



The white matter connectome as an individualized biomarker of language impairment in temporal lobe epilepsy

Erik Kaestner^a, Akshara R. Balachandra^a, Naeim Bahrami^a, Anny Reyes^{a,b}, Sanam J. Lalani^c, Anna Christina Macari^a, Natalie L. Voets^d, Daniel L. Drane^{e,f}, Brianna M. Paul^c, Leonardo Bonilha^g, Carrie R. McDonald^{a,b,h,*}

^a Center for Multimodal Imaging and Genetics, University of California, San Diego, CA, USA

^b San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, USA

^c Department of Neurology, University of California – San Francisco, San Francisco, CA, USA

^d Wellcome Centre for Integrative Neuroimaging, FMRIB Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

^e Departments of Neurology and Pediatrics, Emory University School of Medicine, Atlanta, GA, USA

^f Department of Neurology, University of Washington, Seattle, WA, USA

^g Medical University of South Carolina, Department of Neurology, USA

^h Department of Psychiatry, University of California, San Diego, CA, USA

ABSTRACT

Objective: The distributed white matter network underlying language leads to difficulties in extracting clinically meaningful summaries of neural alterations leading to language impairment. Here we determine the predictive ability of the structural connectome (SC), compared with global measures of white matter tract microstructure and clinical data, to discriminate language impaired patients with temporal lobe epilepsy (TLE) from TLE patients without language impairment. **Methods:** T1- and diffusion-MRI, clinical variables (CVs), and neuropsychological measures of naming and verbal fluency were available for 82 TLE patients. Prediction of language impairment was performed using a robust tree-based classifier (XGBoost) for three models: (1) a CV-model which included demographic and epilepsy-related clinical features, (2) an atlas-based tract-model, including four frontotemporal white matter association tracts implicated in language (i.e., the bilateral arcuate fasciculus, inferior frontal occipital fasciculus, inferior longitudinal fasciculus, and uncinate fasciculus), and (3) a SC-model based on diffusion MRI. For the association tracts, mean fractional anisotropy was calculated as a measure of white matter microstructure for each tract using a diffusion tensor atlas (i.e., AtlasTrack). The SC-model used measurement of cortical-cortical connections arising from a temporal lobe subnetwork derived using probabilistic tractography. Dimensionality reduction of the SC was performed with principal components analysis (PCA). Each model was trained on 49 patients from one epilepsy center and tested on 33 patients from a different center (i.e., an independent dataset). Randomization was performed to test the stability of the results.

Results: The SC-model yielded a greater area under the curve (AUC; .73) and accuracy (79%) compared to both the tract-model (AUC: .54, $p < .001$; accuracy: 70%, $p < .001$) and the CV-model (AUC: .59, $p < .001$; accuracy: 64%, $p < .001$). Within the SC-model, lateral temporal connections had the highest importance to model performance, including connections similar to language association tracts such as links between the superior temporal gyrus to pars opercularis. However, in addition to these connections many additional connections that were widely distributed, bilateral and interhemispheric in nature were identified as contributing to SC-model performance.

Conclusion: The SC revealed a white matter network contributing to language impairment that was widely distributed, bilateral, and lateral temporal in nature. The distributed network underlying language may be why the SC-model has an advantage in identifying sub-components of the complex fiber networks most relevant for aspects of language performance.

1. Introduction

Language impairment is observed in up to ~50% of patients with temporal lobe epilepsy (TLE) (Balter et al., 2016; Reyes et al., 2019) and is typically characterized by deficits in naming and verbal fluency (Allone et al., 2017). Although TLE is an inherently heterogeneous disorder characterized by an array of clinical and cognitive symptoms (Bell et al., 2011), language impairments are most frequently studied in

patients with left TLE (LTLE) (Busch et al., 2005; Raspall et al., 2005; Keary et al., 2007). However, there is now increasing awareness that language impairments are also common in right TLE (RTLE) (Hermann et al., 1997; Alessio et al., 2006; Bell et al., 2001). An early age of seizure onset is associated with pre-surgical language impairment (Lee et al., 2013; Oyegbile et al., 2004), but language function is also affected by other clinical features such as education, handedness, the presence of mesial temporal sclerosis (MTS), anti-epileptic drugs

* Corresponding author at: Altman CTRI building; Floor #: 4W, 9452 Medical Center Drive, La Jolla, CA 92037, USA.

E-mail address: camcdonald@ucsd.edu (C.R. McDonald).

<https://doi.org/10.1016/j.nicl.2019.102125>

Received 19 August 2019; Received in revised form 10 December 2019; Accepted 13 December 2019

Available online 13 December 2019

2213-1582/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

(AEDs), and seizure frequency (Stewart et al., 2014; Sass et al., 1992). Amid this clinical variability, work on cognitive phenotypes (Reyes et al., 2019; Hermann et al., 2007; Dabbs et al., 2009; Rodríguez-Cruces et al., 2018) seeks to understand which commonalities drive shared cognitive impairments across diverse patients with TLE. The concept of cognitive phenotypes is based on the premise that underlying neuro-biological similarities explain similar cognitive performance across patients with distinct clinical profiles.

For language, the underlying neurobiological substrate likely involves a distributed network of fronto-temporal-parietal regions (Catani et al., 2005; Price, 2012; Poeppel et al., 2012). These interconnections rely on the integrity of perisylvian and extra-sylvian white matter tracts, the disruption of which is a key contributor to language impairment (Allone et al., 2017; Leyden et al., 2015). Integrity of this neuroanatomical network has been most thoroughly studied in the context of large white-matter bundles including the arcuate fasciculus (ARC) (Upadhyay et al., 2008), inferior frontal occipital fasciculus (IFOF) (Caverzasi et al., 2014), inferior longitudinal fasciculus (ILF) (Ashtari, 2012), and uncinate fasciculus (UNC) (Hasan et al., 2009), which are known to be affected in TLE (Leyden et al., 2015). However, these large fiber bundles may not capture the entire language network and further, they require generalized assumptions about neuronal architecture across patients.

A comprehensive and individualized white matter approach involving a map of brain network connectivity is the *structural connectome* (SC). The SC is a measure of region-to-region connection strengths derived from each individual patient. This approach may enable a more global and nuanced mapping of white matter networks compared to a summary measure of microstructure derived from long-range white matter bundles, an approach commonly used (Beaulieu, 2014). Recent work on SCs in TLE (Bernhardt et al., 2015) and in other syndromes (Shen et al., 2017) emphasizes the importance of whole-brain connectomes for understanding cognitive co-morbidities in different clinical syndromes and has linked subtle network alterations to language impairments (Sporns et al., 2005; Kim et al., 2014). In TLE, SCs have shown promise in predicting seizure outcomes (Bonilha et al., 2015; Taylor et al., 2018; Gleichgerrcht et al., 2018) as well as for understanding the relationship between network architecture and naming impairment in patients with LTLE (Munsell et al., 2019). Specifically, Munsell et al. identified a distributed, bilateral white matter network of regions contributing to naming performance, even in a cohort of left TLE patients who were all left-dominant for language. These data highlight the distributed nature of the language network and the possible utility of a SC framework for probing this network.

In this study, we evaluate the performance of the SC, white matter association tracts, and clinical variables for correctly classifying TLE patients as language impaired versus non-impaired. We extend previous research by including a more heterogeneous population of TLE patients and a more comprehensive evaluation of language performance. Furthermore, we test the robustness of our model on an independent dataset and compare SC performance to that of conventional models (i.e., clinical variables and global tract-based measures). We hypothesized that the SC would lead to better classification performance, providing an improved understanding of the topology of neuronal networks associated with language performance in TLE.

2. Methods

2.1. Subjects

This study was approved by the Institutional Review Boards at the University of California, San Diego (UCSD) and University of California, San Francisco (UCSF). All participants provided informed consent according to the Declaration of Helsinki. Patients were recruited through referral from the UCSD or UCSF Epilepsy Centers and were all undergoing pre-surgical evaluations. Inclusion criteria for patients included

(1) a TLE diagnosis, (2) age 18 or older, and (3) no dual pathology or mass lesion (i.e., tumors, vascular malformations, focal cortical dysplasia, or other visible lesions on MRI) which could distort white matter anatomy on MRI. Eighty-two patients with medically refractory TLE met inclusion criteria (N = 49 from UCSD; N = 33 from UCSF). TLE diagnosis and side of seizure onset were determined by a board-certified neurologist with expertise in epileptology, in accordance with the criteria defined by the International League Against Epilepsy (Kwan et al., 2010), based on scalp and/or intracranial video-EEG telemetry, seizure semiology, and neuroimaging evaluation. The presence of MTS was determined by inspection of MRI images by a board-certified neuroradiologist with expertise in epilepsy. MRI findings suggested the presence of ipsilateral MTS in 39 patients. Hemispheric language dominance was determined by functional MRI (fMRI), intracarotid amobarbital procedure (IAP), or magnetoencephalography (MEG) and was available for 77% of the sample. Out of the remaining 19 patients with no language laterality information, 5 patients had right TLE and underwent a right ATL or right laser ablation and did not receive a WADA. The remaining patients (n = 13) who have not had language lateralization testing have not had surgery or completed the full surgical workup. For classification of language impairment, neuropsychological data were collected from 61 healthy control subjects for the study that were sex- and age-matched to the patient population. Healthy controls were excluded if they self-reported any history of neurological or psychiatric conditions.

