Prediction of treatment-related language recovery in post-stroke aphasia from neuroimaging and behavioral data

Billot A.[1], Lai S.[2], Varkanitsa M.[1], Braun E.J.[1], Rapp B.[3], Parrish T.B.[4], Kurani A.S.[4], Higgins J.[4], Caplan D.[5], Thompson C.K.[6], Ishwar P.[7], Betke M.[2], Kiran S.[1]

[1]Department of Speech, Language, & Hearing Sciences, Sargent college, Boston University [2] Computer Science Department, Boston University [3] Department of Cognitive Science, Johns Hopkins University [4] Department of Radiology, Feinberg School of Medicine, Northwestern University [5] Neuropsychology Laboratory, Department of Neurology, Massachusetts General Hospital – Harvard Medical School [6] Center for the Neurobiology of Language Recovery, Northwestern University [7] College of Engineering, Boston University

Introduction

Aphasia affects approximately 1/3 of stroke survivors and severely impacts their quality of life as well as their ability to return to work.[1,2]

Predicting recovery is essential to provide a personalized prognosis and help patients and family plan for the future.

A few studies have used machine learning models to predict language recovery after stroke using neuroimaging and behavioral data. However, these studies present different limitations: variable period of recovery across participants, no control of the amount of therapy received, and/or only one type of imaging data investigated.[3-5]

Aim of this study: Investigating the efficacy of two machine learning models to predict treatment-related language recovery using a combination of behavioral, demographic and neuroimaging data

Hypothesis: Model accuracy will be improved by the combination of multimodal neuroimaging, behavioral and demographic variables to predict treatment response group compared to models using single feature sets.

Methods

Participants

55 individuals with aphasia (18F / 37M, age = 58.8 +/- 10.6, months post stroke = 59.0 +/- 47.2) resulting from a single left-hemisphere stroke were recruited in 3 research sites (Boston, Johns Hopkins, and Northwestern Universities)

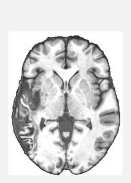
Input features: Demographics, Behavioral and Imaging data

Demographics
- Age
- Months post-stroke onset
- Education

(DM)

Behavior
- Aphasia severity: WAB-R aphasia quotient **(AQ)**
- Cognitive composite scores **(CS)**: visuo-spatial processing + verbal working memory components
  PCA using Doors and People, Corsi, Raven's matrices, SRTT, WAIS Digit span tests

Lesion information



- Lesion size **(LS)**
  Semi-automated lesion drawing on T1-weighted images
- Percentage of spared tissue in **gray** matter regions **(PSg)**
  N = 69 left-hemisphere regions
  AAL atlas, preprocessed with fmriprep
- Percentage of spared tissue in **white** matter regions **(PSw)**
  N = 36 left-hemisphere white matter regions
  BCBtoolkit, Rojkova et al. 2015

Structural connectivity

- DTI - Average Fractional Anisotropy **(FA)**
  N = 10 bilateral + 2 inter-hemispheric white matter tracts
  Computed with AFQ (Yeatman et al. 2012)

Functional connectivity

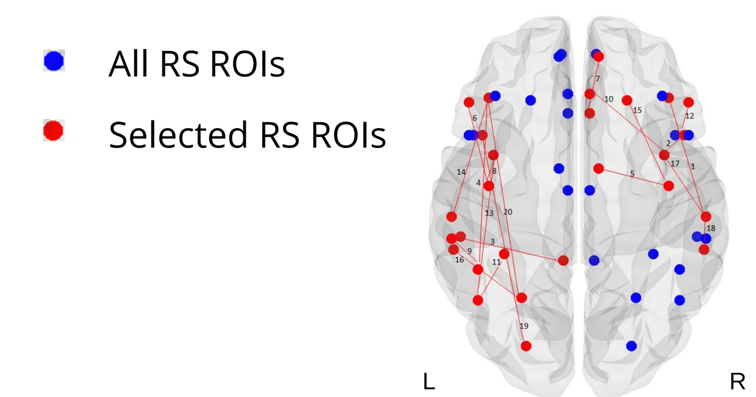
- Resting state fMRI – pairwise bivariate correlations **(RS)**
  N = 625 ROI-to-ROI correlations
  ROIs from language, default mode, salience and dorsal attention networks, AAL atlas.
  Preprocessing with fmriprep and CONN toolbox
  Whitfield-Gabrieli & Nieto-Castanon 2012)

Target: Treatment response

- **12 weeks** of site-specific treatment (BU: Semantic Feature Analysis, JHU: Spell-Study-Spell paradigm, NU : Treatment of Underlying Forms), + **site-specific probes** related to treatment at baseline and post-treatment.
- **Responsiveness to treatment** = **percent change in accuracy** (i.e. average post-treatment accuracy score minus average pre-treatment accuracy score in percentages).
- **Classification** in two groups: **responders** (percent change ≥ 0.25) and **non responders** (percent change < 0.25)

Data analysis

- **Preliminary analysis**: **RS Feature selection**. Preselection of RS features by (i) measuring Pearson Correlation Coefficient between pairwise bivariate correlations and binary treatment response labels, (ii) ranking the features based on the correlation values and (iii) running a set of independent cross-validation experiments to find the best number k of top RS features after which the improvement in prediction performance was likely to be small or negative.



