

Robust Lesion Segmentation on MRI of Patients with Multiple Sclerosis

John Muschelli, PhD - Johns Hopkins Bloomberg School of Public Health
http://johnmuschelli.com/Genentech_Talk_2018.html

April 30, 2018

Overview of Work/Research

- Neuroimaging and R (Neuroconductor Project)
- R Package Development/"Data Science"
- Segmentation/Classification of:
 - White Matter Lesions in Multiple Sclerosis
 - Brain vs. Skull (CT)
 - Brain Hemorrhage/Stroke (CT)

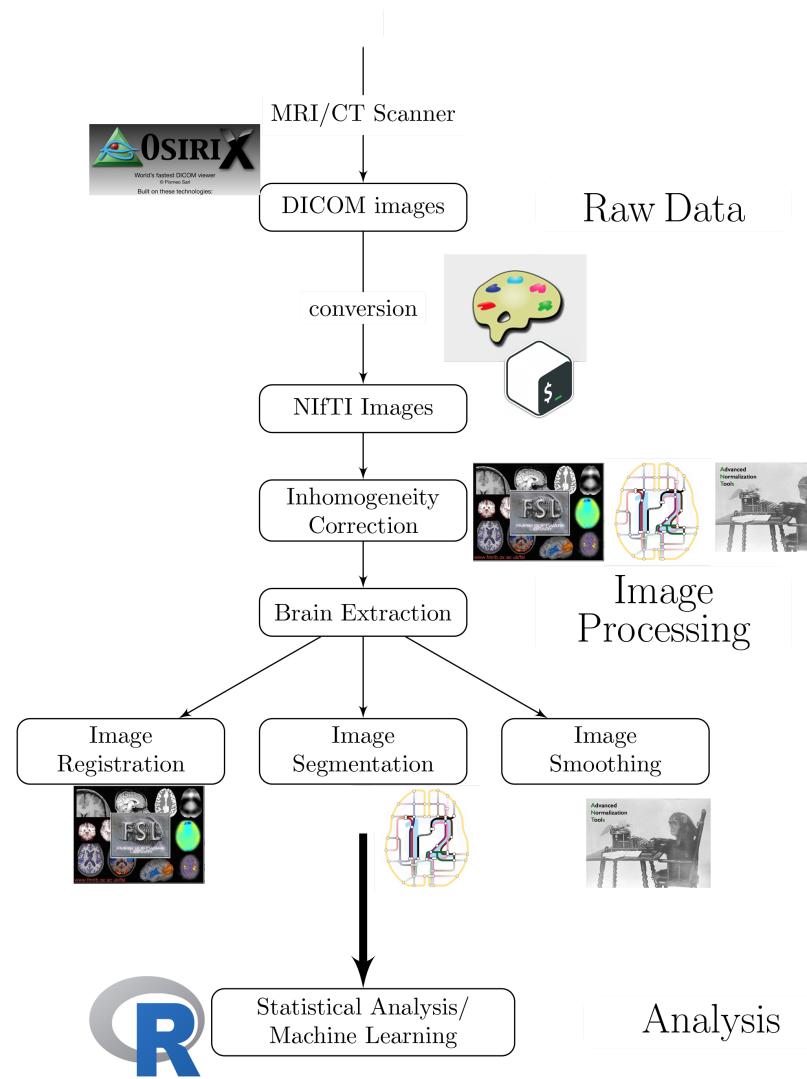
Overview of Work/Research

- Neuroimaging and R (Neuroconductor Project)
- R Package Development
- Segmentation/Classification of:
 - White Matter Lesions in Multiple Sclerosis
 - Brain vs. Skull (CT)
 - Brain Hemorrhage/Stroke (CT)

Brain Image Processing in R

Workflow for an Analysis

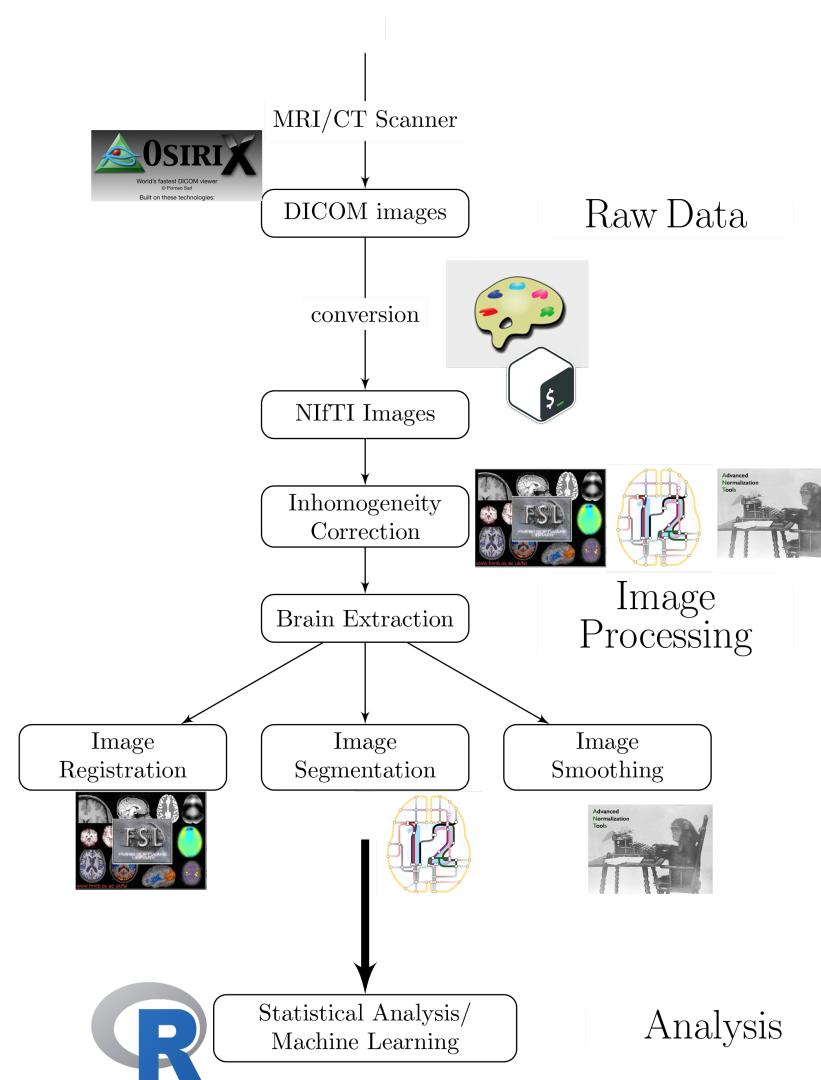
- bash 
- FSL 
- ANTs 
- MRIcroGL 
- OsiriX 
- SPM 12 



Workflow for an Analysis

Multiple pieces of software used

- all different syntax



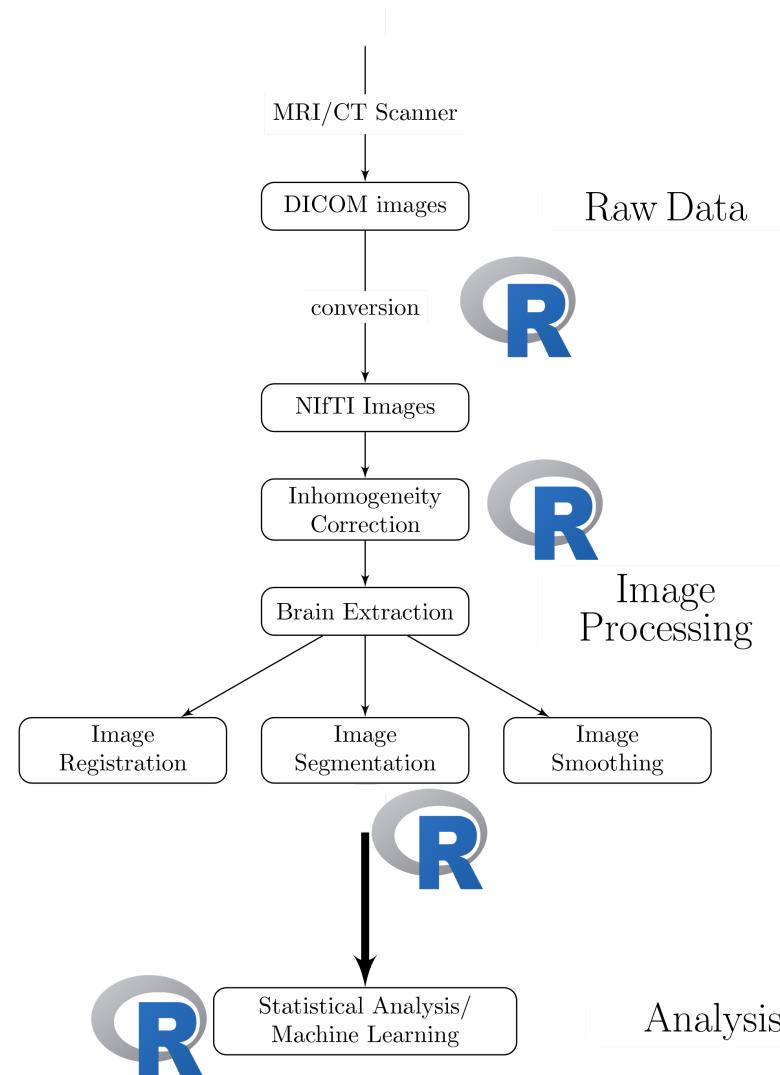
Goal:

Lower the bar to entry

- all R code
 - pipeline tool
 - “native” R code

Complete pipeline

- preprocessing and analysis



What did medical imaging in R have?

