Interpretable machine learning approach to identifying white matter brain differences between healthy aging and Alzheimer's onset through diffusion tensor imaging

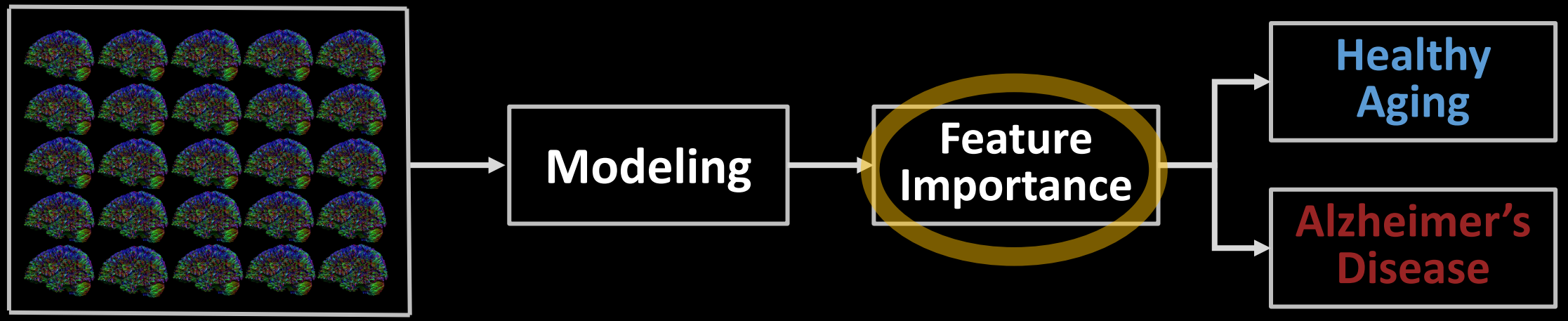
Rastko Stojsin

The Problem

- The pathology of Alzheimer's (neurodegenerative disease) is largely associated with loss of neuronal connections in the brain

- Diffusion tensor imaging helps us estimate the connectivity within white matter via neural tracts

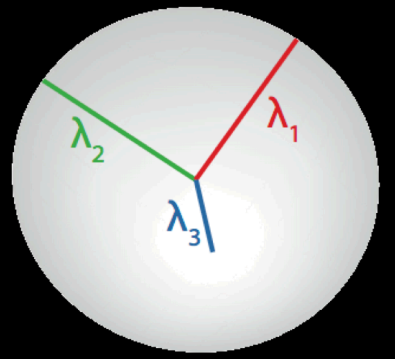
- Identifying the structural differences between a healthy aging brain and one with Alzheimer's could help in better understanding the pathology of Alzheimer's



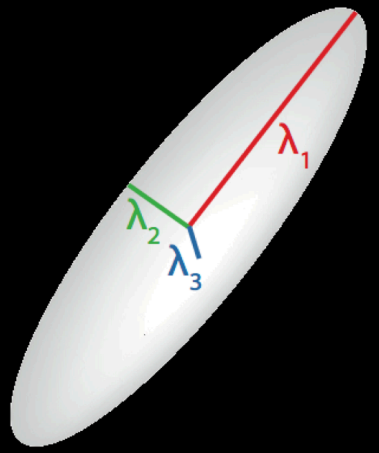
Diffusion Tensor Imaging (DTI)

- MRI technique that uses the anisotropic diffusion of water molecules to estimate the white matter structure of the brain.

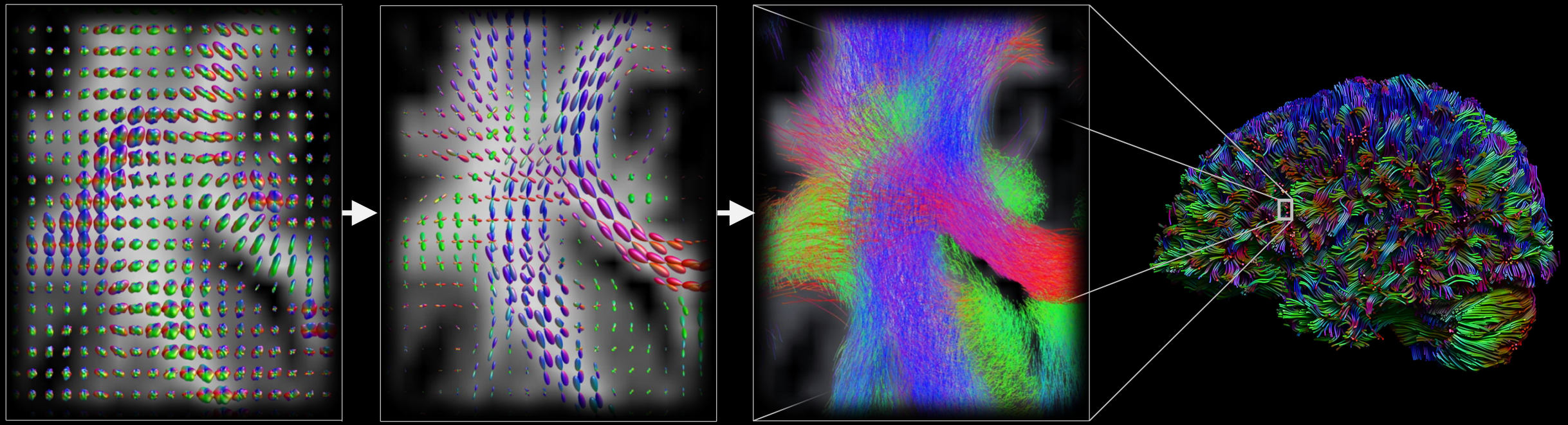
- Water molecules tend to diffuse more freely in the direction of neural tracts rather than across them