2.2. Neuropsychological testing

Tests of language ability (i.e., auditory naming, visual naming and semantic fluency) were obtained as part of a comprehensive neuropsychological evaluation. Language tests included the Boston Naming Test (BNT) (Kaplan et al., 2001), Auditory Naming Test (ANT) (Hamberger and Seidel, 2003), and Category Fluency (CF) subtest of the Delis-Kaplan Executive Function System (DKEFS) (Delis et al., 2001). Naming and semantic fluency were evaluated because they are the most commonly impaired aspects of language in TLE (Hermann et al., 1999; Martin et al., 1990). Conversely, language comprehension and reading are not frequently impaired in TLE (Drane and Pedersen, 2019). For each test, patients' raw scores were converted into z-scores based on the distribution of healthy controls. Patients were classified as "language impaired" if they had at least two out of the three tests with a z-score of -1.5 or lower or at least one test with a z-score of -2 and the remaining tests with a z-score of -1. A similar approach has been used to define impairment in mild cognitive impairment (Jak et al., 2009), Alzheimer's (Albert et al., 2011), and Parkinson's (Litvan et al., 2012). Requiring more than one test in the same cognitive domain (i.e., language) protects against over-classification of impairment due to empirical evidence that abnormal performance on a single test is often observed in healthy individuals due to intra-individual variability (Binder et al., 2009). This approach yields more stable diagnoses over time and it approximates the clinical-decision making process, as clinicians typically examine multiple scores within a domain (Jak et al., 2009). Finally, this methodology has proven useful in previous TLE studies of cognitive impairment (Reyes et al., 2019; Kaestner et al., 2019).

In total, ~60% of TLE patients were classified as language impaired (Language Impaired; TLE-LI); All other patients were classified as "not impaired" (Non-Language Impaired; TLE-NLI). Naming impairment was common across all patients in our TLE-LI group, with every patient impaired on at least one of the naming tests (BNT: 86% impaired in TLE-LI, ANT: 88%). CF deficits were observed in 45% of patients. In the TLE-NLI group, 24% were impaired on the BNT only, 12% on the ANT only, and 9% of CF only. Thus, impairments were not considered pervasive enough to meet our "impairment criteria," with a greater concern for false positives (Edmonds et al., 2015; Edmonds et al., 2016; Bondi et al., 2014).

An estimate of nonverbal IQ (WASI Perceptual Reasoning Index)

(Wechsler, 1999) was obtained to evaluate whether the groups differed in nonverbal cognitive ability. Self-reported symptoms of depression and anxiety were also obtained with the Beck Depression Inventory-II (BDI-II) (Beck et al., 1996) and Beck Anxiety Inventory (BAI) (Beck et al., 1988), respectively, to determine whether mood state differed between the groups. For both the BDI-II and the BAI, higher scores represent greater depressive and anxiety symptoms, respectively.

2.3. Image acquisition

Brain imaging for all patients was performed on a General Electric Discovery MR750 3T scanner with an 8-channel phased-array head coil at the Center for Functional MRI at UCSD or the Surbeck Laboratory for Advanced Imaging at UCSF. Image acquisitions were identical at both centers and included a conventional three-plane localizer, GE calibration scan, a T1-weighted 3D structural scan (TR = 8.08 ms, TE = 3.16 ms, TI = 600 ms, flip angle = 8°, FOV = 256 mm, matrix = 256 × 192, slice thickness = 1 mm isotropic), and a single-shot pulsed-field gradient spin-echo EPI sequence (TE/TR = 96 ms/17 s; FOV = 24 cm, matrix = 128 × 128 × 48; axial). Diffusion-weighted images (DWIs) were acquired with $b = 0$ and $b = 1000 \text{ mm}^2/\text{s}$ with 30 diffusion gradient directions. Two additional $b = 0$ volumes were acquired with either forward or reverse phase-encode polarity for use in nonlinear B_0 correction.

2.4. Image processing

2.4.1. Structural MRI processing

Images were corrected for spatial sensitivity inhomogeneities and for non-linear warping caused by non-uniform fields created by the gradient coils (Jovicich et al., 2006). The cortical surface was reconstructed and parcellated using *FreeSurfer*, 5.3.0 (Dale et al., 1999). Visual inspection was performed on all images to identify topological defects, which were subsequently edited using established software guidelines.

2.4.2. DTI processing

Preprocessing of the diffusion data included corrections for distortions due to magnetic susceptibility (B_0), eddy currents, and gradient nonlinearities, head motion correction and registration to the T1-weighted structural image. For B_0 distortion correction, a reverse gradient method was used (Holland et al., 2010). A detailed description of the image processing is provided elsewhere (McDonald et al., 2014). DTI-derived fractional anisotropy (FA) was calculated based on a tensor fit to the $b = 1000$ data.

2.4.3. Fiber tract calculations

Fiber tract values were derived using a probabilistic diffusion tensor atlas (i.e., AtlasTrack). For tract illustration see Fig. 1A. Because patients with anatomically deforming lesions were excluded from this study, an atlas approach was justified. AtlasTrack is a fully automated method for labeling fiber tracts in individual subjects based on diffusion-weighted images, T1-weighted images, and a probabilistic atlas of fiber tract locations and orientations. An important feature of AtlasTrack is the use of fiber orientation information from the diffusion images to refine tract probability estimates. This individualizes the fiber tract ROIs for each subject and minimizes the contribution from regions that exhibit diffusion orientations inconsistent with the consensus fiber orientation information contained in the atlas. AtlasTrack has been validated in both healthy controls and patients with TLE, and has been shown to be sensitive to microstructural changes in TLE in previous studies (Hagler et al., 2009). For each subject, the T1-weighted structural images were non-linearly registered to a common space and the respective diffusion tensor orientation estimates were compared to the atlas. This resulted in a map of the relative probability that a voxel belongs to a particular tract given its location and similarity of diffusion orientation. Voxels identified with

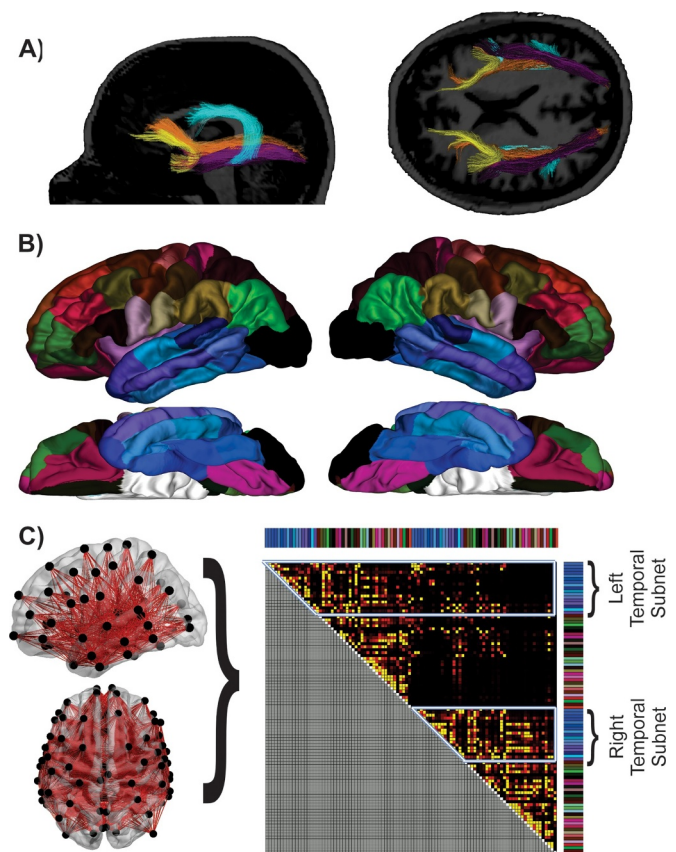


Fig. 1. Neuroanatomical measures of white matter. (A) Illustration on an average brain of the 4 association tracts used: (1) arcuate fasciculus (blue), (2) inferior frontal occipital fasciculus (orange), (3) inferior longitudinal fasciculus (purple), and (4) uncinate fasciculus (yellow). (B) An illustration on an average brain of the ROIs which are interconnected to form the structural connectome. Note that each connection must include at least one ROI in the temporal lobe. (C) Average brain displaying the region-region connections used in this study. On the right is an example connectivity matrix with the temporal lobe used in this study highlighted.

FreeSurfer 5.3.0 as cerebrospinal fluid or gray matter were excluded from the fiber regions of interest (ROIs). Average FA was calculated for each fiber ROI and weighted by fiber probability, so that voxels with low probability of belonging to a given fiber contributed minimally to average values. FA was chosen due to the highly anisotropic nature of the long-range association tracts selected for this study and our previous work on cognitive networks in TLE (Kaestner et al., 2019; Reyes et al., 2019). As a post-hoc investigation, we tested whether adding MD measures to FA significantly changed model performance and found no significant differences (see Supplementary Table 1). A full description of the atlas and detailed steps used to create the atlas are provided in Hagler et al. (Hagler et al., 2009). Specific tracts included in the current analyses are the right and left ARC, IFOF, UNC, and ILF (Fig. 1A).

2.4.4. Structural connectome generation

The FMRIB Diffusion Toolbox (FDT), which is part of the FMRIB Software Library (FSL), was used for local diffusion modeling and performing the connectome-based tractography (Behrens et al., 2003; Behrens et al., 2007). This method differs from the output of AtlasTrack because it theoretically captures more individualized patterns of local fiber tract connections that may not be included in the white matter atlas (i.e., those that are not part of pre-defined large white matter bundles). Probabilistic tractography was performed to calculate connection strength values between cortical regions corresponding to those from a modified version of *FreeSurfer*'s Desikan-Killiany (DK) atlas

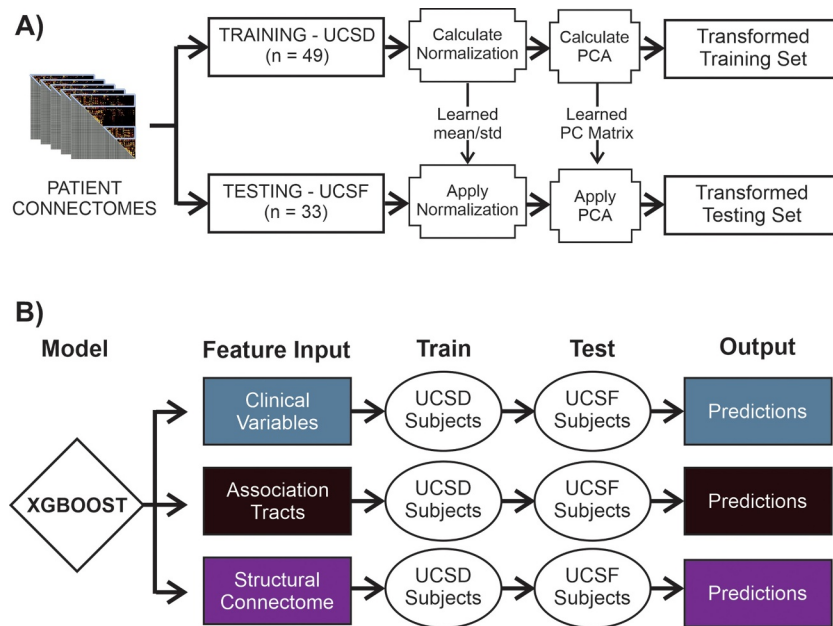


Fig. 2. Diagram of the models used in this study. (A) The connectomes were split into a training group from UCSD and a testing group from UCSF (i.e., an independent dataset). The normalization and PCA calculations were calculated on the training dataset and then applied to the testing dataset. (B) XGBoost was trained on 3 different sets of features: the clinical variables, association tracts, and the structural connectome.