CRAN Task View: Medical Image Analysis

Maintainer: Brandon Whitcher

Contact: bwhitcher at gmail.com

Version: 2016-12-30

URL: <https://CRAN.R-project.org/view=MedicalImaging>

Data Input/Output

DICOM

The industry standard format, for data coming off a clinical imaging device, is [DICOM](#) (Digital Imaging and Communications in Medicine). The DICOM "standard" is very broad and very complicated. Roughly speaking each DICOM-compliant file is a collection of fields organized into two four-byte sequences (group,element) that are represented as hexadecimal numbers and form a *tag*. The (group,element) combination announces what type of information is coming next. There is no fixed number of bytes for a DICOM header. The final (group,element) tag should be the "data" tag (7FE0,0010), such that all subsequent information is related to the image(s).

- The packages [oro.dicom](#), [divest](#), [fmri](#) and [tractor.base](#) (part of the [tractor](#) project) provide R functions that read DICOM files and facilitate their conversion to ANALYZE or NIfTI format.

ANALYZE and NIfTI

Although the industry standard for medical imaging data is DICOM, another format has come to be heavily used in the image analysis community. The [ANALYZE](#) format was originally developed in conjunction with an image processing system (of the same name) at the Mayo Foundation. An Anlayze (7.5) format image is comprised of two files, the "hdr" and "img" files, that contain information about the acquisition and the acquisition itself, respectively. A more recent adaption of this format is known as [NIfTI-1](#) and is a product of the Data Format Working Group (DFWG) from the Neuroimaging Informatics Technology Initiative (NIfTI). The NIfTI-1 data format is almost identical to the ANALYZE format, but offers a few improvements: merging of the header and image information into one file (.nii), re-organization of the 348-byte fixed header into more relevant categories and the possibility of extending the header information.

- The packages [RNifti](#), [AnalyzeFMRI](#), [fmri](#), [tractor.base](#) (part of the [tractor](#) project), [oro.nifti](#), and [neuroim](#) all provide functions that read/write ANALYZE and NIfTI files.

Magnetic Resonance Imaging (MRI)

Diffusion Tensor Imaging (DTI)

- The R package [dti](#) provides structural adaptive smoothing methods for the analysis of diffusion weighted data in the context of the DTI model. Due to its edge preserving properties these smoothing methods are capable of reducing noise without compromizing significant structures (e.g., fibre tracts). The package also provides functions for DTI data processing from input,

Bioinformatics Repository: Bioconductor



- centralized bioinformatics/genomics packages
- large community/number of packages (> 1300)
- published tutorials and workflows
- additional requirements to CRAN (e.g. packages need vignettes)

Processing math: 100%

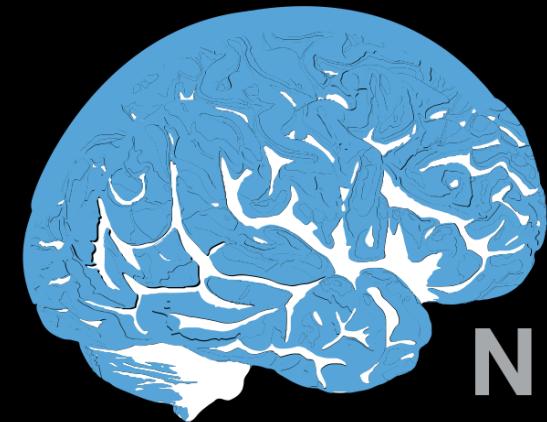
Bioinformatics Repository: Bioconductor

Autocomplete biocViews search:

Show All ↴ entries

Search table:

Package	Maintainer	Title
a4	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Umbrella Package
a4Base	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Base Package
a4Classif	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Classification Package
a4Core	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Core Package
a4Preproc	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Preprocessing Package
a4Reporting	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Reporting Package
ABAEnrichment	Steffi Grote	Gene expression enrichment in human brain regions
ABarray	Yongming Andrew Sun	Microarray QA and statistical data analysis for Applied Biosystems Genome Survey Microrarray (AB1700) gene expression data.
ABSSeq	Wentao Yang	ABSSeq: a new RNA-Seq analysis method based on modelling absolute expression differences
acde	Juan Pablo Acosta	Artificial Components Detection of Differentially Expressed Genes
aCGH	Peter Dimitrov	Classes and functions for Array Comparative Genomic Hybridization data.
ACME	Sean Davis	Algorithms for Calculating Microarray Enrichment (ACME)
ADaCGH2	Ramon Diaz- Uriarte	Analysis of big data from aCGH experiments using parallel computing and ff objects
adSplit	Claudio Lottaz	Annotation-Driven Clustering
	Kasper Daniel	



NEUROCONDUCTOR

An R Platform for
Medical Imaging Analysis
(Muschelli et al. 2018)

Home

About Neuroconductor

Neuroconductor is an open-source platform for rapid testing and dissemination of reproducible computational imaging software. The goals of the project are to:

- provide a centralized repository of R software dedicated to image analysis;
- disseminate quickly software updates;
- educate a large, diverse community of scientists using detailed tutorials and short courses;
- ensure quality via automatic and manual quality controls; and
- promote reproducibility of image data analysis.

Based on the programming language [R](#), Neuroconductor starts with 69 interoperable packages that cover multiple areas of imaging including visualization, data processing and storage, and statistical inference. Neuroconductor accepts new R package submissions, which are subject to a formal review and continuous automated testing.

Install

- [Install Neuroconductor](#)
- [List All Packages](#)
- [Package Dependency Graph](#)
- [Get Support](#)
- [Install R](#)

Use

- [My packages \(38\)](#)
- [Browse releases](#)

Learn

- [Courses](#)
- [Tutorials](#)
- [FAQ](#)
- [Forum](#)

Develop

- [Changes to Your Package](#)
- [List Pending Packages](#)
- [Create Neuroconductor Account](#)

Funding Acknowledgements

Neuroconductor and its maintainers have been partially supported by the R01 grant NS060910 from the National Institute of Neurological Disorders and Stroke at the National Institutes of Health (NINDS/NIH). We are also grateful to the Department of Biostatistics at Johns Hopkins University for providing support during the early stages of development.

Tutorials and Workflows

The screenshot shows a web browser window with the URL johnmuschelli.com/neuroc/ in the address bar. The page content is organized into several sections:

- Navigation Sidebar:** On the left, there is a vertical sidebar with the following items:
 - Guides for Developers
 - Frequently Asked Questions (FAQ)
 - Installation Guides
 - Required Readings
 - Data** (this item is highlighted with a blue background)
 - General Tutorials
 - Disease-specific Tutorials
- Data Section:** This section contains two numbered links:
 1. [Downloading Human Connectome Project Data](#)
 2. [Downloading Functional Connectomes Projects](#)
- General Tutorials Section:** This section is currently empty.
- Structural Imaging Section:** This section contains five numbered links:
 1. [Brain Extraction](#)
 2. [Tissue-Class Segmentation](#)
 3. [Labeling Brain Structures](#)
 4. [Processing Within-Visit Structural MRI](#)
 5. [Cortical Thickness Estimation \(NOT DONE\)](#)
- fMRI Section:** This section contains six numbered links:
 4. [Resting-State fMRI analysis using ANTsR](#)
 5. [Resting-State fMRI analysis using fslr](#)
 6. [Resting-State fMRI analysis using spm12r](#)
- DTI Section:** This section contains three numbered links:
 7. [DTI analysis with fslr](#)
 8. [DTI analysis with Camino](#)
 9. [DTI analysis with Camino with HCP data](#)

Authored R Packages:

- **fslr**
(Muschelli, John, et al. "fslr: Connecting the FSL Software with R." *R JOURNAL* 7.1 (2015): 163-175.)
- **brainR**
(Muschelli, John, Elizabeth Sweeney, and Ciprian Crainiceanu. "brainR: Interactive 3 and 4D Images of High Resolution Neuroimage Data." *R JOURNAL* 6.1 (2014): 42-48.)
- **ichseg**
Muschelli, John, et al. "PItcHPERFeCT: Primary intracranial hemorrhage probability estimation using random forests on CT." *NeuroImage: Clinical* 14 (2017): 379-390.
- **extrantsr**
- **dcm2niir**
- **matlabr**
- **spm12r**
- **freesurfer**
- **itksnapr**
- **stapler**
- **gifti**
- **cifti**
- **papayar**
- **diffR**
- **gcite**
- **rscopus**
- **fedreporter**
- **glassdoor**

Number of Downloads (from `cranlogs`)

Lesion Segmentation of MS

Terminology: Neuroimaging to Data/Statistics

- Segmentation \Leftrightarrow classification
- Image \Leftrightarrow 3-dimensional array
- Mask/Region of Interest \Leftrightarrow binary (0/1) image
- Registration \Leftrightarrow Spatial Normalization/Standarization
 - “Lining up” Brains

Public Dataset with Lesion Segmentation

- “A novel public MR image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus” (Lesjak et al. 2018)
 - Data Published at <http://lit.fe.uni-lj.si/tools.php?lang=eng>
 - 30 subjects with MRI (3T Siemens Trio)
- Manually segmented by 3 expert raters
- Creative-Commons Attribution (CC-BY)
- All analysis and data located: https://github.com/muschelli2/open_ms_data