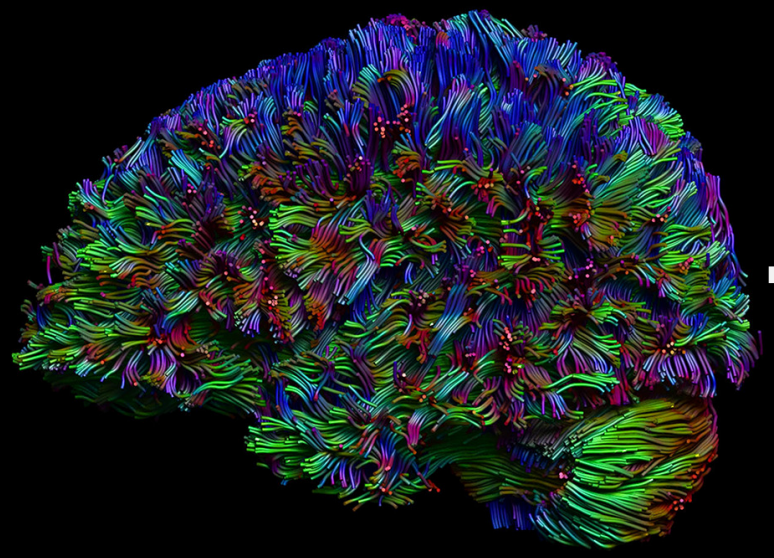
Isotropic



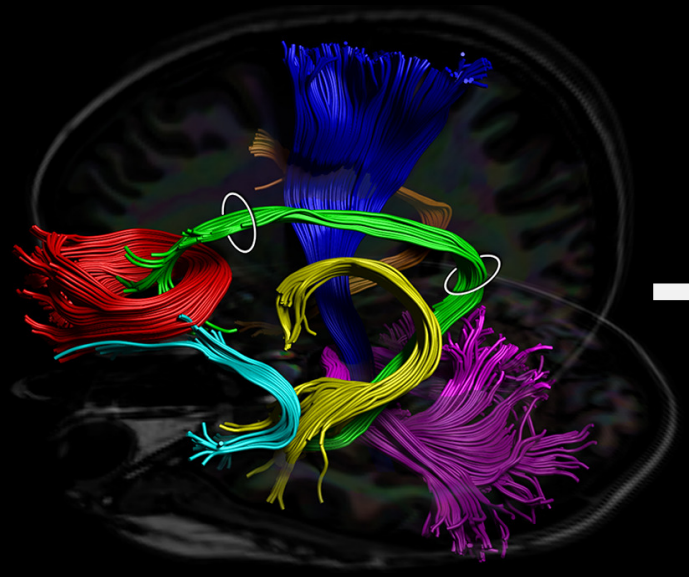
Anisotropic



Brain



Bundles (x49)



Derived Features (x31)

number of tracts
mean length(mm)
span(mm)
curl
elongation
diameter(mm)
volume(mm^3)
trunk volume(mm^3)
branch volume(mm^3)
total surface area(mm^2)
total radius of end regions(mm)
total area of end regions(mm^2)
irregularity
area of end region 1(mm^2)
radius of end region 1(mm)
irregularity of end region 1
area of end region 2(mm^2)
radius of end region 2(mm)
irregularity of end region 2
qa, nqa, dti_fa, md, ad, rd, gfa
iso, rdi, nrdi02L, nrdi04L, nrdi06L

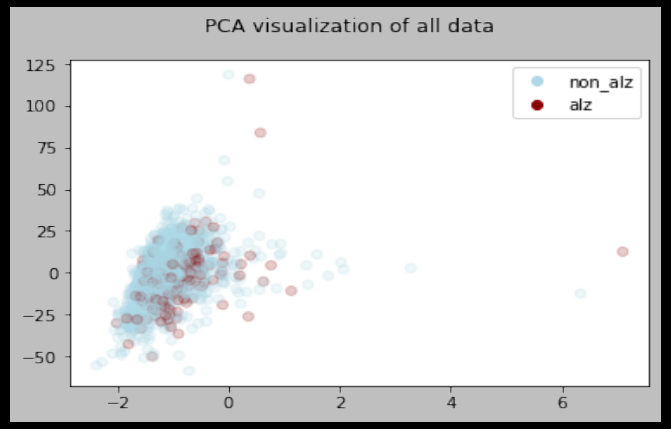
Each brain has 49 bundles
Each bundle has 31 derived features
Each brain has 1521 derived features total (49*31)

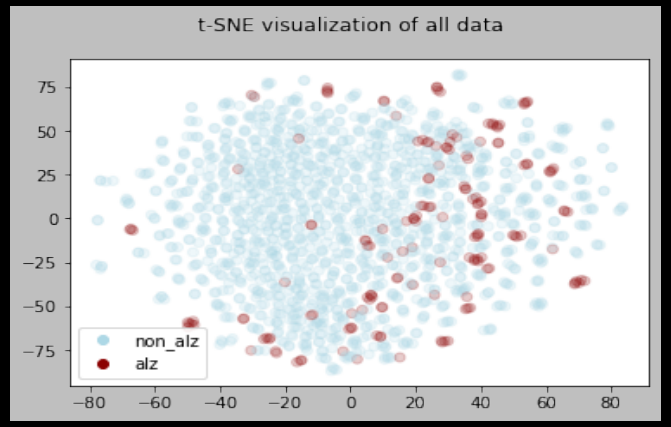
Imbalanced Data Issue

1613 - Total number of brain scans

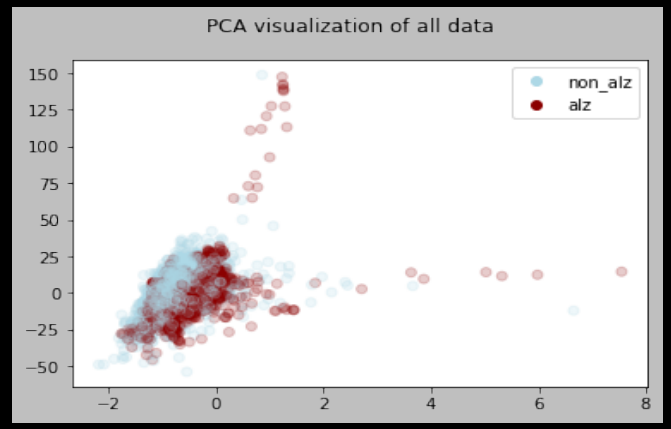
1450 - Scans of non-Alzheimer’s Patients