(Fig. 1B). The cortical ROIs were obtained using Freesurfer's automatic parcellation process applied to each patient's T1-weighted image. These connection strength values were then compiled into 2-D symmetric $n \times n$ matrices to yield brain connectomes for each patient.

Probabilistic fiber tracking was performed using FDT (Behrens et al., 2003; Behrens et al., 2007). A GPU-accelerated implementation of FDT's BEDPOSTX was run to estimate the diffusion parameters at each voxel of the DWIs (Hernández et al., 2013). Probabilistic tractography was then performed using FDT's PROBTRACKX2 with the following parameters: 5000 samples, 2000 steps per sample, 0.5 mm step length, 0.2 curvature threshold, loop checking enabled on paths. Path distributions were also corrected for the inherent linear bias towards longer pathways in tractography algorithms (Hagmann et al., 2007). PROBTRACKX2 generates connectivity distributions from user-specified seed regions, in which voxels in the output brain image have values representing the number of streamlines (i.e., connection strength values) passing through them from the specified seed region. A full description of FDT's tractography implementation can be found elsewhere (Behrens et al., 2003; Behrens et al., 2007).

The cortical seed regions (i.e., ROIs) fed to PROBTRACKX2 were acquired from Freesurfer's automatic cortical parcellation process applied to T1-weighted images (Dale et al., 1999). The initial parcellation was performed with the DK (Desikan et al., 2006) atlas, which was modified to create more fine-grained ROIs. The following DK ROIs were split orthogonally to the long axis of the parcellation: *middle temporal*, *superior temporal*, *inferior temporal*, *fusiform*, *postcentral*, *precentral*, *middle frontal*, and *superior frontal*. The resulting atlas contained 98 ROIs (49 for each hemisphere). This number of ROIs is in line with the recommendations of studies which show that as the number of ROIs increase, connectomes become sparse with less reproducible data (Bonilha et al., 2015; Prčkovska et al., 2016). Choosing broad, functionally agnostic ROIs corresponds with the regional investigation of epilepsy across much of the existing literature allowing us to investigate broad circuits and avoid false-negative results.

The cortical parcellations were transformed from Freesurfer's conformed space to each subject's diffusion space using an affine transformation performed with FSL's FLIRT. To create a 98×98 symmetric connectivity matrix, the connectivity between each pair of source and destination ROIs were averaged. These connectivity values were also normalized by the sum of the number of voxels of the source and destination ROIs to account for differences in head size between subjects.

In summary, the structural connectivity means the number of probabilistic streamlines reaching ROI A when ROI B was seeded, averaged with the opposite direction, divided by the number of voxels in ROIs A and B, and corrected by the distance travelled by the fibers. After obtaining the connectivity matrices, the analysis was restricted to connections including at least one ROI in the left or right temporal lobes (Fig. 1C). This temporal lobe subnetwork was selected for analysis because TLE has been shown to affect connectivity both within the temporal lobe and between temporal and extratemporal regions (Besson et al., 2014).

2.5. Language impairment prediction models

XGBoost (Bernhardt et al., 2015) (v0.81), a type of decision tree algorithm (Chen and Guestrin, 2016), was selected for classification. Similar to other such decision tree models as random forest, XGBoost constructs many shallow trees (i.e., weak learners) that each, by themselves, do not provide optimal classification results. However, by assembling these weak learners, good classification performance is achieved. XGBoost is theoretically more robust than other decision tree algorithms to potential outliers that may exist in relatively small datasets. XGBoost also improves upon the random forest algorithm by using gradient boosting to minimize the training error, thereby focusing on the mistakes made by the previous trees and correcting the internal model to account for these outliers. Furthermore, XGBoost introduces regularization terms, which protect against overfitting to the training data by making the model more conservative and simpler. Hyperparameter optimization involves systematically tuning the internal parameters of the machine learning model to arrive at a set of parameters that yield maximum model performance. Hyperparameter optimization was performed to tune the performance of each individual model.

We tested each language prediction model by training an XGBoost classifier on UCSD patients (i.e., training set) and testing the model on UCSF patients (i.e., testing set), as the data from the second institution functioned as an external, independent dataset. Three XGBoost models were created for comparison: clinical variables (CV-model), tract-based (Tract-model), and structural connectome (SC-model). The framework is displayed in Fig. 2.

CV-model: The following clinical variables were included in our model: age, education, sex, handedness, MTS status, side of seizure

onset, age of onset, number of current AEDs, and seizure frequency (calculated as the total number of focal seizures with impaired awareness and generalized seizures per month).

Tract-model: The following right and left hemisphere temporal lobe fiber tracts were selected due to evidence of their disruption in TLE (Shen et al., 2017) and likely involvement in language processing: ARC, IFOF, ILF, and UNC.

SC-model: To reduce the high-dimensional nature of the SC (i.e., the SC had 4753 connections per subject), we applied principal component analysis (PCA) to lower the number of connections used as features in the model. PCA finds the directions in which the observations have the most variance and thus, best differentiate the data. These directions are known as the principal components (PCs) that can then be used in the model. The number of PCs that yielded maximum accuracy was assessed through hyperparameter optimization; it was found that having 40 PCs maximized accuracy of the connectome model. These 40 features represent <1% of the original 4753 features, a sizeable reduction.

2.6. Neuroanatomical interpretation of PCA

After assessing model performance, we sought to connect the important PCs back to overall white matter architecture. Mathematically, PCs are computed as linear combinations of the original features. A weight matrix for the PCs was extracted and the connections contributing to each PC were ranked. We took both a PC-centric approach and a Connection-centric approach to test whether our results were invariant to the method employed. The PC-centric approach identified the distribution of top-ranked connections that made up each important PC. For this, we took the top 1% of connections contributing to each important PC. In the Connection-centric approach, we examined the top connections across the important PCs by summing each connection's weight across all the important PCs and identifying the top 1% of connections. To understand the distribution of the identified connections from both of these approaches we identified the proportion of connections involving a set of regions across the brain. The brain was split into lateral temporal, ventromedial temporal, lateral frontal, inferior parietal, and superior frontoparietal cortex. Importantly, because we were using a temporal subnetwork, each connection contained at least one connection from the temporal lobe.

2.7. Statistical analysis

Analysis of variance (ANOVA) and Fisher's exact test were used to test for differences in demographic and clinical variables among the TLE-LI and TLE-NLI patients in the training and testing sets. Performance of each model was evaluated with receiver operating characteristic (ROC) curves, area under the ROC curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The thresholds for the predictions were chosen based on the point on the ROC curve that yielded maximum accuracy. Significant differences in model performance were assessed by creating 95% confidence intervals with 1000 bootstrapped samples by leaving 2 random patients out of the training dataset.

3. Results

3.1. Patient demographics and clinical variables

Table 1 shows the clinical group data for the four subgroups in this study (training TLE-LI and TLE-NLI as well as testing TLE-LI and TLE-NLI). There were differences in education [$F(3,79) = 3.736, p = 0.01$] among the groups. Pairwise comparisons revealed higher education in the training TLE-NLI group relative to the testing TLE-LI group ($p < .01$). No other characteristics reached significance. Additionally, we tested if the number of patients on zonisamide/topiramate differed

between groups and found that it did not ($p > .05$). Side of seizure onset approached significance (Fisher's Exact Test: 10.897, $p = .06$). Whereas the testing TLE-NLI was split 50% LTLE and 50% RTLE, the training TLE-NLI was 29% LTLE and 71% RTLE. However, a large portion of TLE-LI had RTLE onset in both the training (36%) and testing dataset (29%). A measure of nonverbal ability, the WASI Perceptual Reasoning Index, revealed no significant differences across groups ($p > .05$). In addition, the language-impaired and non-impaired groups did not differ in BDI-II [$F(3,79) = 1.73, p = .16$], BAI [$F(3,77) = 0.79, p = .50$], or in language laterality (Fisher's Exact = 1.89, $p = .45$).

3.2. Model performances

ROC curves for the three models are displayed in Fig. 3A, with the AUC for each model in Fig. 3B. The SC-model had a higher AUC (.73) compared to the CV-model (.59) and the Tract-model (.54). Next, we compared each model's optimal performance by identifying prediction thresholds based on the point on the ROC curve that yielded maximum accuracy. These additional model performance metrics are displayed in Table 2. The SC-model outperformed the other models across most of these metrics. For accuracy, the SC-model performed the best (79%), followed by the Tract-model (70%), with the CV-model having the poorest performance (64%). Across the other measures, the SC-model was also generally the highest; Tract- and CV-models had similar performances.