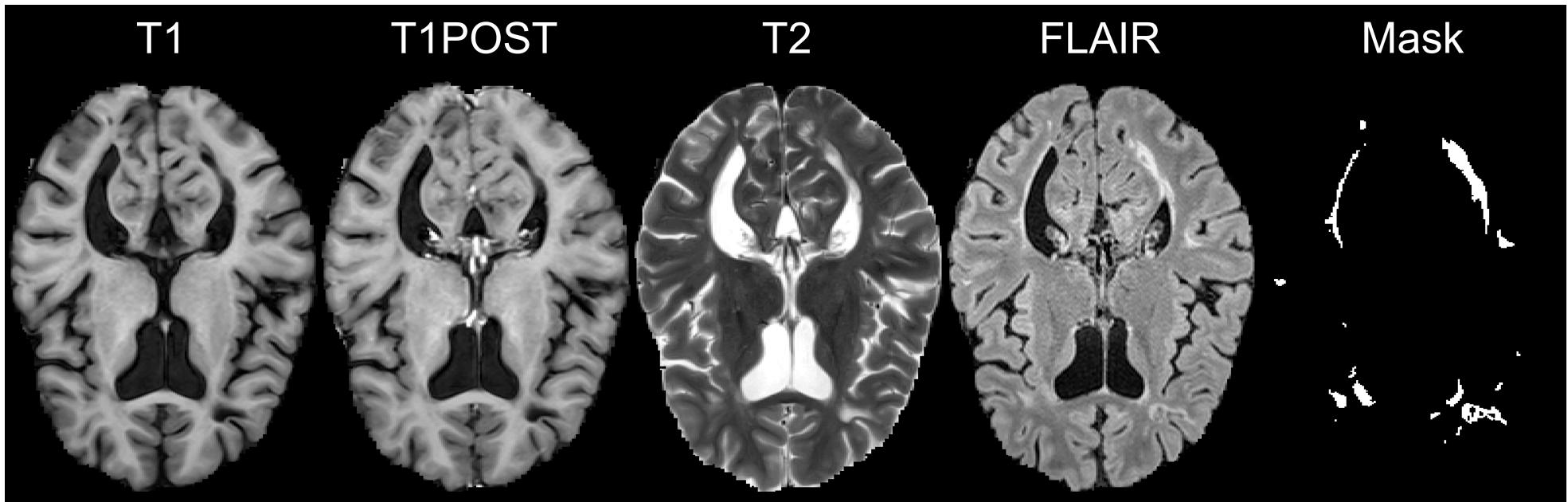
Demographic Data

- On many different therapies (9 no therapy), age IQR: 33 - 42, EDSS IQR: 1.5 - 4

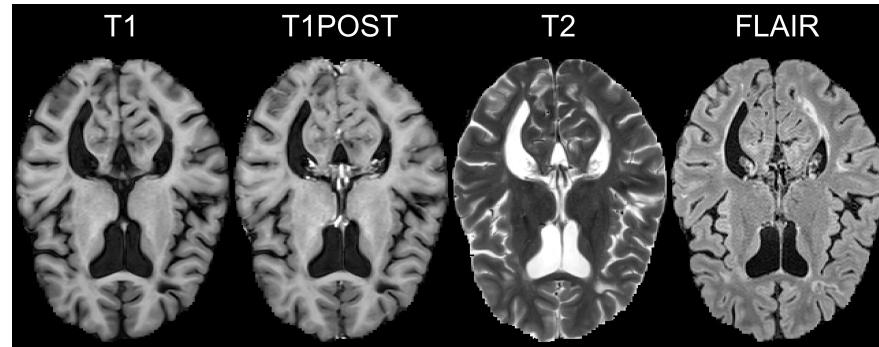
Variable	Overall
n	30
Age (mean (sd))	39.27 (10.12)
sex = M (%)	7 (23.3)
EDSS (mean (sd))	2.61 (1.88)
Lesion_Volume (mean (sd))	17.40 (16.13)
MS_Subtype (%)	
Clinically Isolated Syndrome	2 (6.7)
Progressive-relapsing	1 (3.3)
Relapsing-remitting	24 (80.0)
Secondary-progressive	2 (6.7)
Unspecified	1 (3.3)

Imaging Data

- 2D T1 (TR=2000ms, TE=20ms, TI=800ms) and after gadolinium
- 2D T2 (TR=6000ms, TE=120ms), 3D FLAIR (TR=5000ms, TE=392ms, TI=1800 ms)
 - Fluid attenuated inversion recovery - reduce signal of fluids
- All had flip angle of 120°



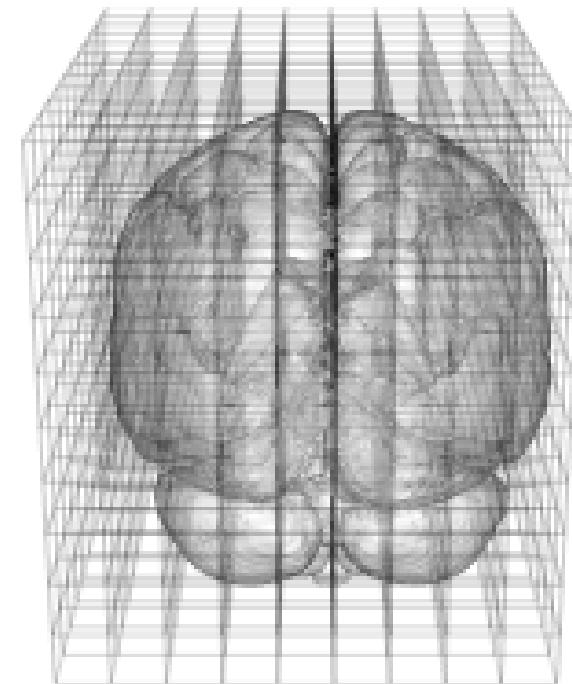
Project Goal



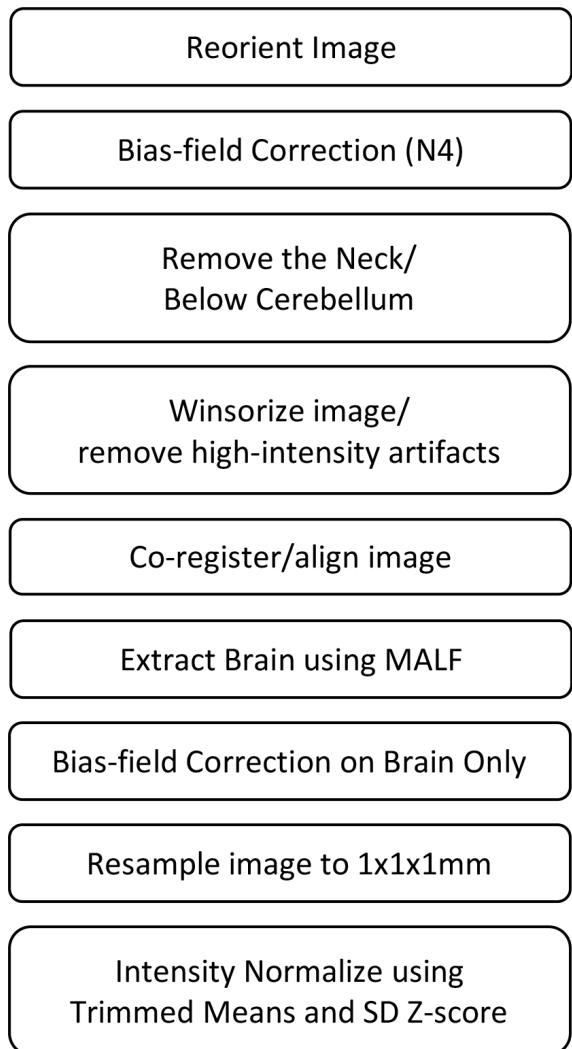
Predictors



Image Representation: voxels (3D pixels)



Step 1: Image Processing: Workflow

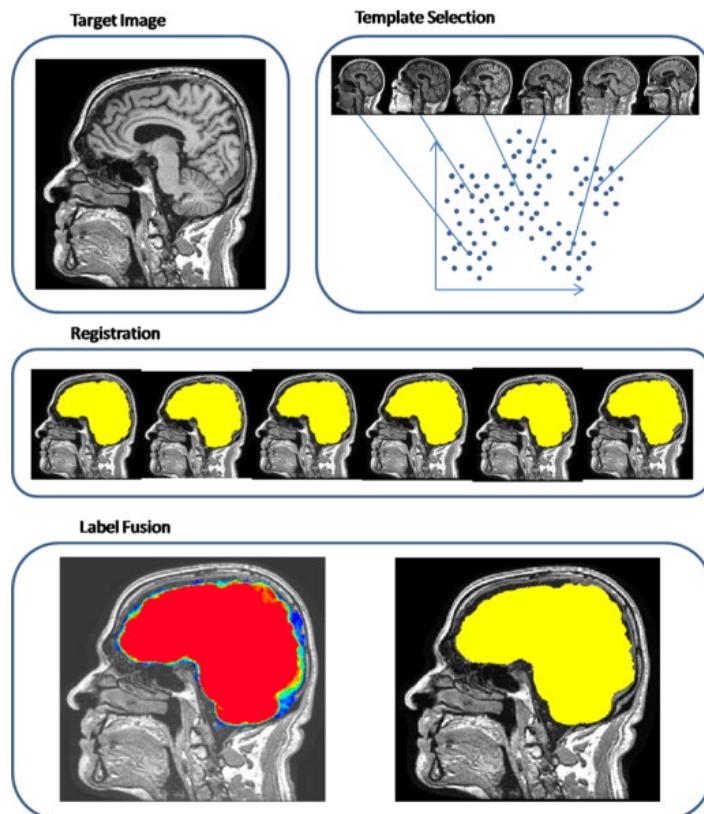


The N4 (Tustison et al. 2010) EM-style model assumed is:
 $\log(x(v)) = \log(u(v)) + \log(f(v))$

- x: given image
- u: uncorrupted image
- f: bias field
- v: location in the image