List of 20 RS connections included in the analyses

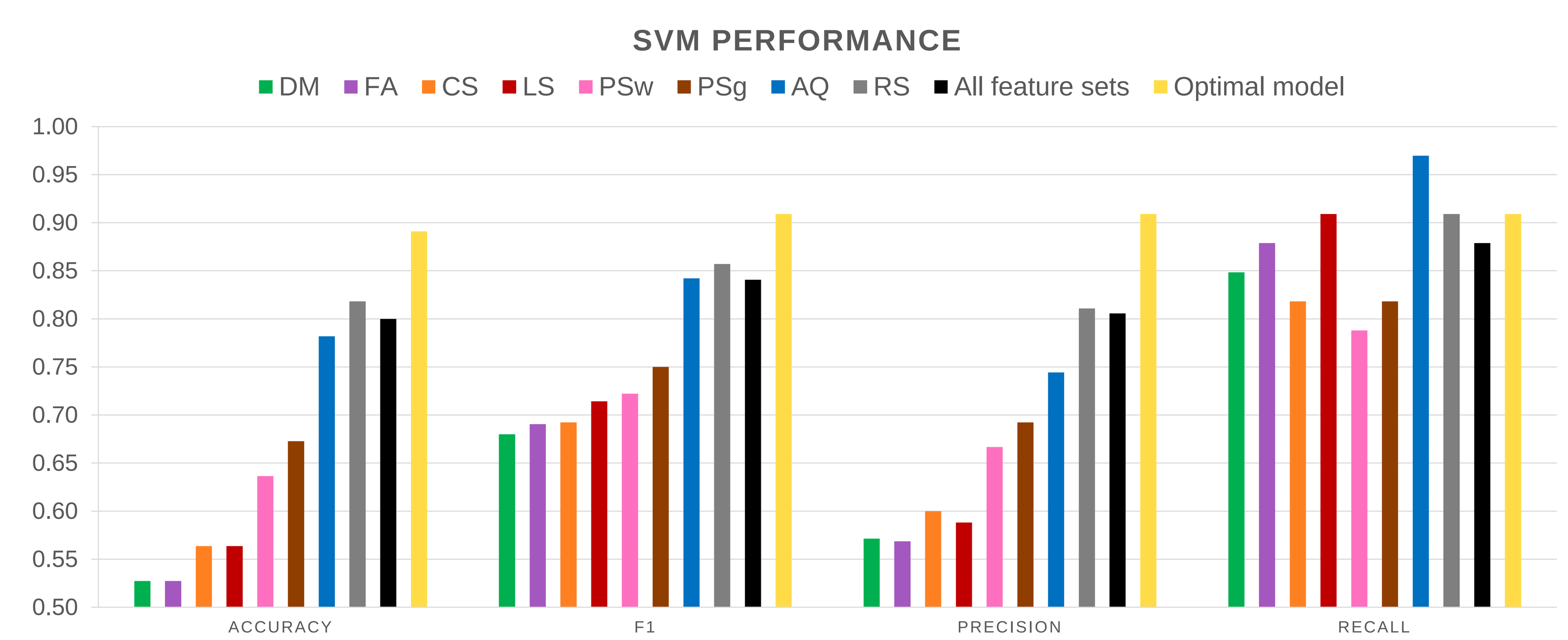
1.	R IFGorb - R STG	11.	L FUS - L AG
2.	R INS - R supTP	12.	R IFGtri - R supTP
3.	L PCC - L ITG	13.	L INS - L AG
4.	L IFGorb - L IPG	14.	L IFGorb - L STG
5.	R PCG - R SMA	15.	R PCG - R SFG
6.	L PCG - L IFGtri	16.	L SPG - L MTG
7.	R SFGmedial - R ACCsub	17.	R INS - R STG
8.	L PCG - L INS	18.	R STG - R MTG
9.	L IPG - L SMG	19.	L SOG - L midTP
10.	R INS - R ACCpre	20.	L IFGorb - L SPG