Randomization testing was performed to test the stability of model performances using 1000 replications (Table 3). This approach confirmed the previous results. AUC remained superior for the SC-model (73% +/- 2%), which was significantly higher than the Tract-model (54% +/- 3%; $p < .001$) and the CV-model (59% +/- 3%; $p < .001$). Similarly, accuracy remained superior for the SC-model (79% +/- 3%) which was significantly higher than the Tract-model (63% +/- 3%; $p < .001$) and the CV-model (65% +/- 2%; $p < .001$).

3.3. Feature importance from each model

Fig. 4 displays the main features that contributed to each model's performance, with the feature importance calculated by sklearn (Pedregosa et al., 2011). This value ranges from 0 (no importance) to 1 and sums to 1 across features within each model. For the CV-model, 8 of the 9 features contributed to the model, with only handedness making no measurable contribution. The features varied from .02 to .28 and in order are: patient age, seizure frequency, age of onset, education, sex, side of seizure onset, number of AEDs, and MTS status. For the Tract-model all 8 tracts contributed to model performance, ranging from .08 to .18. The top 2 tracts were both left hemisphere tracts, the left IFO and the ARC. However, overall, the distribution of tract importance was distributed fairly evenly between the left and right hemisphere tracts. For the SC-model only 9 out of the 40 PCs contributed to model performance. Qualitatively this broke down into two groups, a high contribution group of 4 PCs (HCG; importance range: .12-.32) and a low contribution group of 5 PCs (all importance: .04).

Next we sought to understand the anatomical distribution of the PCs. The PC-centric approach (Fig. 5), described in the methods, focused on understanding which connections made the most contribution to each individual important PC. We examined both the HCG PCs (i.e., only the top 4 PCs) and an "all contributing PCs" group (ACG; i.e., all nine contributing PCs). The largest percentage of important connections arose from the lateral temporal (HCG: 39%, ACG: 38%) compared to the ventromedial temporal (HCG: 21%, ACG: 24%) regions bilaterally. Of the temporal-to-extratemporal connections, the temporal to lateral frontal (HCG: 12%, ACG: 14%) and the temporal to superior frontoparietal cortex (HCG: 14%, ACG: 13%) had a higher number of connections contributing to model performance. The remaining two regions, the temporal to parietal (HCG: 9%, ACG: 6%) and temporal to occipital (HCG: 5%, ACG: 5%), had fewer contributions. Due to the bilateral

Table 1
Demographics and clinical variables

	Training data (UCSD)		Testing data (UCSF)		ANOVA	p-value
	Language impaired	No impairment	Language impaired	No impairment		
N	28	21	21	12		
Age (years)	38.2 (15.0)	38.4 (13.1)	29.2 (10.6)	33.3 (10.8)	2.503	.07
Education (Years)	13.4 (2.2)	14.8 (2.3)	12.7 (1.8)	13.8 (1.9)	3.736	.01¹
Age of Onset	18.2 (12.4)	26.2 (17.0)	17.1 (13.6)	23.3 (12.3)	1.985	.12
Duration (years)	20.0 (17.8)	12.2 (14.2)	12.1 (10.6)	10.0 (8.0)	2.200	.10
Number of AEDs	2.4 (1.0)	2.2 (0.7)	2.3(1.0)	2.3 (1.2)	.163	.92
Seizure frequency	9.5 (19.5)	10.8 (22.0)	11.0 (19.1)	3.9 (3.9)	.433	.73
BDI-II	15.1 (9.5)	17.8 (13.2)	13.0 (6.9)	10.4 (6.2)	1.73	.16
BAI	16.3 (12.7)	16.7 (13.1)	12.8 (9.0)	11.8 (8.7)	0.79	.50
					Fisher's Exact	p-value
Sex: M/F	11/17	9/12	10/11	6/6	.644	.91
Handedness: L/R/A	2/25/1	2/19/0	3/17/1	0/12/0	3.573	.18
MTS: Yes/No	14/14	9/12	11/10	5/7	.679	.90
Onset Side: L/R/Bilateral	14/10/4	6/15/0	13/6/2	6/6/0	10.897	.06
Language Side: L/R/B	16/4/1	10/2/2	9/4/4	9/0/2	1.89	.45
Neuropsychological Tests					ANCOVA*	p-value
BNT T-score	29.7 (8.6)	39.7 (10.0)	32.9 (9.4)	47.8 (10.5)	12.76	<.001
BNT Raw score	41.4 (9.6)	51.7 (5.6)	40.9 (9.9)	55.3 (2.5)		
ANT T-score	34.6 (12.2)	55.8 (10.9)	33.7 (14.1)	53.9 (4.89)	20.13	<.001
D-KEFS CF T-score	40.5 (7.2)	49.6 (9.8)	41.4 (13.4)	49.7 (8.9)	4.09	.010
Perceptual Reasoning IQ [#]	97.1 (17.5)	104.3 (12.9)	88.8 (15.6)	97.6 (14.8)	1.78	.16

TLE: temporal lobe epilepsy; F: females; M: males; L: left; R: right; A: ambidextrous; MTS: mesial temporal sclerosis; AEDs: antiepileptic drugs; standard deviations are presented inside the parentheses; BDI-II: Beck Depression Inventory-II; BAI: Beck Anxiety Inventory; BNT: Boston Naming Test; ANT: Auditory Naming Test;

¹ Pairwise comparisons revealed higher education in the training TLE-NLI group relative to the testing TLE-LI group.

* ANCOVA controlling for education.

[#] Two-subtest IQ based on performance on WASI Matrix Reasoning and Block Design Subtests.

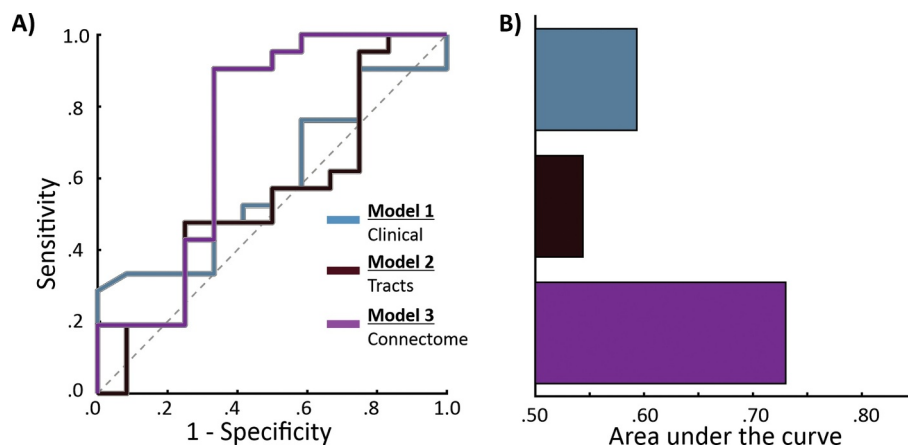


Fig. 3. ROC curves and Area Under the Curve comparing model performance when discriminating TLE-LI from TLE-NLI. (A) The ROC curves associated with 3 XGBoost models. (B) The area under the curve associated with each ROC curve.

Table 2
Model performances when trained on UCSD data and tested on UCSF data.

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Clinical Variables	0.59	0.64	0.86	0.25	0.67	0.50
Tracts	0.54	0.70	0.95	0.25	0.69	0.75
Connectomes	0.73	0.79	0.86	0.67	0.82	0.73

PPV: positive predictive value; NPV: negative predictive value; AUC: area under the ROC curve.

nature of the tract-model feature importance, we also assessed how interconnected the hemispheres were in the SC-model. Both the HCG and ACG had greater left-left connectivity (intra-hemispheric; HCG: 41%, ACG: 33%) compared to right-right connectivity (intra-hemispheric; HCG: 29%, ACG: 27%). Interestingly, both groups also had a high number of left-right connections (inter-hemispheric; HCG: 31%, ACG: 40%).

The connection-centric approach broadly replicated the findings from the PC-centric approach. This approach focused on summing

connection contribution across important PCs to understand which individual connections contributed the most these PCs. Again, the lateral temporal connections that contributed to model performance (HCG: 51%, ACG: 47%) were more numerous than the ventromedial temporal connections (HCG: 13%, ACG: 23%). Interestingly, the left-left connections were greater in this connection-centric approach than in the PC-centric approach. Left-left connections (HCG: 78%, ACG: 56%) were once again more numerous than right-right connections (HCG: 7%, ACG: 19%). There were fewer left-right connections (HCG: 15%, ACG: 26%). Finally, to visually illustrate the fiber connections which contributed the most to the important PCs, Fig. 6 displays the fiber tracts on an average brain from the top three region-region connections.

4. Discussion

Here we present evidence that structural networks derived from diffusion MRI provide superior prediction of language impairment in TLE compared to clinical features, and may serve as a stronger

Table 3
Model performances when trained on UCSD in a 1000-fold leave-2-out approach and tested on UCSF.

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Clinical Variables	0.59 +/- 0.03	0.65 +/- 0.02	0.82 +/- 0.10	0.36 +/- 0.15	0.69 +/- 0.04	0.54 +/- 0.12
Tracts	0.54 +/- 0.03	0.63 +/- 0.03	0.83 +/- 0.10	0.29 +/- 0.18	0.68 +/- 0.04	0.48 +/- 0.08
Connectomes	0.73 +/- 0.02 [#]	0.79 +/- 0.03 [#]	0.92 +/- 0.06	0.57 +/- 0.09 [#]	0.79 +/- 0.03	0.82 +/- 0.11 [#]

AUC: area under the ROC curve; PPV: positive predictive value; NPV: negative predictive value;

[#] = Connectome significantly better than Tracts and Clinical Variables.

+/- = Standard deviation on 1000 bootstrapped samples.