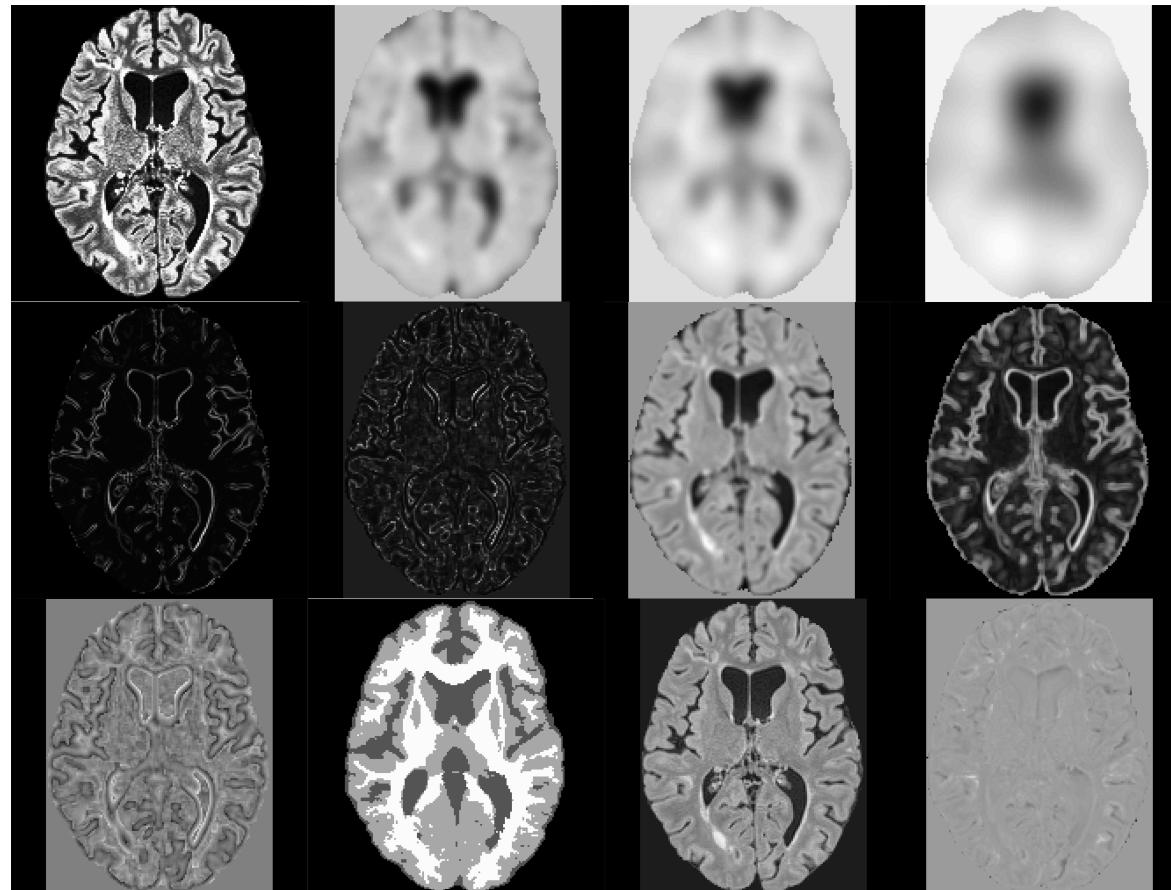
Step 1: Image Processing: MALF

Figure from Multi-Atlas Skull Stripping method paper (Doshi et al. 2013):



- Register templates to an image using the T1 for that subject
- Apply transformation to the label/mask
- Average each voxel over all templates
 - there are “smarter” (e.g. weighted) ways

Step 2: Create Predictors for each Sequence



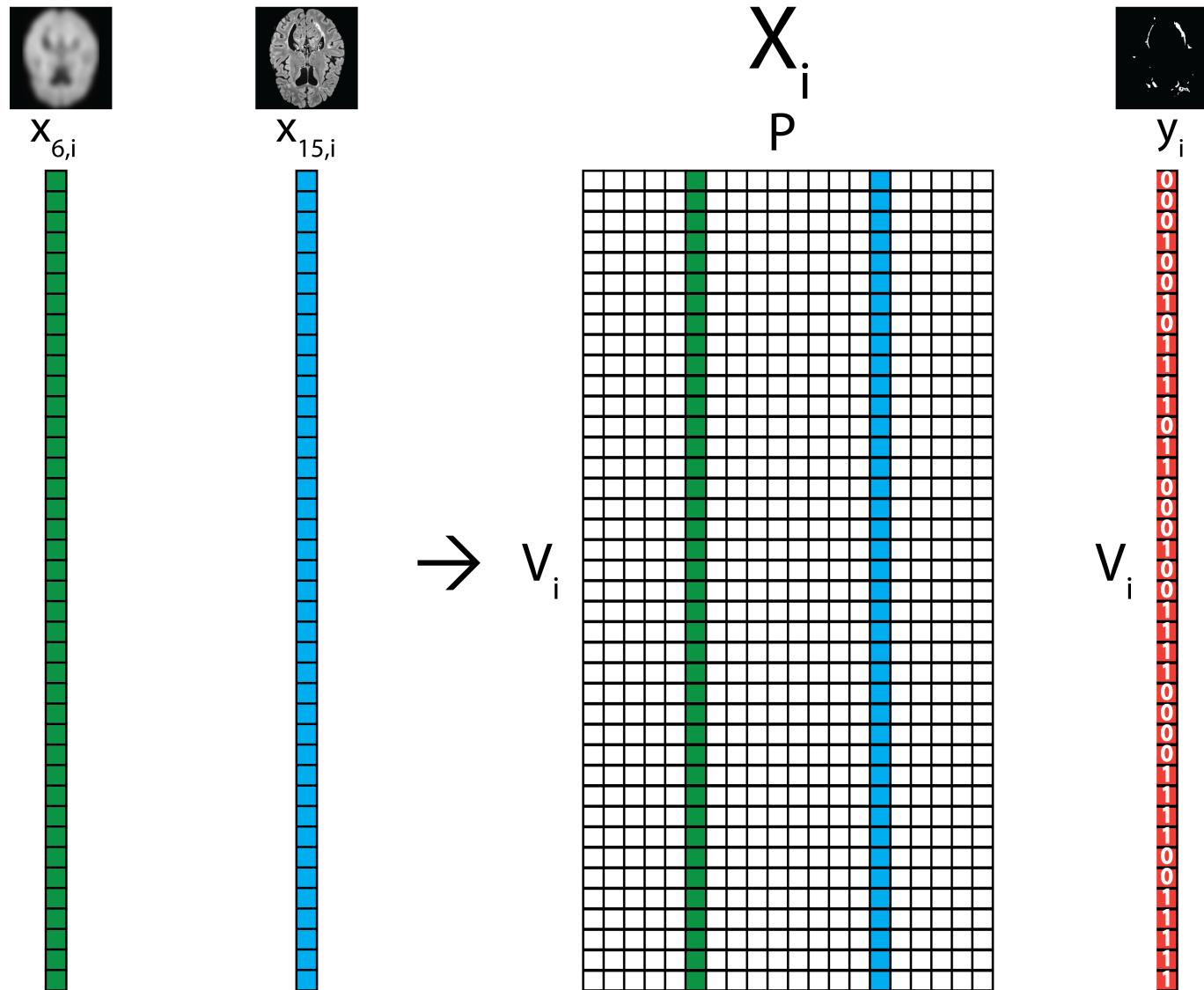
- Predictors created with intensity-normalized data
 - Quantile images, smoothers, local moments
- Tissue class probability with local moments:
MALF and FAST (Zhang, Brady, and Smith 2001)
- Z-score to a population template

A package to do all this: `smri.process`

- GitHub package ([muschelliij2/smri.process](https://github.com/muschelliij2/smri.process))

```
65 processed = smri_prenormalize(  
66   x = files,  
67   outdir = idf$proc_dir,  
68   gold_standard = gold_standard,  
69   gs_space = "FLAIR",  
70   reg_space = "FLAIR",  
71   malf_transform = "SyN",  
72   verbose = 2,  
73   outprefix = outprefix,  
74   probs = c(0, 0.995),  
75   num_templates = 15,  
76   force_registration = FALSE)  
77  
78 outprefix = file.path(  
79   idf$malf_dir,  
80   "T1")  
81  
82 all_resampled = seg_normalize(  
83   prenormalize = processed,  
84   template = "none",  
85   verbose = TRUE,  
86   force_registration = FALSE,  
87   outprefix = outprefix  
88 )  
89 normalized = all_resampled$normalized  
90  
91 pred = norm_predictors(  
92   normalized = normalized,  
93   normalization = "trimmed_z"  
94 )
```

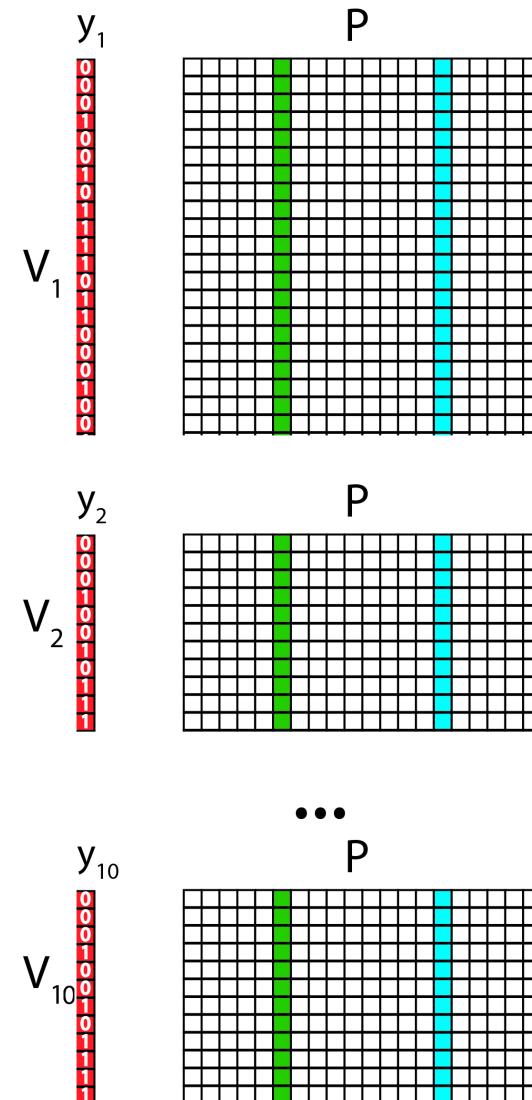
Data Structure for One Patient



Step 3: Aggregate Data

Training Data Structure

- Sample 10% of the voxels (save computation time)
- Stack together 14 randomly selected patients, stratified by age (over median) and volume
- Train model/classifier on this design matrix
- Smooth the probability map
- Test on 16 hold out



Step 4: Fit Models / Classifier

Let $y_i(v)$ be the presence / absence of lesion for voxel v from person i .