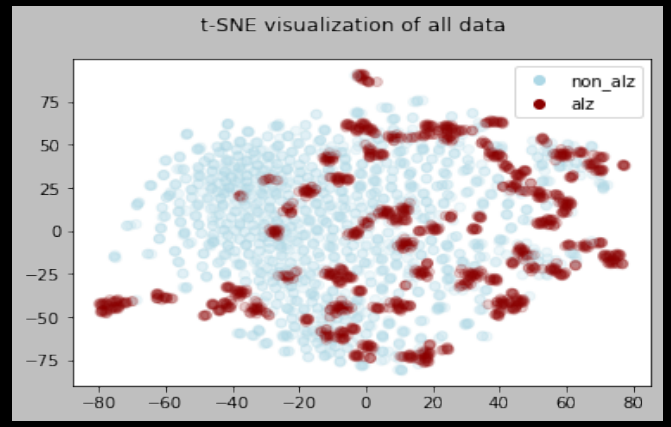
163 - Scans of Patients with Alzheimer's (now/in future scans)





Synthetic Minority Oversampling Technique (SMOTE)





The Process

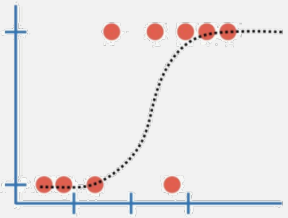
data support

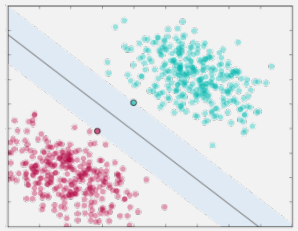
derived features

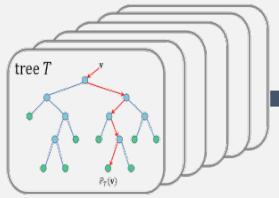
SMOTE

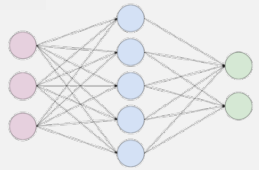
no up-sample

modeling









...