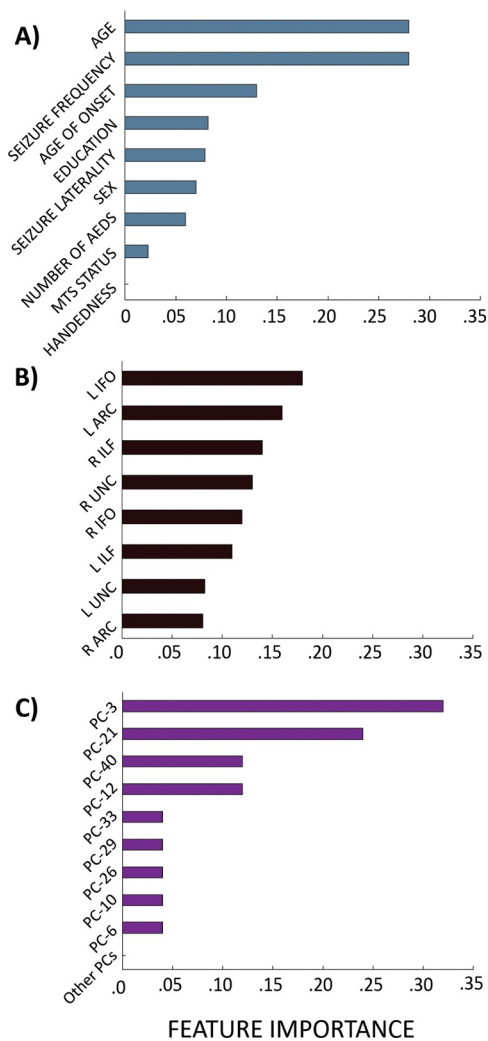


Fig. 4. Feature importance plots in each model. (A) Feature importance for the clinical model. (B) Feature importance for the tract model. (C) Feature importance for structural connectome model. Note that of the 40 PCs included in the model, only 9 PCs made a contribution.

biomarker of language impairment than measures of global microstructure derived from white matter tracts. White matter damage is a contributing factor to cognitive impairment in TLE (Leyden et al., 2015) and other syndromes (Shen et al., 2017). We purport that a network-based approach measuring cortical-cortical connections, such as the SC, may enable a more precise understanding of the specific microstructural alterations that lead to language impairment in TLE. The temporal connections highlighted by the SC temporal-subnetwork approach are widely distributed and bilateral, but with a majority arising from lateral as opposed to ventromedial temporal lobe regions. These findings are supported by a previous SC study with a more homogenous group of patients (i.e., all LTLE patients who were left-hemisphere

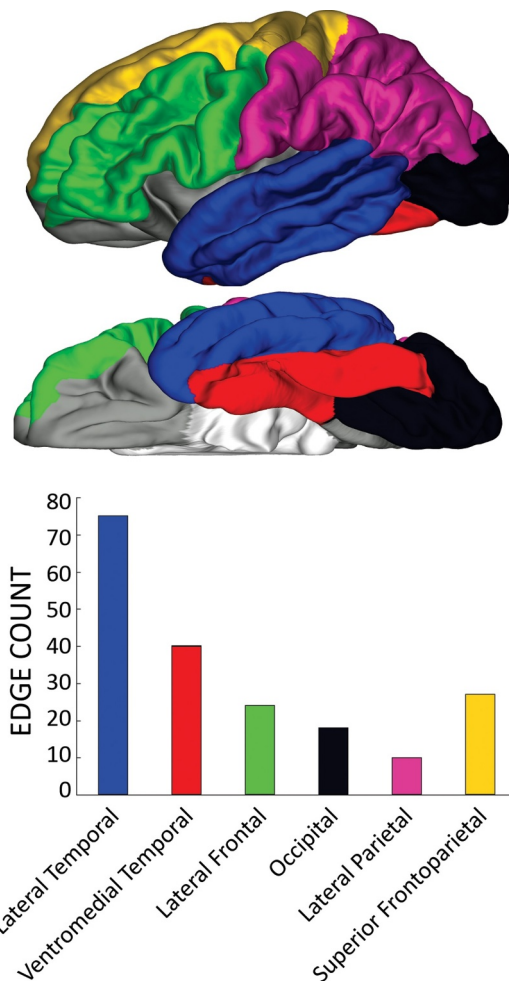


Fig. 5. Top white matter connections contributing to structural connectome performance. Distribution of connections (edges) in each of the color-coded regions (displayed brain is unilateral but connections were counted bilaterally) emphasizing a lateral temporal focus.

dominant for language (Munsell et al., 2019)), suggesting that this pattern is not specific to our particular TLE sample. We add to this literature by demonstrating that a network-based SC approach may capture the extent of the language network better than either clinical features or global measures of white matter tract integrity.

4.1. Clinical predictors of language impairment in TLE

Language impairment in TLE has been classically associated with left-hemisphere seizure onsets and therefore has been most extensively studied in LTLE (Busch et al., 2005; Raspall et al., 2005; Keary et al., 2007). However, recent studies examining patients with RTLE have identified a high proportion of patients with language impairment (Hermann et al., 1997; Alessio et al., 2006; Bell et al., 2001) and a

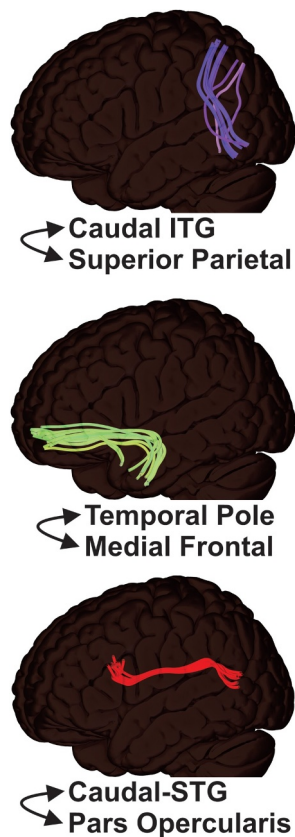


Fig. 6. Top 3 connections identified as contributing to the structural connectome PCs. Illustration of connection lines between the 3 ROIs which contributed to the 9 important PCs in the structural connectome. All 3 connections were left-left.

number of studies have found indistinguishable levels of language impairment in patients with RTLE versus LTLE (Ogden-Epker and Cullum, 2001; Langfitt and Rausch, 1996; Cherlow and Serafetinides, 1976; Stafiniak et al., 1990). In this paper, 33% of patients with RTLE and 55% of those with LTLE fell into the language impaired group. Thus, side-of-seizure onset had relatively low importance in the clinical model. This highlights the need to consider language impairment in patients with RTLE rather than exclude these patients from studies addressing pre-surgical language function. Although we believe that a bilateral white matter network contributes to language, it is likely that language impairment in our RTLE patients also suggests the presence of bilateral temporal lobe pathology (Seidenberg et al., 2005; Kaaden et al., 2011; Hermann et al., 2002).

Clinical features beyond side of seizure onset have been shown to affect the organization of the language network (Stewart et al., 2014; Sass et al., 1992). Our model identified most of the clinical features as having some contribution to model performance. An early age of seizure onset in particular has been shown to increase the likelihood of disrupted language function (Lee et al., 2013; Hermann et al., 2002) and was found to be one of the most important features in the clinical model. This is most likely due to disruption of white matter by seizures during critical stages of language development (Seidenberg et al., 2005; Kaaden et al., 2011; Hermann et al., 2002). It is of note that age and seizure frequency were the two top features contributing to the clinical model. Seizure frequency, like age of seizure onset, may provide clues as to the extent of underlying network disruption. Regarding age, the measures of language included in our study require speeded lexical production and retrieval, both of which are known to be influenced by age (Albert et al., 1988). Collectively, the language impaired and non-impaired patients proved to have a high degree of overlap in their

clinical/demographic characteristics, which resulted in modest performance of the clinical model. We purport that these clinical variables provide clues as to the likelihood of language impairment, whereas measures of white matter microstructure provide a more direct measure of the integrity and topology of the underlying language network.

4.2. The importance of white matter measures for understanding language impairment

Given the heterogeneity of the clinical features, an alternative approach to explaining shared cognitive deficits (i.e., cognitive phenotypes) is to identify shared neurobiological abnormalities underlying these deficits. Indeed, across imaging modalities neurobiological commonalities are found for cognitive phenotypes. Using language impairment in TLE as an example, the lateral temporal lobe was found to have increased white matter path length (Reyes et al., 2019) and lower functional activations (Kaestner et al., 2019). Here we focus on white matter as its disruption, and therefore cognitive disruption, may be driven by poor development of (or damage to) white matter tracts in patients (Seidenberg et al., 2005; Kaaden et al., 2011; Hermann et al., 2002).

There are many methods of characterizing white matter integrity, and determining the clinical usefulness of various measures is an ongoing goal. Perhaps the most common approach is to derive a single measure of microstructural integrity for white matter bundles, here represented by the tract-model, and relating it to cognitive ability (Leyden et al., 2015). An alternative approach, here represented by the SC-model, focuses on individualized network patterns (Munsell et al., 2017), measured as pairwise gray matter connections that may not be captured in the large association tracts. In this study, the tract-model performed with high sensitivity but low specificity for detecting language impairment in TLE. This is in line with our previous study on language impairment (Kaestner et al., 2019), which revealed widespread damage of fiber bundles which did not differ significantly between language impaired and non-impaired individuals. This pattern implies widespread damage to white matter in TLE, and raises the concern that a summary measure of tract integrity is insufficient for isolating pathology associated with language impairment. Future efforts may focus on identifying subsections of fiber bundles associated with particular cognitive operations. Efforts at this more precise subdividing of fiber bundles have indeed shown promise (Voets et al., 2017; Duffau, 2015).

The SC-model had a significantly higher overall accuracy and ROC performance driven mainly by a higher specificity (i.e., excluding patients who are not language impaired). Anatomically, the SC-model replicated the lateral temporal lobe as the most important region in identifying language impairment in TLE. Although our SC identified widespread connections that are not included in our association tracts, it is noteworthy that two of the top three connections identified as contributing to the PCs appear to approximate the ARC and UNC (see Fig. 6). The ARC (Upadhyay et al., 2008) and UNC (Lu et al., 2002) have both been associated with naming performance (McDonald et al., 2008) and with semantic fluency (Wang et al., 2010). Thus, our SC approach appears to both capture white matter association tracts known to be important to language, while also including less obvious and/or understood connections that may also be critical for language performance.

