Predicting Aphasia Severity and Type Post-Stroke Using Structural MRI and Machine Learning

Olga Buivolova[1,2] (obuivolova@hse.ru), Yuan Wei[1](yyaun@hse.ru), Victor Karpychev[3] (karpuchevvictor@gmail.com), Roman Cheremin[2] (drcher2008@gmail.com), Natalia Ermakova[1] (nermakova@hse.ru), Georgii Gorshkov[1] (ggorshkov@hse.ru), Jana Kokoreva[1] (ykokoreva@hse.ru), Alexey Petrushevskii[2] (shevsky@mail.ru), Ksenia Radusheva[1] (kvkhlebnikova@edu.hse.ru), Ilia Smirnov[1] (Smirnov.i.d148825@mail.ru), Olga Soloukhina[1] (OSoloukhina@gmail.com), Sergey Subbotin[4] (SubbotinSA@pirogov-center.ru), Dmitry I. Ignatov[5] (dignatov@hse.ru), Olga Dragoy[1,6] (odragoy@hse.ru)

1 Center for Language and Brain, HSE University, Moscow, Russia
2 Center for Speech Pathology and Neurorehabilitation, Moscow, Russia
3 Department of Neurology, School of Medicine, Emory University, Atlanta, Georgia, USA
4 Digital Development Lab, National Medical and Surgical Center named after N.I. Pirogov, Moscow, Russia
5 Laboratory for Models and Methods of Computational Pragmatics, HSE University, Moscow, Russia
6 Institute of Linguistics, Russian Academy of Sciences, Moscow, Russia

Introduction

Aphasia (a language disorder) affects approximately one-third of stroke survivors (Flowers et al., 2016), arising from disrupted language networks and producing highly heterogeneous impairment patterns due to interactions among neuroanatomical, clinical, and demographic variables (Plowman et al., 2012). Machine learning (ML) enables multivariate models that integrate diverse factors to predict post-stroke language outcomes (Yourganov et al., 2016). While ML shows promise for predicting aphasia severity, classification of distinct aphasia syndromes remains underexplored – only one prior study has addressed this using the Western classification (Yourganov et al., 2015). Luria's neuropsychologically grounded taxonomy (for review, see Akhutina, 2016), which offers a principled bridge between focal structural damage and qualitatively different language breakdown patterns, has not been previously examined in ML research. Furthermore, the added value of multimodal MRI over structural MRI (sMRI) alone is debated: multimodal approaches face practical barriers including acquisition artifacts and limited clinical standardization (Tournier et al., 2011). At the same time, sMRI is a clinical gold standard. This study had two objectives:
(1) to develop and validate a robust prognostic model capable of predicting overall severity and syndrome types according to Luria's classification, and (2) to identify the specific neural correlates most critical to these predictions using only demographic and sMRI data.

Methods

We had two cohorts of participants. The primary cohort comprised 304 native Russian-speaking individuals with chronic post-stroke speech and language disorders (114 female, 190 male; mean age 55.1±10.7 years) ≥3 months post unilateral left-hemisphere stroke, with an independent validation cohort of 41 participants (10 female, 31 male; mean age 52.3±12.4 years) meeting the same inclusion criteria. Aphasia severity (6-point ordinal scale from mild to very severe) and disorder type (7 categories) were assessed using the Assessment of Speech in Aphasia within Luria's neuropsychological framework. Structural MRI lesions were manually delineated, normalized to MNI space, and overlaid with gray (116 regions) and white matter (55 tracts) atlases; feature reduction retained regions damaged in ≥5% of participants, yielding 42 gray- and 36 white-matter features plus total lesion volume and three demographic variables. Through systematic evaluation of multiple ML algorithms and the introduction of a novel Robust Generalization Index, we identified an optimal, generalizable model (TabPFN).

Results

Our findings reveal a clear dissociation between the prediction of severity and syndrome type. While aphasia severity was robustly predicted, particularly when consolidated into mild versus non-mild categories (accuracy 85.4%, sensitivity 93.1%, AUC=0.90) with performance linked interpretably to chronicity and gray- and white-matter integrity, the classification of discrete aphasia syndromes proved more challenging, achieving only moderate accuracy (validation accuracy 56%). SHAP analysis provided computational validation of Luria's framework: efferent motor aphasia was predicted by Broca's area damage with preserved posterior language structures; sensory aphasia by superior temporal gyrus and arcuate fasciculus damage; afferent motor aphasia by postcentral gyrus damage with intact superior temporal gyrus. Age showed class-specific predictive directions, and 59% of misclassified participants had mixed diagnoses, with the algorithm correctly identifying secondary diagnoses in 15% of cases.

Discussion

This work delivers the validated computational tool for prognosis within Luria's clinically and anatomically relevant classification system, addressing a critical gap in post-stroke language rehabilitation. By integrating sMRI data with a rigorous ML framework, we demonstrate that meaningful predictions of both aphasia severity and type can be derived from structural MRI alone – despite the inherent complexity and heterogeneity of post-stroke speech and language disorders. Furthermore, the model provides novel, data-driven insights into brain structure-function relationships, identifying specific gray and white matter correlates that distinguish among Luria's aphasia types and severity levels.

References

Akhutina, T. (2016). Luria’s classification of aphasias and its theoretical basis. *Aphasiology*, *30*(8), 878-897.

Flowers, H. L., Skoretz, S. A., Silver, F. L., Rochon, E., Fang, J., Flamand-Roze, C., & Martino, R. (2016). Poststroke aphasia frequency, recovery, and outcomes: a systematic review and meta-analysis. Archives of physical medicine and rehabilitation, 97(12), 2188-2201.

Plowman, E., Hentz, B., & Ellis, C. (2012). Post-stroke aphasia prognosis: A review of patient-related and stroke-related factors. *Journal of evaluation in clinical practice*, *18*(3), 689-694.

Tournier, J. D., Mori, S., & Leemans, A. (2011). Diffusion tensor imaging and beyond. *Magnetic resonance in medicine*, *65*(6), 1532.

Yourganov, G., Smith, K. G., Fridriksson, J., & Rorden, C. (2015). Predicting aphasia type from brain damage measured with structural MRI. *Cortex*, *73*, 203-215.

Yourganov, G., Fridriksson, J., Rorden, C., Gleichgerrcht, E., & Bonilha, L. (2016). Multivariate connectome-based symptom mapping in post-stroke patients: networks supporting language and speech. *Journal of Neuroscience*, *36*(25), 6668-6679.