evaluation & inference

Feature Importance

Non-Alzheimer's

Alzheimer's

The Models

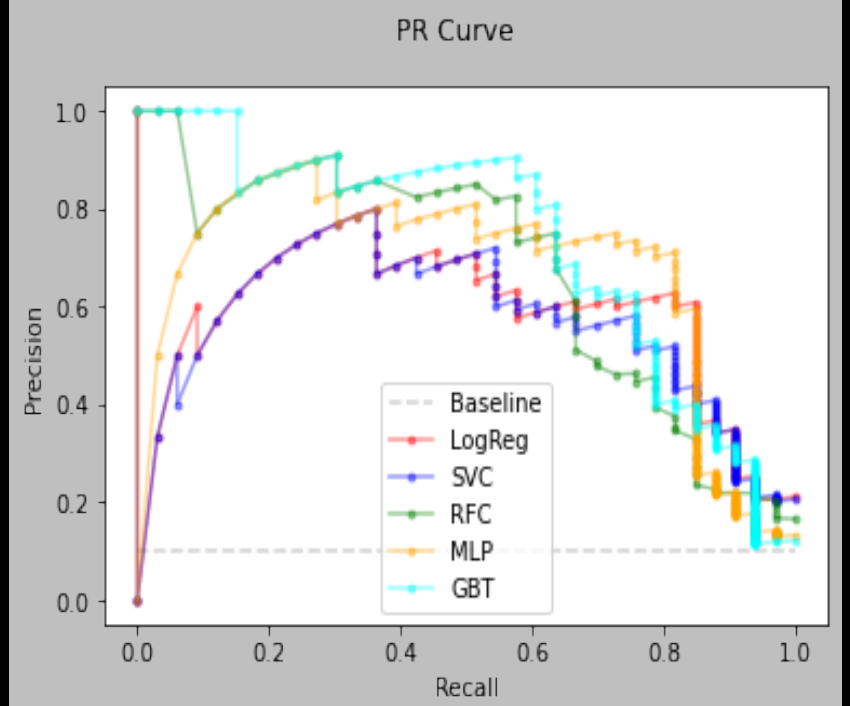
Logistic Regression

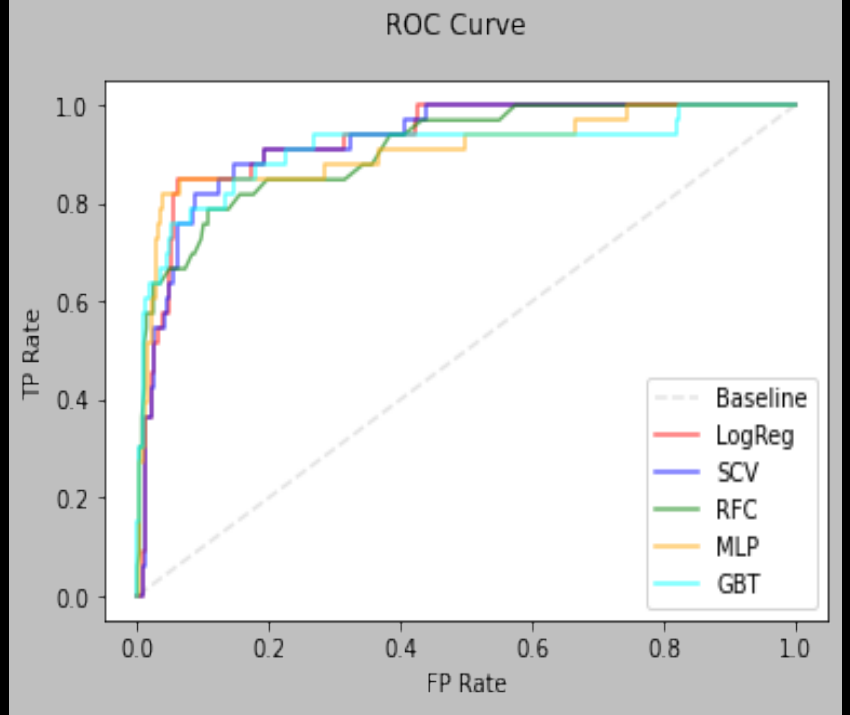
Support Vector Classification

Random Forrest

Multi-Layer Perceptron

Gradient Boosted Trees





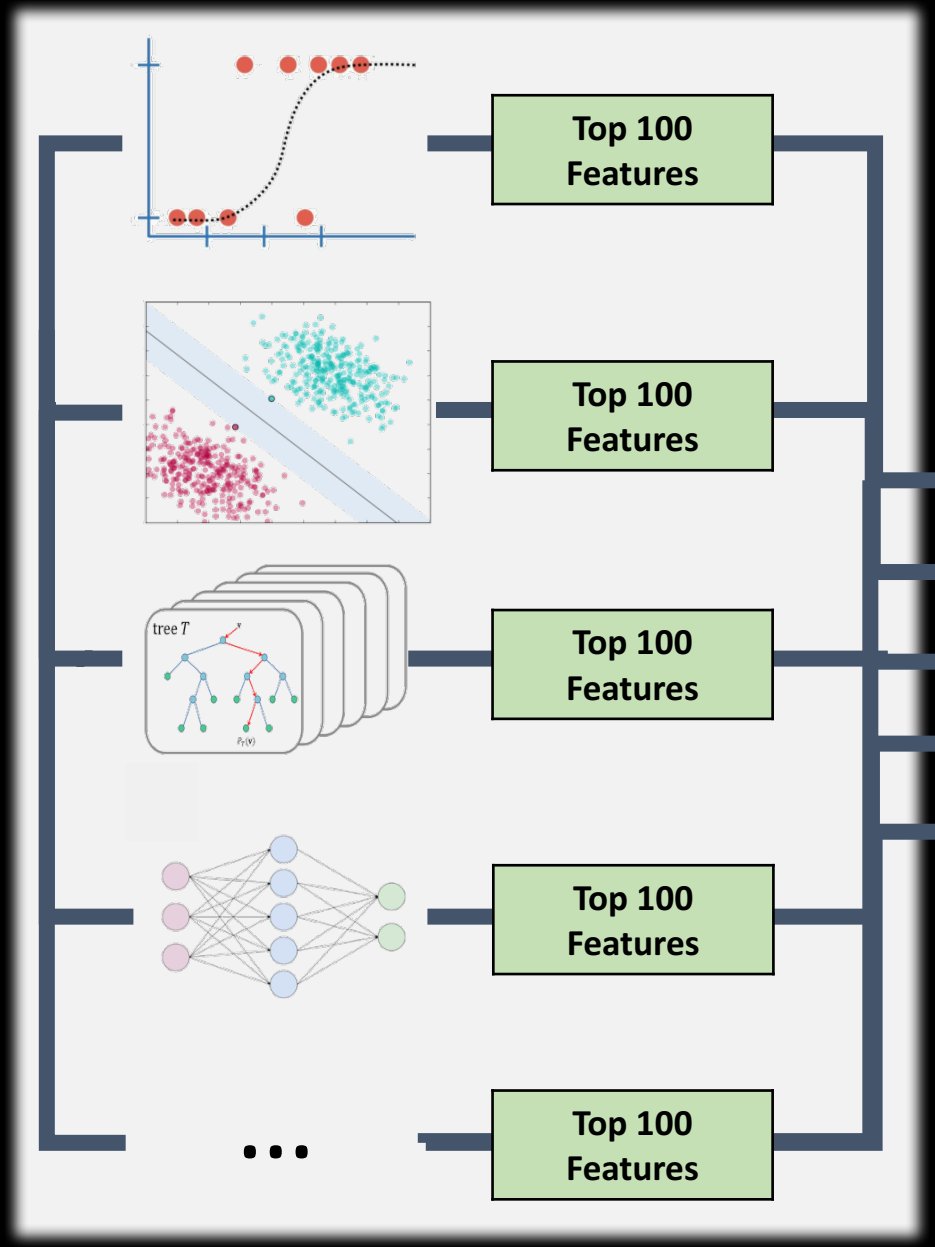
Test Performance

Average Test Precision
LGRG : 0.5911
SVC : 0.5737
RFC : 0.6698
MLP : 0.6790
GBT : 0.7317