- **Training and validation:** Random Forest (RF) and Support Vector Machine (SVM) were used to classify participants into responders and non responders. All feature sets combinations were tested. For each feature sets combination, leave-one-out round-robin was used to train and test the model. Hyper-parameters were tuned on the training set using leave-one-out cross-validation.
- **Model performance metrics:** Accuracy, F1 (harmony mean between precision and recall), precision (positive predictive value) and recall (sensitivity)

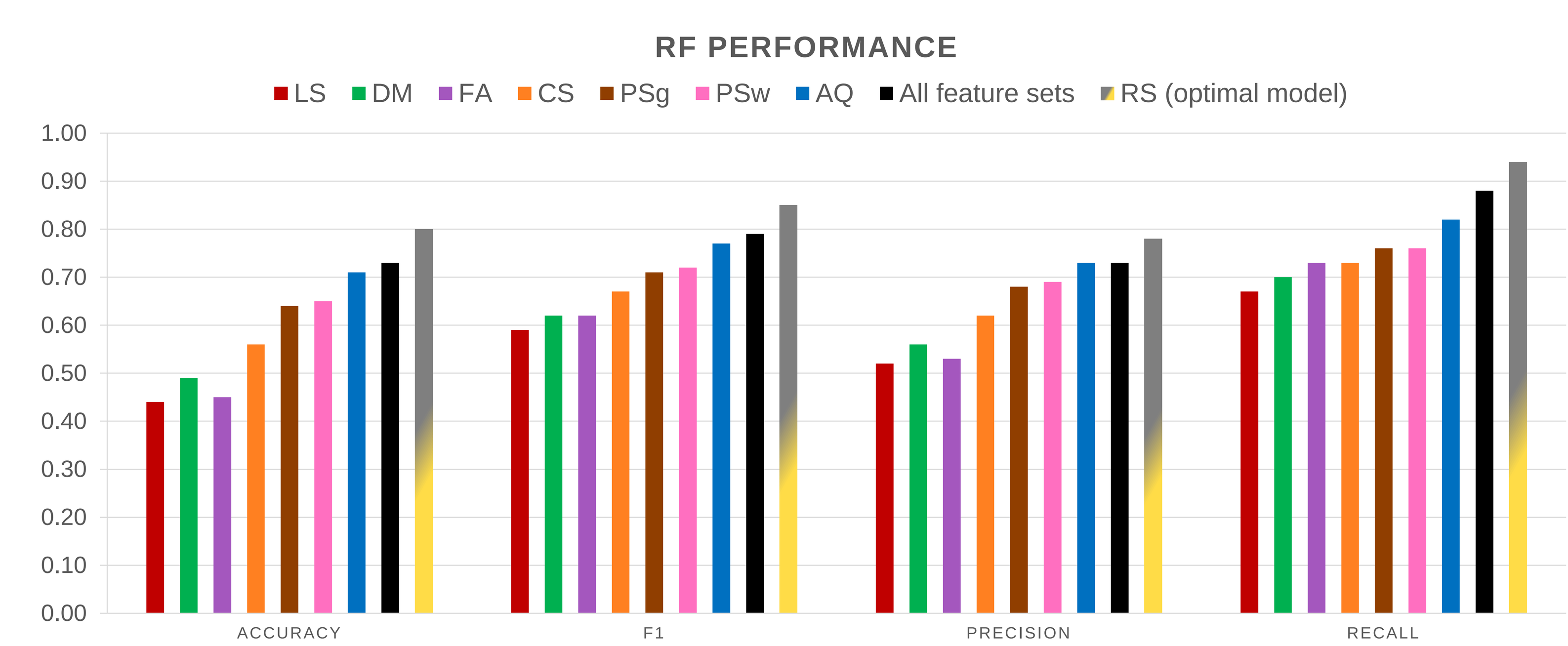
Results

RF and SVM models performance

Comparison of models including a single feature set, all feature sets and the optimal model that resulted in the best F1 score.