4.3. The complex and bilateral nature of language networks in epilepsy

Due to the long duration of epilepsy suffered by many patients with TLE, (re)organization of the language network is of keen interest for understanding pre-surgical language function, language dominance, and estimating risk for post-surgical language decline (Balter et al., 2016). Language lateralization is typically evaluated functionally, with either IAP or fMRI, both of which aim to categorize language trarily

as either left-dominant, right-dominant, or bilateral (Binder, 2011; Janecek et al., 2013). This is the approach generally followed in clinical practice and there is evidence that this approach holds utility for predicting postoperative language decline (Szaflarski et al., 2017). However, studies using fMRI have demonstrated that this trinary categorization of language is oversimplified, with some studies positing up to 15 different patterns of possible language organization in TLE (Berl et al., 2014) and there is evidence that the prediction of postoperative language outcome is contingent on the amount of tissue resected displaying pre-surgical fMRI-activation (You et al., 2019). Studies of healthy controls find that while in the overall population language is more left- than right-hemisphere dominant, there exists a continuum of laterality rather than discrete categories (Knecht et al., 2002). Even in patients categorized as left-dominant for language, the right-hemisphere structural networks are important; in a recent DTI study using graph theory measures, a bilateral distribution of nodes was found to explain 60% of the variance in naming performance (Munsell et al., 2019). This is supported by fMRI data implicating distributed bilateral networks associated with language regions in both resting state (Doucet et al., 2017) and language task-derived (Trimmel et al., 2018) networks. Measures of broadly distributed, bilateral functional global networks also have predictive power for post-surgical outcome (Audrain et al., 2018). In the present study using a heterogeneous TLE sample, we found bilateral connections contributing to SC-model and tract-model performance; in the SC-model this included not only left-left connections (which predominated) but also right-right and inter-hemispheric connections.

However, following right temporal lobectomies, only ~5% of patients display a significant decline in naming (Busch et al., 2018). This suggests that the right-sided nodes and connections in patients with typical (i.e., left hemisphere) language dominance are likely contributing but not essential, at least for language functions typically tested post-surgically. Understanding right-sided language contributions in TLE is complicated by the uncertainty of patient-by-patient reorganization in many studies. As the field continues to advance in understanding the additional connections beyond the left perisylvian network contributing to language, it will be important to consider the interplay between functional and anatomical measures of connectivity in the context of a bi-hemispheric network approach (Chu et al., 2018). Recent studies have begun to bridge the relationship between functional and anatomical organization of the language network in TLE (Osipowicz et al., 2016; Chang et al., 2017). For example, TLE patients with a rightward shift in language activations on fMRI have been shown to be more likely to have preserved language if they also show a rightward asymmetry in integrity of the ARC (Chang et al., 2017). In navigating the transition from trinary to continuum thinking, the SC approach, and functional connectivity approaches, can further exploit the complexity of the language network and begin to move toward individualized patterns rather than discrete categories.

4.4. Limitations and future directions

Although the overall performance of the SC-model was significantly higher than the alternative models, performance still can be improved. Machine learning approaches are powerful tools for understanding the unseen patterns in data but require many examples in the training data. Although our sample size of 82 patients is higher than in many previous machine learning and connectome studies in TLE, and we took precautions to avoid overfitting, the accumulation of larger amounts of neuroimaging data will be necessary to ensure the generalizability of our findings. Large-scale efforts such as ENIGMA-Epilepsy may afford this opportunity (Whelan et al., 2018). Similarly, though most machine learning efforts focus on a single institution, we used a model trained on one institutions' data and tested on another. Future studies will need to continue expanding to additional institutions to ensure broad generalizability.

Here we classified patients as language impaired if they were impaired on any two of three language measures of naming and fluency. However, these three measures are unlikely to have completely overlapping neuro-anatomical substrates, potentially creating confusion in the model. However, individual language tasks can be non-specific and may capture deficits outside the language network, which could also obscure the relevant networks. For example, although the vast majority of language impaired individuals in our study were impaired on both visual and auditory naming (providing greater confidence of a naming deficit), there may be a subset of RTLE patients that are impaired on BNT due to a visual object recognition rather than language impairment (Drane et al., 2015; Drane et al., 2008). Our requirement that patients be impaired on at least two language measures was implemented to mitigate this likelihood. However, future studies with a larger sample size that evaluate impairment on each individual test may identify sub-phenotypes associated with specific network alterations, achieving more precise categorizations.

5. Conclusion

A recent review of connectomics in TLE noted that further development of the field would require comparisons of connectome approaches with more conventional measures (Tavakol et al., 2019). Here we add to this literature by demonstrating that the SC may capture the extent of the language network better than either clinical features or global measures of fiber bundle microstructure. While current clinical or track-based approaches may oversimplify the language network, they have the virtue of being easy to implement and interpret. As network methodologies increase in popularity, efforts are needed to easily extract and interpret meaningful biomarkers. Here we used PCA for dimensionality reduction. In other studies graph theory measures have proven to be effective in predicting seizure outcome (Taylor et al., 2018) and in characterizing cognitive phenotypes in TLE (Reyes et al., 2019). As the field matures, there will need to be convergence on accepted methods for connectome generation (Sotiropoulos and Zalesky, 2019). Here we chose an anatomically-based, cognitively agnostic method for connectome reconstruction, with a number of ROIs in line with recommendations from previous studies (Bonilha et al., 2015; Prčkovska et al., 2016). Alternative approaches including functionally defined ROIs that take into account the cognitive construct under study may further improve performance. Further studies are needed to determine the best dimensionality reduction and connectome generation approaches to white matter data and to identify the neural networks that underlie specific cognitive impairments in TLE.

Funding

Supported by NIH/NINDS R01 NS065838 (CRM); R21 NS107739 (CRM; LB); F31 NS111883-01 (AR); R01 NS088748 (DLD).

Declaration of Competing Interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.nicl.2019.102125](https://doi.org/10.1016/j.nicl.2019.102125).

References

- Albert, MS, DeKosky, ST, Dickson, D, et al., 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement* 7 (3), 270–279.
- Albert, MS, Heller, HS, Milberg, W, 1988. Changes in naming ability with age. *Psychol. Aging* 3 (2), 173.