Test F1 Score
LGRG : 0.5806
SVC : 0.6000
RFC : 0.6153
MLP : 0.6984
GBT : 0.7037

Test ROC AUC
LGRG : 0.9303
SVC : 0.9261
RFC : 0.9102
MLP : 0.9084
GBT : 0.9102

Feature Importance by Committee

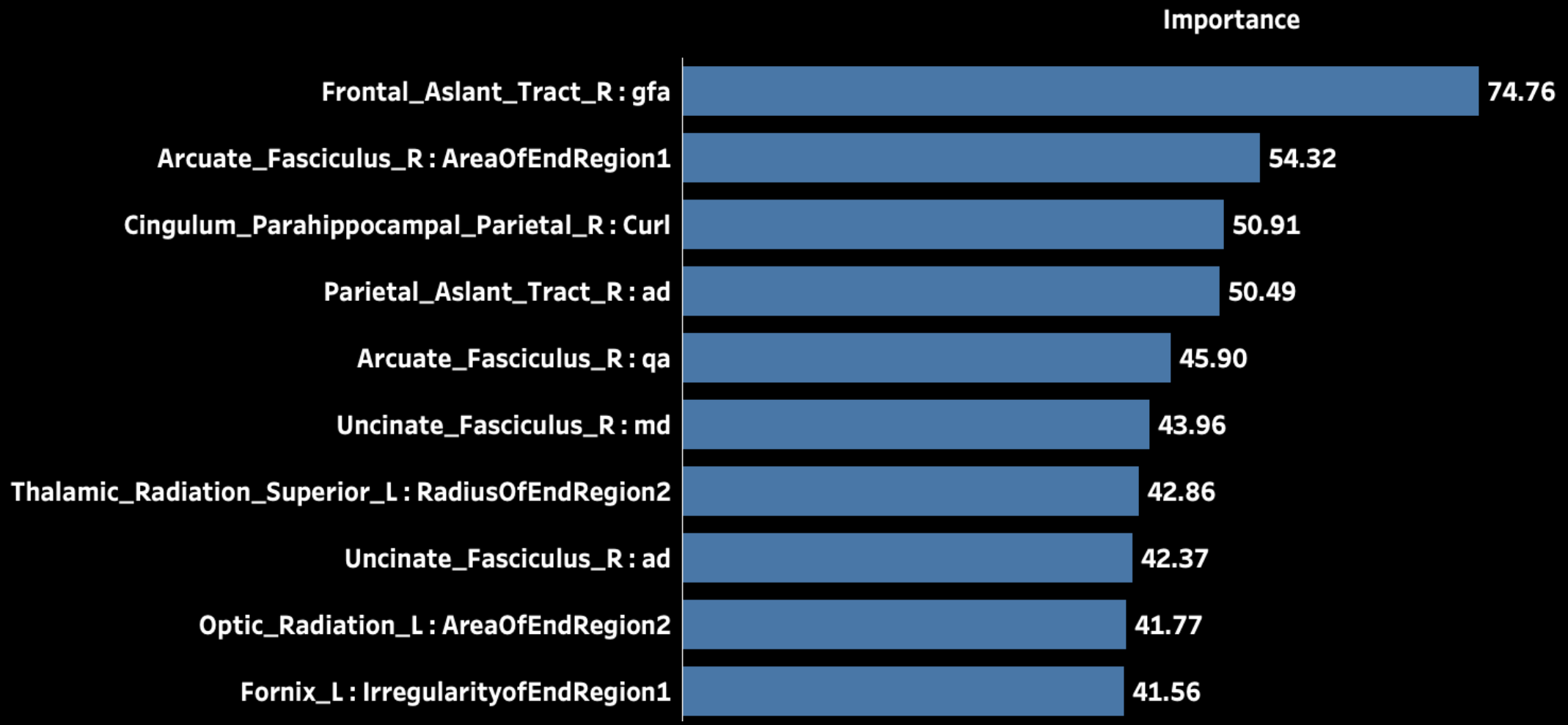


Model Feature Importance Weight	
Gradient Boosted Trees	0.220
Multi-Layer Perceptron	0.218
Random Forest	0.192
Support Vector	0.188
Logistic Regression	0.182

$\sum$

*based on models' test F1 Score

Relative Feature Importance by Committee

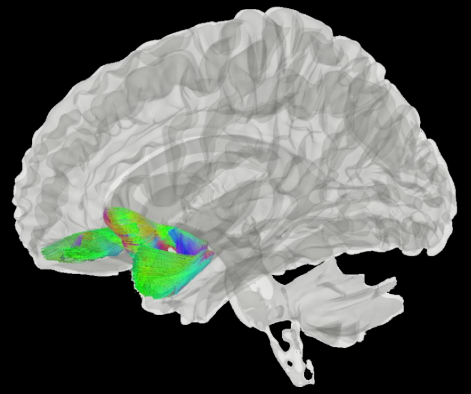


*appendix for full graph & specific model influence

Measure Importance by Committee

*appendix for full graph & specific model influence

Bundle Importance by Committee



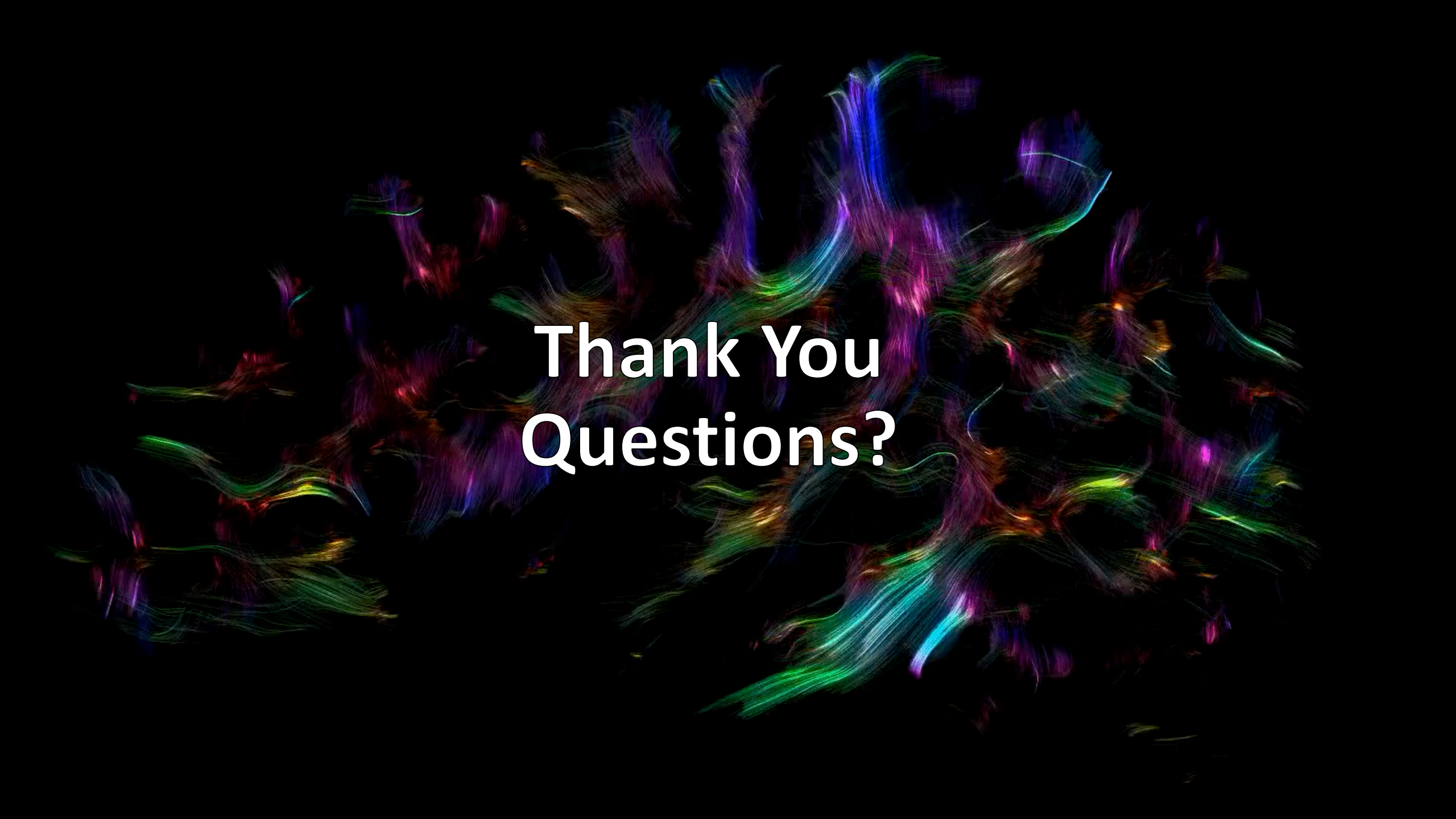
"consistent abnormalities in the **uncinate fasciculus** are found in **Alzheimer's disease**, semantic dementia, and temporal lobe epilepsy"