General model form: $P(Y_i(v) = 1) \propto f(X_i(v))$

- Previous work - **OASIS** (Sweeney et al. 2013):

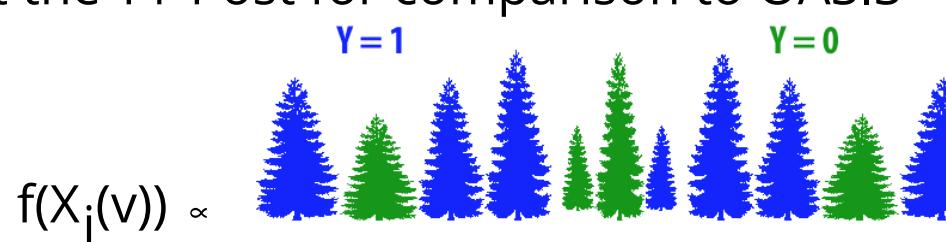
$$f(X_i(v)) = \exp \left\{ \beta_0 + \sum k x_k(v) \beta_k + x_k(v) \times x_{10,k} \beta_{10,k} + x_k(v) \times x_{20,k} \beta_{20,k} \right\}$$

$k \in \{T1, T2, FLAIR, PD\}$.

- With the original model w/o T1Post and a re-trained model

Models Fit on the Training Data

- 85 predictors were generated
- Random Forests (Wright and Ziegler 2017), (Breiman 2001)
 - With 5 fold cross-validation, default 500 trees, mtry: \sqrt{p}
 - With and without the T1-Post for comparison to OASIS



For each model (RF with and w/o T1Post and OASIS retrained or not)

- Estimate a probability cutoff on training data
- Predict on test data, assess performance acrosss all voxels in the brain

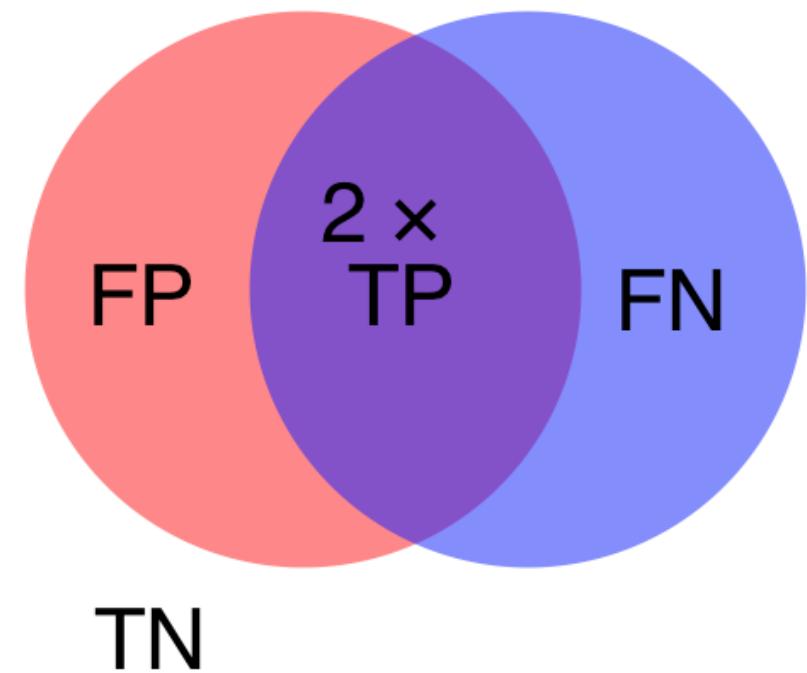
Assessing Performance

For each test scan, and over all test scans, we can calculate the following 2-by-2 table, where cells represent number of voxels and corresponding Venn diagram:

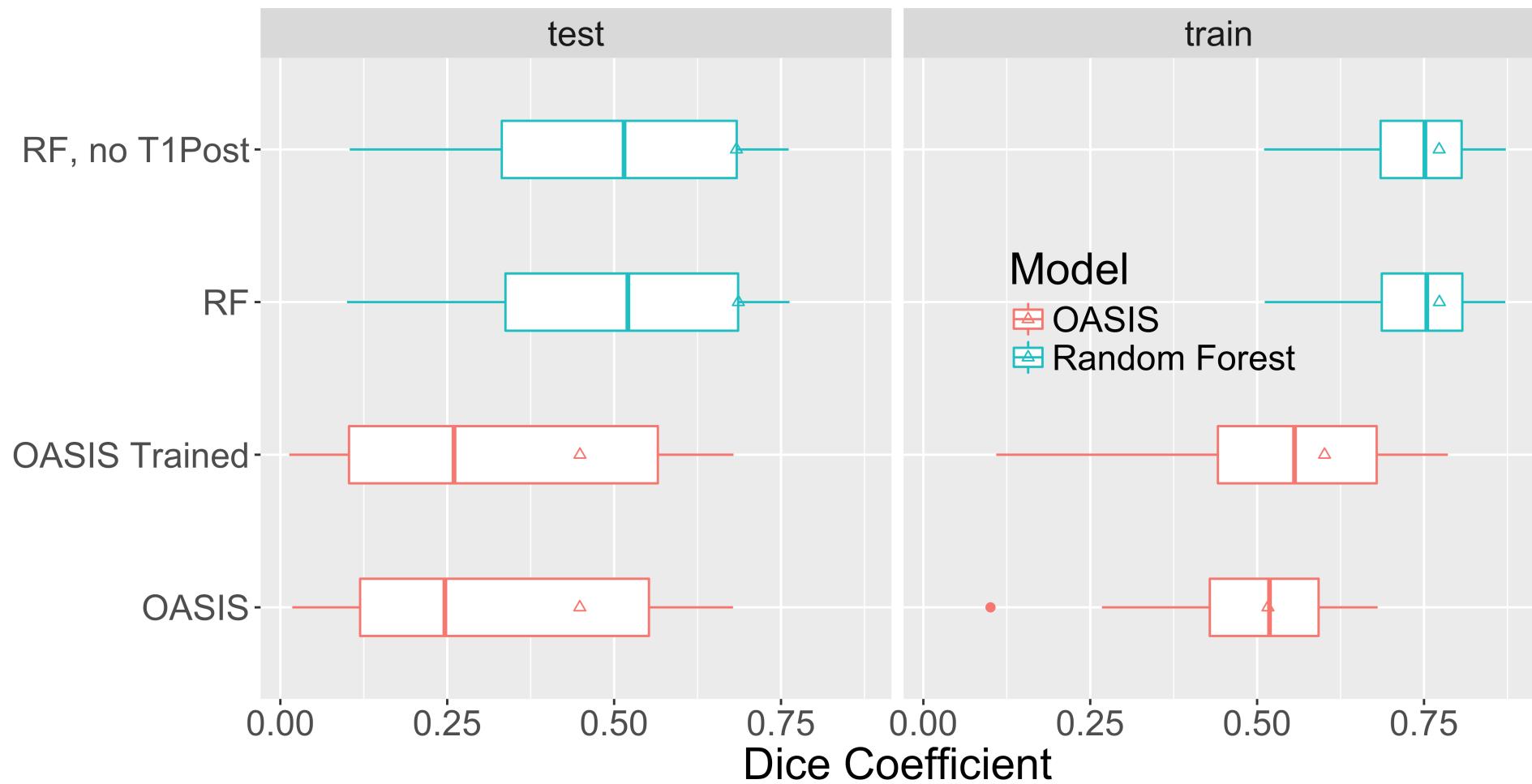
		Manual	
		0	1
Auto	0	TN	FN
	1	FP	TP

Dice Coeffiicent (Dice 1945):