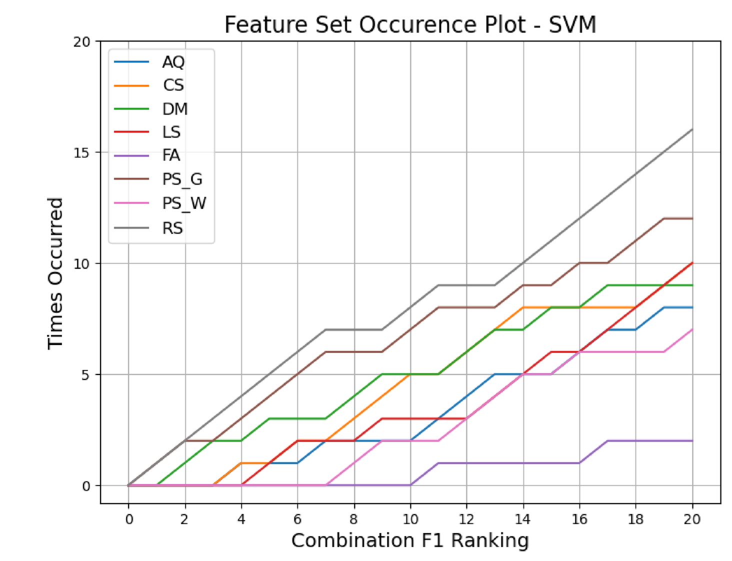


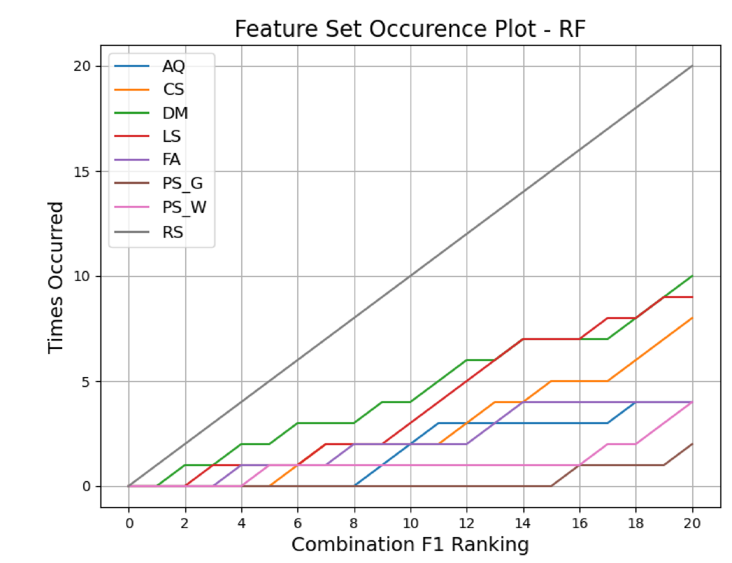
➔ Optimal model = Percent spared in gray matter regions + Resting State feature sets
➔ Highest F1 score: 0.91



➔ Optimal model = Resting-state (RS) feature set
➔ Highest F1 score: 0.85

Feature sets occurrence in the top best 20 models (SVM & RF)





➔ Resting State feature set occurs consistently across the combinations that result in the best F1 scores. Percent spared in gray matter regions feature set seems to contribute to most of the best SVM models but not to the best RF models.

Conclusion

- Random Forest and Support Vector Machine models can **predict with high accuracy** if an individual with chronic aphasia may show some **improvement after language treatment** or not.
- Across models, **resting-state fMRI data** is a **strong predictor** of responsiveness to treatment in chronic aphasia.
- Resting-state fMRI data is the only single feature set that outperforms the model including all feature sets.
- The combination of multimodal neuroimaging, behavioral and demographic data may not be necessary to achieve high prediction performance.
- Future analyses are needed to test these machine learning models on independent datasets.

Acknowledgements

We thank all individuals with aphasia who participated in this study and their families. This study was funded by the NIH/NIDCD Clinical Research Center Grant, P50DC012283, the Hariri Institute Artificial Intelligence Research (AIR) Initiative and the Institute for Health System Innovation Policy (IHSIP) at Boston University.

References

[1] Lam, J. M., & Wodchis, W. P. (2010). The relationship of 60 disease diagnoses and 15 conditions to preference-based health-related quality of life in Ontario hospital-based long-term care residents. *Medical care*, 380-387.

[2] Daniel, K., Wolfe, C. D., Busch, M. A., & McKevitt, C. (2009). What are the social consequences of stroke for working-aged adults? A systematic review. *Stroke*, 40(6), e431-e440.

[3] Saur, D., Ronneberger, O., Kümmerer, D., Mader, I., Weiller, C., & Klöppel, S. (2010). Early functional magnetic resonance imaging activations predict language outcome after stroke. *Brain*, 133(4), 1252-1264.

[4] Hope, T. M., Seghier, M. L., Leff, A. P., & Price, C. J. (2013). Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage: clinical*, 2, 424-433.

[5] Aguilar, O. M., Kerry, S. J., Ong, Y. H., Callaghan, M. F., Crinion, J., Woodhead, Z. V. J., ... & Hope, T. M. (2018). Lesion-site-dependent responses to therapy after aphasic stroke. *Journal of Neurology, Neurosurgery & Psychiatry*, 89(12), 1352-1354.

Corresponding author: Anne Billot
abillot@bu.edu