- Von der Heide, R; Skipper, L; Klobusicky, E; Olson, IR (2013)

*appendix for full graph & specific model influence

Future Work

- Confirm plausibility of results with neurologist
- Explore bundles of interest further (more derived features / actual scans)
- Add / change models to confirm robustness of committee importance determination methodology
- Incorporate available non-brain scan related data (handedness, sex, etc.)


Thank You
Questions?

Appendix (MLP Feature Importance Calculation)

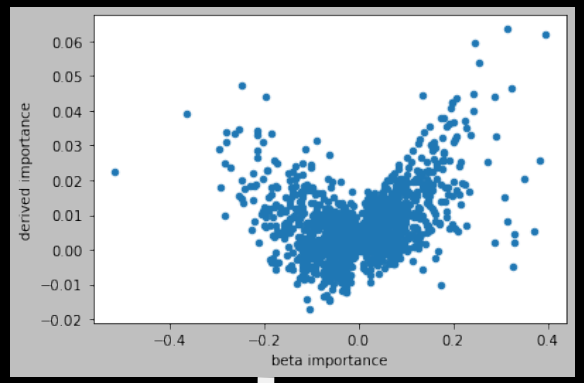
$$\text{Feature Importance}_j = f1\ Score - \frac{1}{n}\sum_{i=1}^{n} f1\ Score_{ij}$$

- The difference between the the baseline F1 score and the average F1 score calculated when taking a random permutation of the given feature in the test set.

- Low Difference ~ Model is insensitive to permutations of feature ~ feature is not very important
- High Difference ~ model is sensitive to permutations of feature ~ feature is important

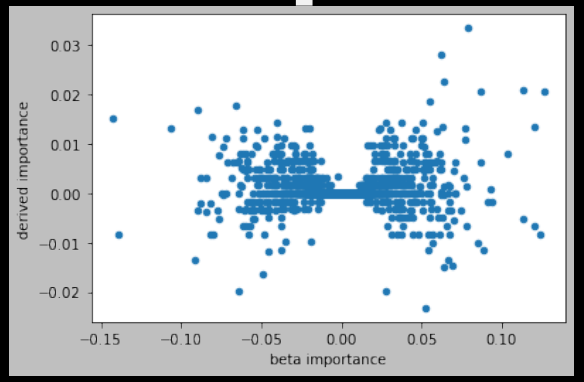
Appendix (MLP Feature Importance Sanity Check)

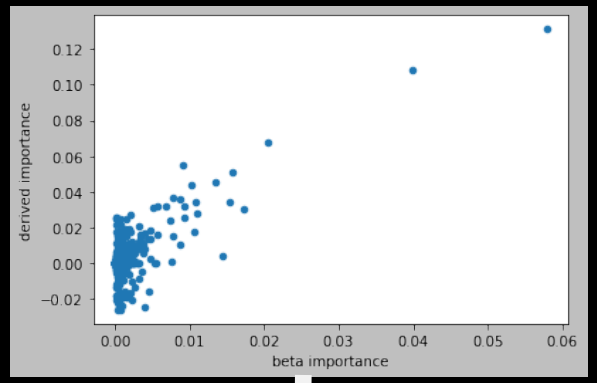
- Use model agnostic feature importance to estimate well known model feature importance metrics