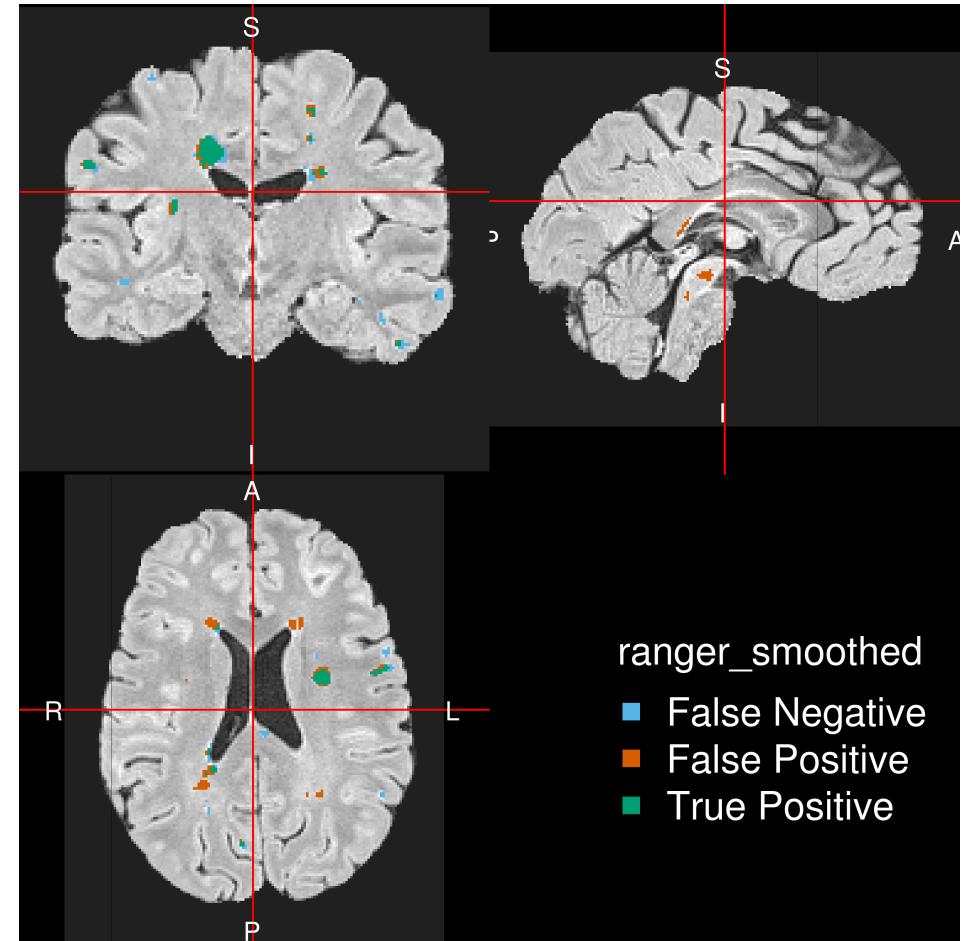
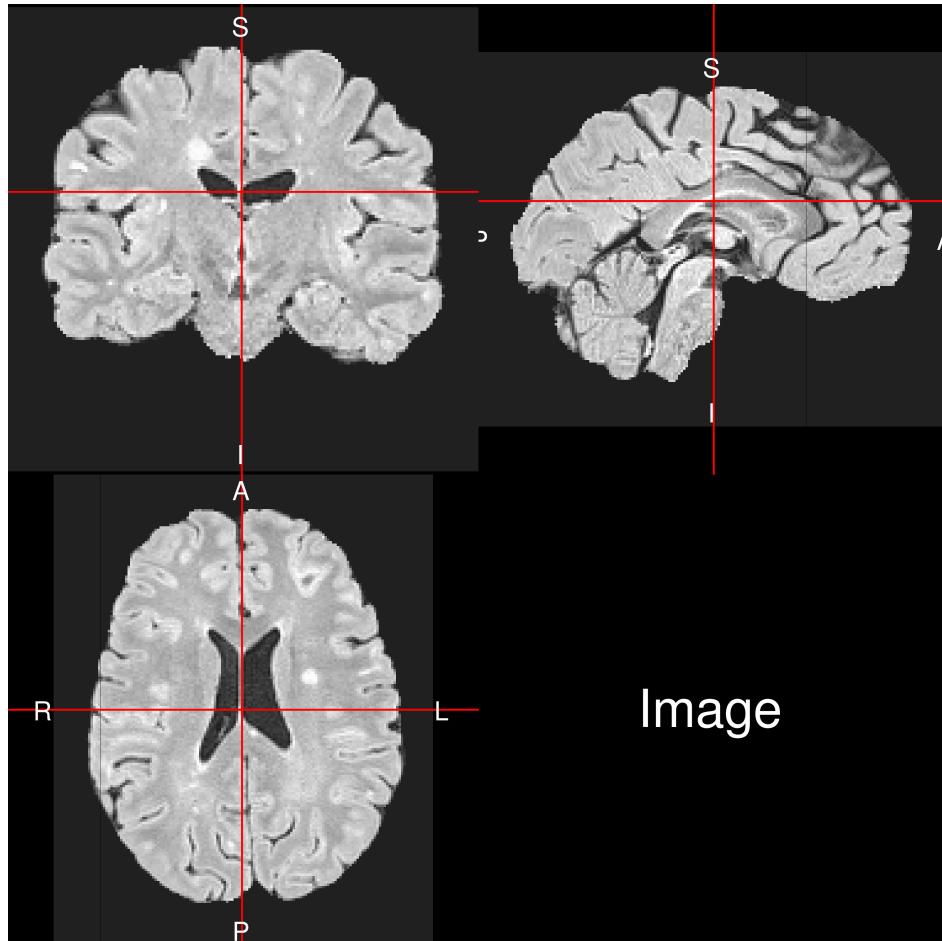
$$\text{Dice} = \frac{2 \times \text{TP}}{\text{TP} + \text{FN} + \text{FP}}$$



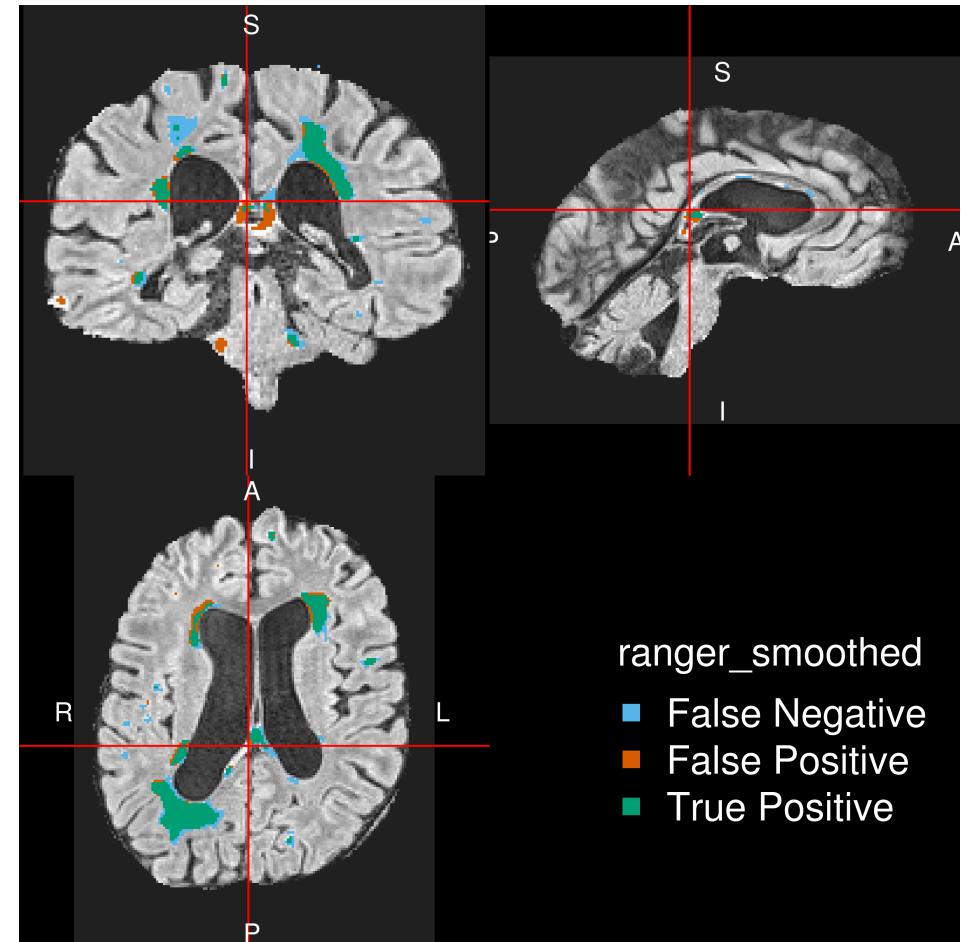
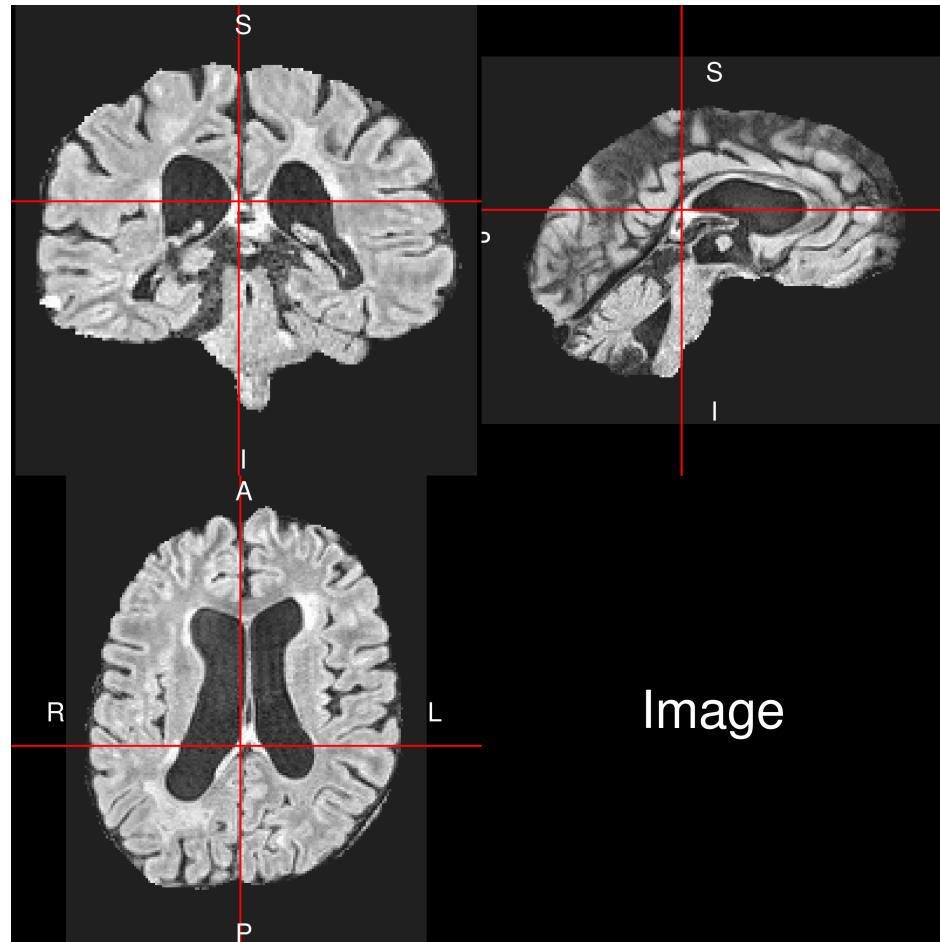
Dice Results (Triangle is population Dice)



