistical structure of lesions and behaviour in post-stroke aphasia. Brain Conn. 2020;2:2:125–141. doi: 10.1089/brain.2020.0235

57. R Core Team. R: A language and environment for statistical computing. 2018;74. doi: 10.1007/s41592-018-0235-4

58. Davatzikos C. Machine learning in neuroimaging: progress and challenges. Neuroimage. 2019;197:652–656. doi: 10.1016/j.neuroimage.2018.10.003

59. Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage. 2018;16:180(Pt A):68–77. doi: 10.1016/j.neuroimage.2017.06.061

60. Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognit. 2015;48:2839–2846.

61. R Core Team. R: A language and environment for statistical computing. 2018;74. doi: 10.1007/s41592-018-0235-4

62. Davatzikos C. Machine learning in neuroimaging: progress and challenges. Neuroimage. 2019;197:652–656. doi: 10.1016/j.neuroimage.2018.10.003

63. Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage. 2018;16:180(Pt A):68–77. doi: 10.1016/j.neuroimage.2017.06.061

64. Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognit. 2015;48:2839–2846.