- Alessio, A, Bonilha, L, Rorden, C, et al., 2006. Memory and language impairments and their relationships to hippocampal and perirhinal cortex damage in patients with medial temporal lobe epilepsy. *Epilepsy Behav.* 8 (3), 593–600. <https://doi.org/10.1016/j.yebeh.2006.01.007>.
- Allone, C, Lo Buono, V, Corallo, F, et al., 2017. Neuroimaging and cognitive functions in temporal lobe epilepsy: a review of the literature. *J. Neurol. Sci.* 381, 7–15. <https://doi.org/10.1016/j.jns.2017.08.007>.
- Ashtari, M., 2012. Anatomy and functional role of the inferior longitudinal fasciculus: a search that has just begun. *Dev. Med. Child Neurol.* 54 (1), 6–7.
- Audrain, S, Barnett, AJ, McAndrews, MP, 2018. Language network measures at rest indicate individual differences in naming decline after anterior temporal lobe resection. *Hum. Brain Mapp.* 39 (11), 4404–4419. <https://doi.org/10.1002/hbm.24281>.
- Balter, S, Lin, G, Leyden, KM, Paul, BM, McDonald, CR, July 2016. Neuroimaging correlates of language network impairment and reorganization in temporal lobe epilepsy. *Brain Lang.* <https://doi.org/10.1016/j.bandl.2016.06.002>.
- Beaulieu, C., 2014. The biological basis of diffusion anisotropy. *Diffusion MRI*. pp. 155–183 Elsevier.
- Beck, AT, Epstein, N, Brown, G, Steer, RA, 1988. An inventory for measuring clinical anxiety: psychometric properties. *J. Consult. Clin. Psychol.* 56 (6), 893.
- Beck AT, Steer RA, Brown GK. Beck depression inventory-II. *San Antonio.* 1996;78(2):490–498.
- Behrens, TE, Berg, HJ, Jbabdi, S, Rushworth, MF, Woolrich, MW, 2007. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage* 34 (1), 144–155. <https://doi.org/10.1016/j.neuroimage.2006.09.018>.
- Behrens, TE, Woolrich, MW, Jenkinson, M, et al., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50 (5), 1077–1088. <https://doi.org/10.1002/mrm.10609>.
- Bell, B, Lin, JJ, Seidenberg, M, Hermann, B, 2011. The neurobiology of cognitive disorders in temporal lobe epilepsy. *Nat. Rev. Neurol.* 7 (3), 154–164. <https://doi.org/10.1038/nrneuro.2011.3>.
- Bell, BD, Hermann, BP, Woodard, AR, et al., 2001. Object naming and semantic knowledge in temporal lobe epilepsy. *Neuropsychology* 15 (4), 434.
- Berl, MM, Zimmaro, LA, Khan, OI, et al., 2014. Characterization of atypical language activation patterns in focal epilepsy: language activation patterns. *Ann. Neurol.* 75 (1), 33–42. <https://doi.org/10.1002/ana.24015>.
- Bernhardt, BC, Bonilha, L, Gross, DW, 2015. Network analysis for a network disorder: The emerging role of graph theory in the study of epilepsy. *Epilepsy Behav.* 50, 162–170. <https://doi.org/10.1016/j.yebeh.2015.06.005>.
- Besson, P, Dinkelaeker, V, Valabregue, R, et al., 2014. Structural connectivity differences in left and right temporal lobe epilepsy. *Neuroimage* 100, 135–144. <https://doi.org/10.1016/j.neuroimage.2014.04.071>.
- Binder, JR., 2011. Functional MRI is a valid noninvasive alternative to Wada testing. *Epilepsy Behav.* 20 (2), 214–222.
- Binder, LM, Iverson, GL, Brooks, BL, 2009. To Err is human: “Abnormal” neuropsychological scores and variability are common in healthy adults. *Arch. Clin. Neuropsychol.* 24 (1), 31–46. <https://doi.org/10.1093/arclin/acn001>.
- Bondi, MW, Edmonds, EC, Jak, AJ, et al., 2014. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J. Alzheimers Dis.* 42 (1), 275–289.
- Bonilha, L, Gleichgerrcht, E, Fridriksson, J, et al., 2015. Reproducibility of the structural brain connectome derived from diffusion tensor imaging. *PLoS One* 10 (8), e0135247. <https://doi.org/10.1371/journal.pone.0135247>.
- Bonilha, L, Jensen, JH, Baker, N, et al., 2015. The brain connectome as a personalized biomarker of seizure outcomes after temporal lobectomy. *Neurology* 84 (18), 1846–1853. <https://doi.org/10.1212/WNL.0000000000001548>.
- Busch, RM, Frazier, TW, Haggerty, KA, Kubu, CS, 2005. Utility of the boston naming test in predicting ultimate side of surgery in patients with medically intractable temporal lobe epilepsy. *Epilepsia* 46 (11), 1773–1779. <https://doi.org/10.1111/j.1528-1167.2005.00300.x>.
- Busch, RM, Hogue, O, Kattan, MW, et al., 2018. Nomograms to predict naming decline after temporal lobe surgery in adults with epilepsy. *Neurology* 91 (23), e2144–e2152.
- Catani, M, Jones, DK, Fyhtche, DH, 2005. Perisylvian language networks of the human brain. *Ann. Neurol.* 57 (1), 8–16. <https://doi.org/10.1002/ana.20319>.
- Caverzasi, E, Papinutto, N, Amirbekian, B, Berger, MS, Henry, RG, 2014. Q-ball of inferior fronto-occipital fasciculus and beyond. *PLoS One* 9 (6), e100274.
- Chang, Y-HA, Kemmotsu, N, Leyden, KM, et al., 2017. Multimodal imaging of language reorganization in patients with left temporal lobe epilepsy. *Brain Lang.* 170, 82–92. <https://doi.org/10.1016/j.bandl.2017.03.012>.
- Chen, T, Guestrin, C, 2016. XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>. KDD '16. New York, NY, USA: ACM.
- Cherlow, DG, Serafetinides, EA., 1976. Speech and memory assessment in psychomotor epileptics. *Cortex* 12 (1), 21–26.
- Chu, S-H, Parhi, KK, Lenglet, C, 2018. Function-specific and enhanced brain structural connectivity mapping via joint modeling of diffusion and functional MRI. *Sci. Rep.* 8 (1). <https://doi.org/10.1038/s41598-018-23051-9>.
- Dabbs, K, Jones, J, Seidenberg, M, Hermann, B, 2009. Neuroanatomical correlates of cognitive phenotypes in temporal lobe epilepsy. *Epilepsy Behav.* 15 (4), 445–451. <https://doi.org/10.1016/j.yebeh.2009.05.012>.
- Dale, AM, Fischl, B, Sereno, MI, 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* 9 (2), 195–207. <https://doi.org/10.1006/nimg.1998.0396>.
- Dale, AM, Fischl, B, Sereno, MI, 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* 9 (2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>.
- Delis, DC, Kaplan, E, Kramer, JH, 2001. *Delis-Kaplan Executive Function System*. Examiners Manual. Psychological Corporation.
- Desikan, RS, Ségonne, F, Fischl, B, et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31 (3), 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>.
- Doucet, GE, He, X, Sperling, MR, Sharan, A, Tracy, JI, 2017. From “rest” to language task: Task activation selects and prunes from broader resting-state network. *Hum. Brain Mapp.* 38 (5), 2540–2552. <https://doi.org/10.1002/hbm.23539>.
- Drane, DL, Loring, DW, Voets, NL, et al., 2015. Better object recognition and naming outcome with MRI-guided stereotactic laser amygdalohippocampotomy for temporal lobe epilepsy. *Epilepsia* 56 (1), 101–113.
- Drane, DL, Ojemann, GA, Aylward, E, et al., 2008. Category-specific naming and recognition deficits in temporal lobe epilepsy surgical patients. *Neuropsychologia* 46 (5), 1242–1255.
- Drane, DL, Pedersen, NP., 2019. Knowledge of language function and underlying neural networks gained from focal seizures and epilepsy surgery. *Brain Lang.* 189, 20–33. <https://doi.org/10.1016/j.bandl.2018.12.007>.
- Duffau, H., 2015. Stimulation mapping of white matter tracts to study brain functional connectivity. *Nat. Rev. Neurol.* 11 (5), 255–265. <https://doi.org/10.1038/nrneuro.2015.51>.
- Edmonds, EC, Delano-Wood, L, Clark, LR, et al., 2015. Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. *Alzheimers Dement* 11 (4), 415–424.
- Edmonds, EC, Eppig, J, Bondi, MW, et al., 2016. Heterogeneous cortical atrophy patterns in MCI not captured by conventional diagnostic criteria. *Neurology* 87 (20), 2108–2116.
- Gleichgerrcht, E, Munsell, B, Bhatia, S, et al., 2018. Deep learning applied to whole-brain connectome to determine seizure control after epilepsy surgery. *Epilepsia* 59 (9), 1643–1654. <https://doi.org/10.1111/epi.14528>.
- Hagler, DJ, Ahmadi, ME, Kuperman, J, et al., 2009. Automated white-matter tractography using a probabilistic diffusion tensor atlas: Application to temporal lobe epilepsy. *Hum. Brain Mapp.* 30 (5), 1535–1547. <https://doi.org/10.1002/hbm.20619>.
- Hagler, DJ, Ahmadi, ME, Kuperman, J, et al., 2009. Automated white-matter tractography using a probabilistic diffusion tensor atlas: Application to temporal lobe epilepsy. *Hum. Brain Mapp.* 30 (5), 1535–1547. <https://doi.org/10.1002/hbm.20619>.
- Hagmann, P, Kurlant, M, Gigandet, X, et al., 2007. Mapping human whole-brain structural networks with diffusion MRI. *Sporns O. PLoS One* 2 (7), e597. <https://doi.org/10.1371/journal.pone.0000597>.
- Hamberger, MJ, Seidel, WT, 2003. Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *J. Int. Neuropsychol. Soc.* 9 (03). <https://doi.org/10.1017/S135561770393013X>.
- Hasan, KM, Iftikhar, A, Kamali, A, et al., 2009. Development and aging of the healthy human brain uncinate fasciculus across the lifespan using diffusion tensor tractography. *Brain Res.* 1276, 67–76.
- Hermann, B, Davies, K, Foley, K, Bell, B, 1999. Visual confrontation naming outcome after standard left anterior temporal lobectomy with sparing versus resection of the superior temporal gyrus: a randomized prospective clinical trial. *Epilepsia* 40 (8), 1070–1076.
- Hermann, B, Seidenberg, M, Bell, B, et al., 2002. The neurodevelopmental impact of childhood-onset temporal lobe epilepsy on brain structure and function. *Epilepsia* 43 (9), 1062–1071.
- Hermann, B, Seidenberg, M, Lee, E-J, Chan, F, Rutecki, P, 2007. Cognitive phenotypes in temporal lobe epilepsy. *J. Int. Neuropsychol. Soc.* 13 (01). <https://doi.org/10.1017/S135561770707004X>.
- Hermann, BP, Seidenberg, M, Schoenfeld, J, Davies, K, 1997. Neuropsychological characteristics of the syndrome of mesial temporal lobe epilepsy. *Arch. Neurol.* 54 (4), 369–376. <https://doi.org/10.1001/archneur.1997.00550160019010>.
- Hernández, M, Guerrero, GD, Cecilia, JM, García, JM, Inuggi, A, 2013. Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using GPUs. *PLoS One* 8 (4), 61892. <https://doi.org/10.1371/journal.pone.0061892>.
- Holland, D, Kuperman, JM, Dale, AM, 2010. Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. *Neuroimage* 50 (1), 175–183. <https://doi.org/10.1016/j.neuroimage.2009.11.044>.
- Jak, AJ, Bondi, MW, Delano-Wood, L, et al., 2009. Quantification of five neuropsychological approaches to defining mild cognitive impairment. *Am. J. Geriatr. Psychiatry* 17 (5), 368–375.
- Janecek, JK, Swanson, SJ, Sabsevitz, DS, et al., 2013. Language lateralization by fMRI and Wada testing in 229 patients with epilepsy: Rates and predictors of discordance. *Epilepsia* 54 (2), 314–322. <https://doi.org/10.1111/epi.12068>.
- Jovicich, J, Czanner, S, Greve, D, et al., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 30 (2), 436–443. <https://doi.org/10.1016/j.neuroimage.2005.09.046>.
- Kaaden, S, Quesada, CM, Urbach, H, et al., 2011. Neurodevelopmental disruption in early-onset temporal lobe epilepsy: evidence from a voxel-based morphometry study. *Epilepsy Behav.* 20 (4), 694–699.
- Kaestner, E, Reyes, A, Macari, AC, et al., 2019. Identifying the neural basis of a language-impaired phenotype of temporal lobe epilepsy. *Epilepsia*.
- Kaplan E, Goodglass H, Weintraub S. *Boston Naming Test*. Pro-ed; 2001.
- Keary, TA, Frazier, TW, Busch, RM, Kubu, CS, Iampietro, M, 2007. Multivariate neuropsychological prediction of seizure lateralization in temporal epilepsy surgical cases: NP prediction of seizure lateralization. *Epilepsia* 48 (8), 1438–1446. <https://doi.org/10.1111/j.1528-1167.2007.01098.x>.
- Kim, J, Parker, D, Whyte, J, et al., 2014. Disrupted structural connectome is associated with both psychometric and real-world neuropsychological impairment in diffuse traumatic brain injury. *J. Int. Neuropsychol. Soc.* 20 (9), 887–896. <https://doi.org/10.1002/hbm.20619>.