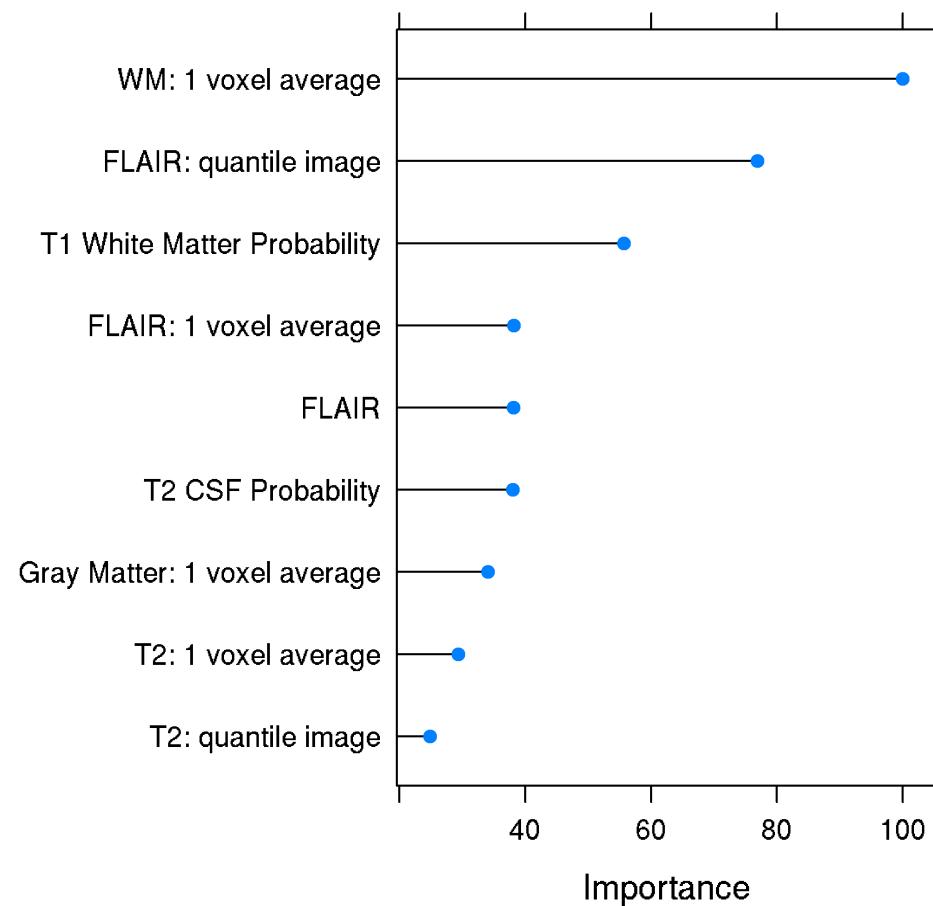
Patient with Median DSI (0.63) in Test



Patient with High DSIs (0.73) in Test

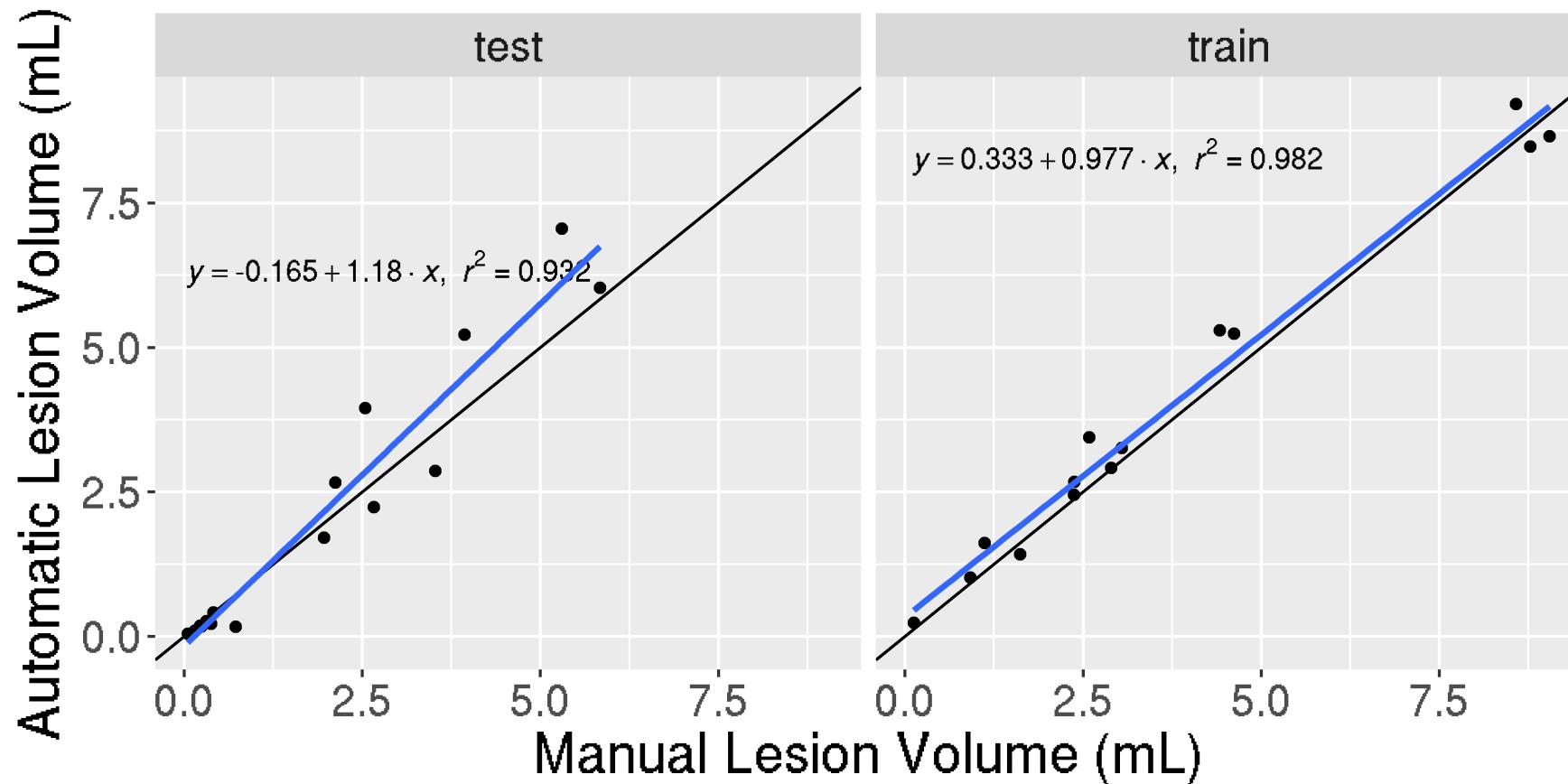


Variable Importance

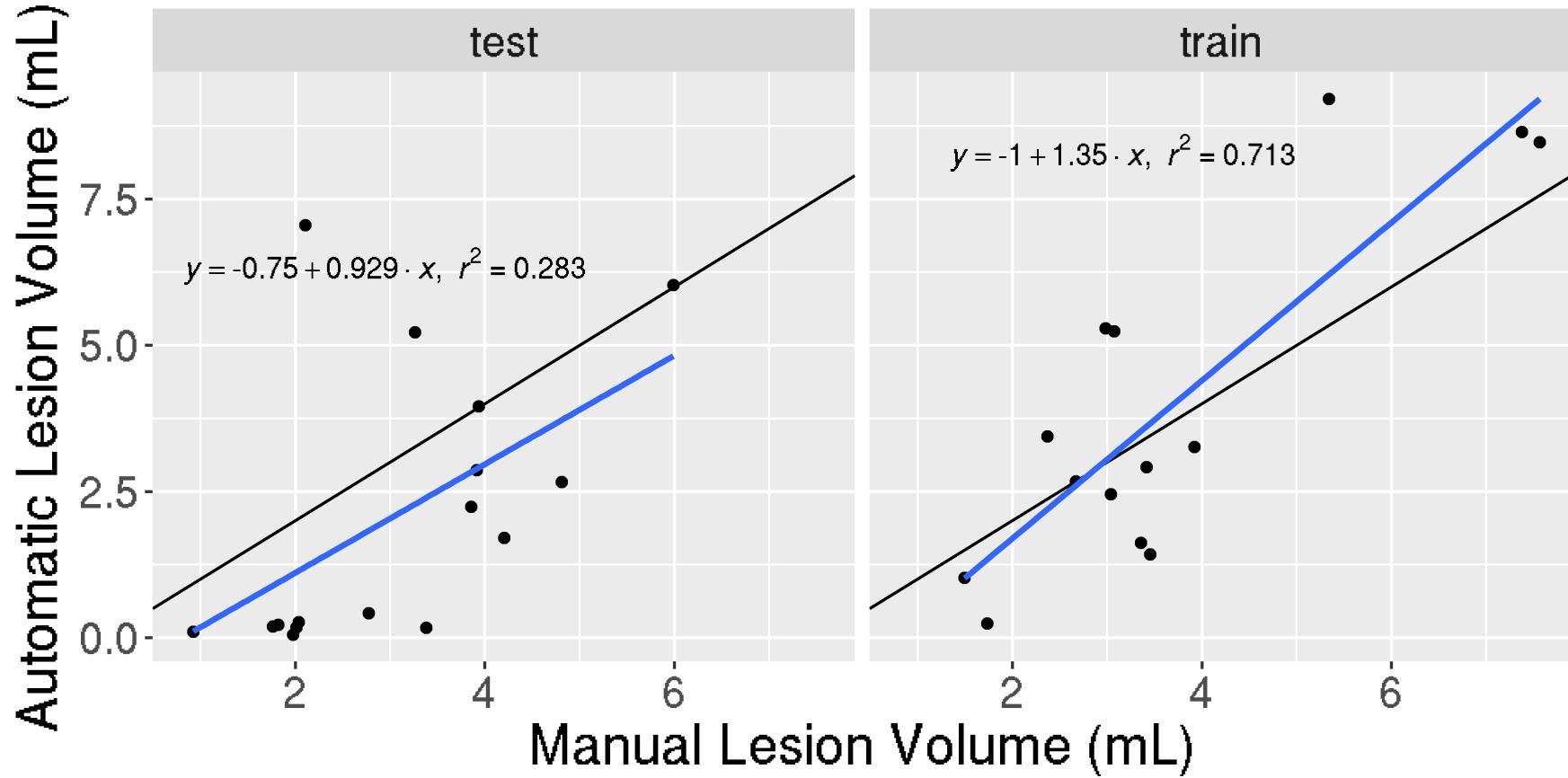


- Top predictors in RF model
- T1Post not in there
- Tissue segmentations are important predictors
 - FLAIR as well

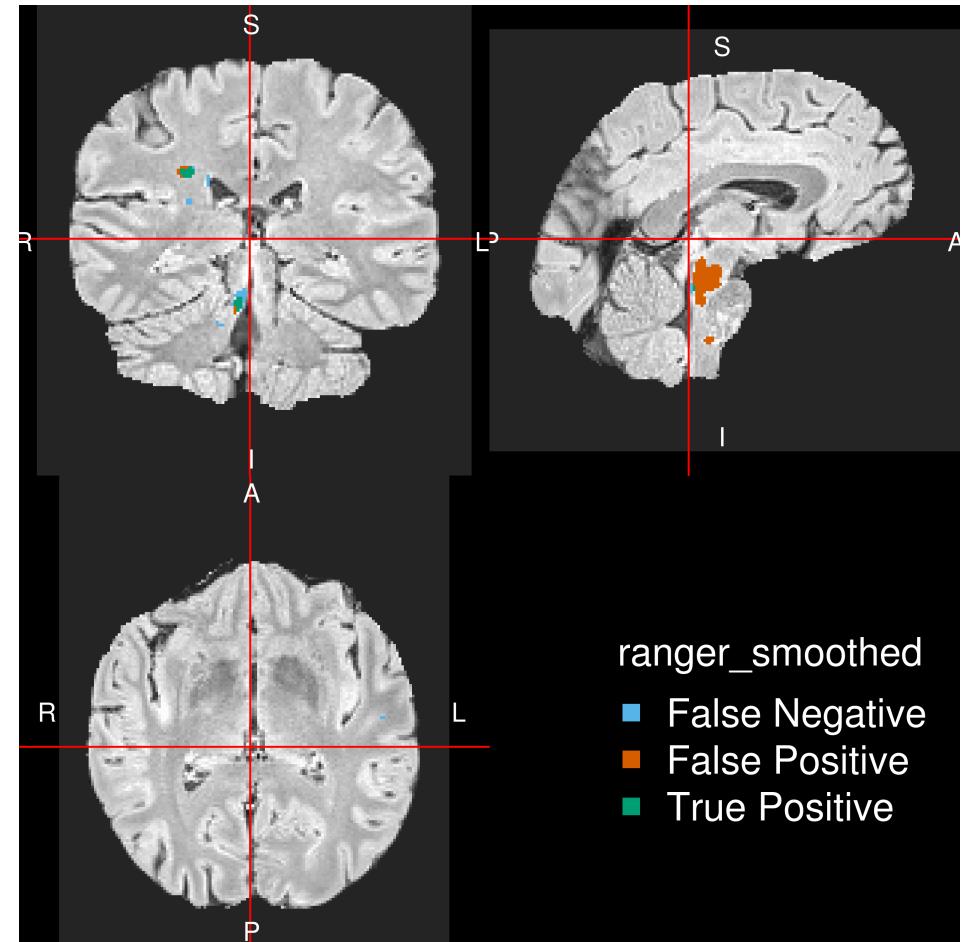
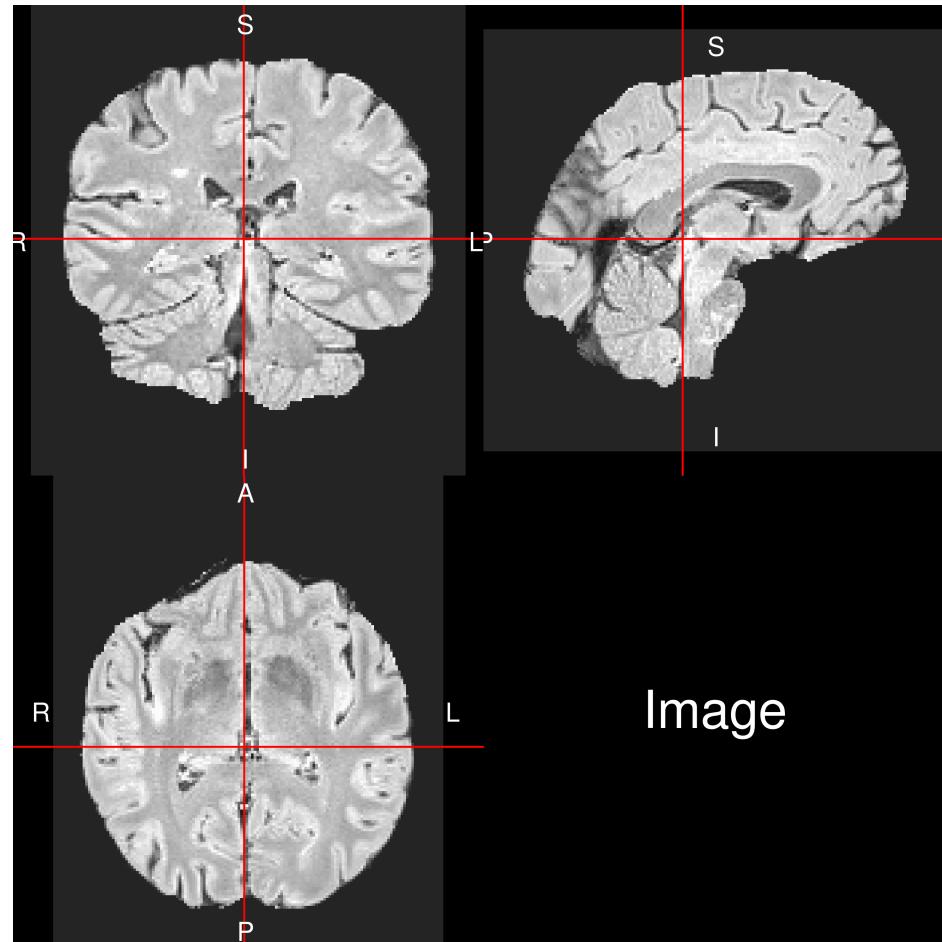
RF Predicted Volume Estimates True Volume



OASIS: not so much



Brain Stem Lesions Estimated



Conclusions of Lesion Analyses

- We can segment MS lesions reasonably well
- Better models with larger samples
- Needs to be more stable/accurate for a biomarker
 - Location may also be relevant and not taken into account
 - Is the brain stem an area we should focus on or remove from assessment?

Next Steps/Questions

- Run new processing the 131 patients from OASIS paper
- Gray matter injury estimation
- Is EDSS the clinical score we should be correlating with?
- “Black hole” lesions using the T1-post image, these may show “active” lesions

Thank You

- Breiman, Leo. 2001. "Random Forests." 45 (1). Springer:5–32.
- Dice, Lee R. 1945. "Measures of the Amount of Ecologic Association Between Species." 26 (3):297–302. <http://www.jstor.org/stable/1932409>.
- Doshi, Jimit, Guray Erus, Yangming Ou, Bilwaj Gaonkar, and Christos Davatzikos. 2013. "Multi-Atlas Skull-Stripping." 20 (12). Elsevier:1566–76.
- Lesjak, Žiga, Alfiia Galimzianova, Aleš Koren, Matej Lukin, Franjo Pernuš, Boštjan Likar, and Žiga Špiclin. 2018. "A Novel Public MR Image Dataset of Multiple Sclerosis Patients with Lesion Segmentations Based on Multi-Rater Consensus." 16 (1). Springer:51–63.
- Muschelli, John, Adrian Gherman, Jean-Philippe Fortin, Brian Avants, Brandon Whitcher, Jonathan D Clayden, Brian S Caffo, and Ciprian M Crainiceanu. 2018. "Neuroconductor: An R Platform for Medical Imaging Analysis." .
- Sweeney, Elizabeth M, Russell T Shinohara, Navid Shiee, Farrah J Mateen, Avni A Chudgar, Jennifer L Cuzzocreo, Peter A Calabresi, Dzung L Pham, Daniel S Reich, and Ciprian M Crainiceanu. 2013. "OASIS Is Automated Statistical Inference for