- 10.1017/s1355617714000812.
- Knecht, S, Flöel, A, Dräger, B, et al., 2002. Degree of language lateralization determines susceptibility to unilateral brain lesions. *Nat. Neurosci.* 5 (7), 695–699. <https://doi.org/10.1038/nn868>.
- Kwan, P, Arzimanoglou, A, Berg, AT, et al., 2010. Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia* 51 (6), 1069–1077.
- Langfitt, JT, Rausch, R, 1996. Word-finding deficits persist after left anterotemporal lobectomy. *Arch. Neurol.* 53 (1), 72–76.
- Lee, C-Y, Tabesh, A, Benitez, A, Helpert, JA, Jensen, JH, Bonilha, L, 2013. Microstructural integrity of early- versus late-myelinating white matter tracts in medial temporal lobe epilepsy. *Epilepsia* 54 (10), 1801–1809. <https://doi.org/10.1111/epi.12353>.
- Leyden, KM, Kucukboyaci, NE, Puckett, OK, Lee, D, Loi, RQ, 2015. What does diffusion tensor imaging (DTI) tell us about cognitive networks in temporal lobe epilepsy. *Quant. Imaging Med. Surg.* 5 (2), 17.
- Litvan, I, Goldman, JG, Tröster, AI, et al., 2012. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: movement disorder society task force guidelines. *Mov. Disord.* 27 (3), 349–356. <https://doi.org/10.1002/mds.24893>.
- Lu, LH, Crosson, B, Nadeau, SE, et al., 2002. Category-specific naming deficits for objects and actions: semantic attribute and grammatical role hypotheses. *Neuropsychologia* 40 (9), 1608–1621.
- Martin, RC, Loring, DW, Meador, KJ, Lee, GP, 1990. The effects of lateralized temporal lobe dysfunction on normal and semantic word fluency. *Neuropsychologia* 28 (8), 823–829.
- McDonald, CR, Ahmadi, ME, Hagler, DJ, et al., 2008. Diffusion tensor imaging correlates of memory and language impairments in temporal lobe epilepsy. *Neurology* 71 (23), 1869–1876.
- McDonald, CR, Leyden, KM, Hagler, DJ, et al., 2014. White matter microstructure complements morphometry for predicting verbal memory in epilepsy. *Cortex* 58, 139–150. <https://doi.org/10.1016/j.cortex.2014.05.014>.
- Munsell, BC, Hofesmann, E, Delgado, J, Styner, M, Bonilha, L, 2017. Identifying sub-network fingerprints in structural connectomes: a data-driven approach. In: Wu, G, Laurienti, P, Bonilha, L, Munsell, BC (Eds.), *Connectomics in Neuroimaging 10511*. Springer International Publishing, Cham, pp. 79–88. https://doi.org/10.1007/978-3-319-67159-8_10.
- Munsell, BC, Wu, G, Fridriksson, J, et al., 2019. Relationship between neuronal network architecture and naming performance in temporal lobe epilepsy: a connectome based approach using machine learning. *Brain Lang.* 193, 45–57. <https://doi.org/10.1016/j.bandl.2017.08.006>.
- Ogden-Epker, M, Cullum, CM., 2001. Quantitative and qualitative interpretation of neuropsychological data in the assessment of temporal lobectomy candidates. *Clin. Neuropsychol.* 15 (2), 183–195.
- Osipowicz, K, Sperling, MR, Sharan, AD, Tracy, JI, 2016. Functional MRI, resting state fMRI, and DTI for predicting verbal fluency outcome following resective surgery for temporal lobe epilepsy. *J. Neurosurg.* 124 (4), 929–937. <https://doi.org/10.3171/2014.9.JNS131422>.
- Oyegbile, TO, Dow, C, Jones, J, et al., 2004. The nature and course of neuropsychological morbidity in chronic temporal lobe epilepsy. *Neurology* 62 (10), 1736–1742.
- Pedregosa, F, Varoquaux, G, Gramfort, A, et al., 2011. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Poepfel, D, Emmorey, K, Hickok, G, Pyllkanen, L, 2012. Towards a new neurobiology of language. *J. Neurosci.* 32 (41), 14125–14131. <https://doi.org/10.1523/JNEUROSCI.3244-12.2012>.
- Prčková, V, Rodrigues, P, Puigdemívil Sanchez, A, et al., 2016. Reproducibility of the structural connectome reconstruction across diffusion methods. *J. Neuroimaging* 26 (1), 46–57. <https://doi.org/10.1111/jon.12298>.
- Price, CJ., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62 (2), 816–847. <https://doi.org/10.1016/j.neuroimage.2012.04.062>.
- Raspall, T, Doñate, M, Boget, T, et al., 2005. Neuropsychological tests with lateralizing value in patients with temporal lobe epilepsy: Reconsidering material-specific theory. *Seizure* 14 (8), 569–576. <https://doi.org/10.1016/j.seizure.2005.09.007>.
- Reyes, A, Kaestner, E, Bahrami, N, et al., 2019. Cognitive phenotypes in temporal lobe epilepsy are associated with distinct patterns of white matter network abnormalities. *Neurology* 92 (17), e1957–e1968.
- Reyes, A, Kaestner, E, Bahrami, N, et al., 2019. Cognitive phenotypes in temporal lobe epilepsy are associated with distinct patterns of white matter network abnormalities. *Neurology* 92 (17), e1957–e1968. <https://doi.org/10.1212/WNL.0000000000007370>.
- Rodríguez-Cruces, R, Velázquez-Pérez, L, Rodríguez-Leyva, I, et al., 2018. Association of white matter diffusion characteristics and cognitive deficits in temporal lobe epilepsy. *Epilepsy Behav.* 79, 138–145. <https://doi.org/10.1016/j.yebeh.2017.11.040>.
- Sass, KJ, Sass, A, Westerveld, M, et al., 1992. Specificity in the correlation of verbal memory and hippocampal neuron loss: Dissociation of memory, language, and verbal intellectual ability. *J. Clin. Exp. Neuropsychol.* 14 (5), 662–672. <https://doi.org/10.1080/01688639208402854>.
- Seidenberg, M, Kelly, KG, Parrish, J, et al., 2005. Ipsilateral and contralateral MRI volumetric abnormalities in chronic unilateral temporal lobe epilepsy and their clinical correlates. *Epilepsia* 46 (3), 420–430.
- Shen, X, Finn, ES, Scheinost, D, et al., 2017. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* 12 (3), 506–518. <https://doi.org/10.1038/nprot.2016.178>.
- Sotiropoulos, SN, Zalesky, A, 2019. Building connectomes using diffusion MRI: why, how and but. *NMR Biomed.* 32 (4), e3752. <https://doi.org/10.1002/nbm.3752>.
- Sporns, O, Tononi, G, Kötter, R, 2005. The human connectome: A structural description of the human brain. *PLoS Comput. Biol.* 1 (4), 0245–0251. <https://doi.org/10.1371/journal.pcbi.0010042>.
- Stafiniak, P, Saykin, AJ, Sperling, MR, et al., 1990. Acute naming deficits following dominant temporal lobectomy: prediction by age at 1st risk for seizures. *Neurology* 40 (10), 1509.
- Stewart, CC, Swanson, SJ, Sabsevitz, DS, Rozman, ME, Janeczek, JK, Binder, JR, 2014. Predictors of language lateralization in temporal lobe epilepsy. *Neuropsychologia* 60, 93–102. <https://doi.org/10.1016/j.neuropsychologia.2014.05.021>.
- Szafarski, JP, Gloss, D, Binder, JR, et al., 2017. Practice guideline summary: use of fMRI in the presurgical evaluation of patients with epilepsy: report of the guideline development, dissemination, and implementation subcommittee of the american academy of neurology. *Neurology* 88 (4), 395–402.
- Tavakoli, S, Royer, J, Lowe, AJ, et al., 2019. Neuroimaging and connectomics of drug-resistant epilepsy at multiple scales: From focal lesions to macroscale networks. *Epilepsia* 60 (4), 593–604. <https://doi.org/10.1111/epi.14688>.
- Taylor, PN, Sinha, N, Wang, Y, et al., 2018. The impact of epilepsy surgery on the structural connectome and its relation to outcome. *NeuroImage Clin.* 18 (January), 202–214. <https://doi.org/10.1016/j.nicl.2018.01.028>.
- Trimmel, K, van Graan, AL, Caciagli, L, et al., 2018. Left temporal lobe language network connectivity in temporal lobe epilepsy. *Brain* 141 (8), 2406–2418. <https://doi.org/10.1093/brain/aww164>.
- Upadhyay, J, Hallock, K, Ducros, M, Kim, D-S, Ronen, I, 2008. Diffusion tensor spectroscopy and imaging of the arcuate fasciculus. *Neuroimage* 39 (1), 1–9.
- Voets, NL, Bartsch, A, Plaha, P, 2017. Brain white matter fibre tracts: a review of functional neuro-oncological relevance. *J. Neurol. Neurosurg. Psychiatry* 88 (12), 1017–1025. <https://doi.org/10.1136/jnnp-2017-316170>.
- Wang, X, Lang, S, Hong, LU, Lin, MA, Yan-ling, MAO, Yang, F, 2010. Changes in extra-temporal integrity and cognition in temporal lobe epilepsy: a diffusion tensor imaging study. *Neurol. India* 58 (6), 891.
- Wechsler, D, WASI, 1999. (Wechsler adult scale–Reduced). NY Psychol. Corp.
- Whelan, CD, Altmann, A, Botía, JA, et al., 2018. Structural brain abnormalities in the common epilepsies assessed in a worldwide ENIGMA study. *Brain* 141 (2), 391–408. <https://doi.org/10.1093/brain/aww341>.
- You, X, Zachery, AN, Fanto, EJ, et al., 2019. fMRI prediction of naming change after adult temporal lobe epilepsy surgery: activation matters. *Epilepsia* 60 (3), 527–538. <https://doi.org/10.1111/epi.14656>.