Logistic Regression
Correlation = 0.54

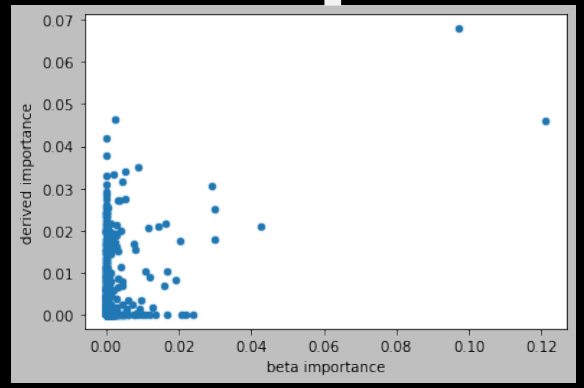
Support Vector
Correlation = 0.19



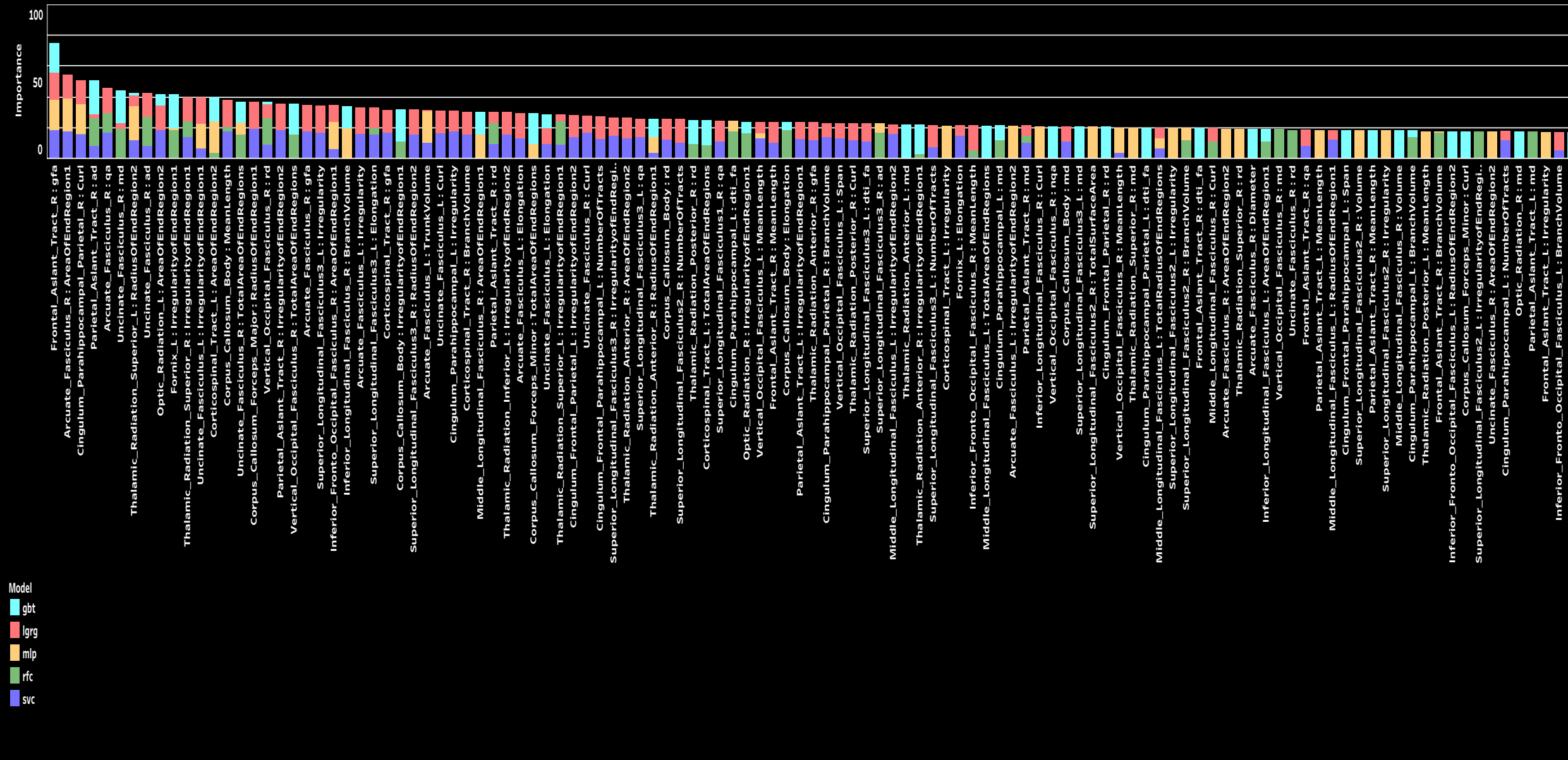


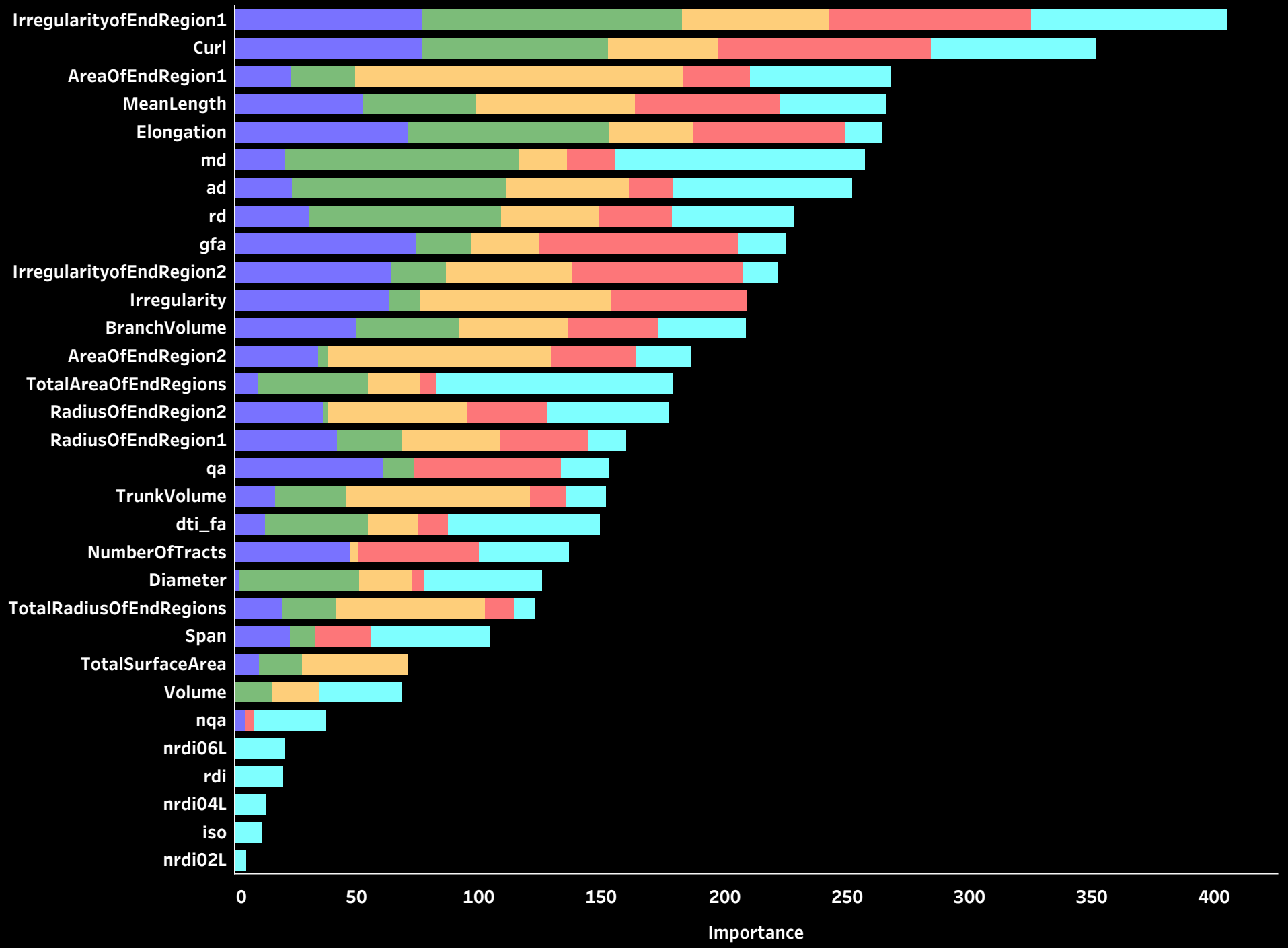
Random Forest
Correlation = 0.77

Gradient Boosted Trees
Correlation = 0.38



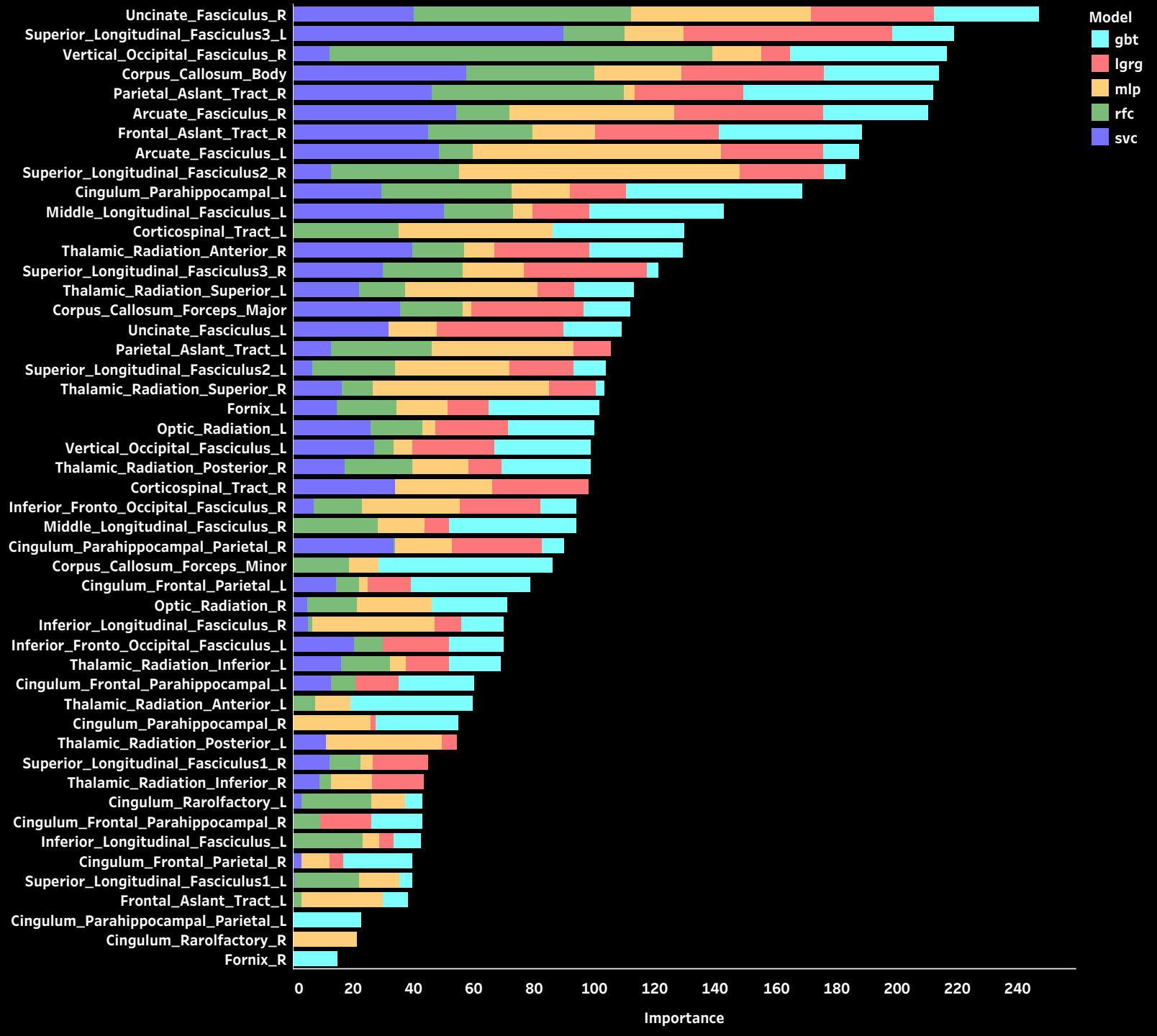
Appendix (bundle/measure)





Appendix Measure

Appendix Bundle



References

-"Alzheimer's Disease Fact Sheet". *National Institute on Aging*. Retrieved 25 January 2021.
-Cavaliere C, Aiello M, Di Perri C, Fernandez-Espejo D, Owen AM and Soddu A (2015) Diffusion tensor imaging and white matter abnormalities in patients with disorders of consciousness. *Front. Hum. Neurosci*. **8**:1028. doi: 10.3389/fnhum.2014.01028
- Yeh, F. C., Panesar, S., Fernandes, D., Meola, A., Yoshino, M., Fernandez-Miranda, J. C., ... & Verstynen, T. (2018). Population-averaged atlas of the macroscale human structural connectome and its network topology. NeuroImage, 178, 57-68. - http://brain.labsolver.org/
- Von der Heide, R; Skipper, L; Klobusicky, E; Olson, IR (2013). "Dissecting the uncinate fascicles: disorders, controversies, and a hypothesis". *Brain*. **136**(Pt6): 1692–707. doi:10.1093/brain/awt094
- https://radiopaedia.org/articles/diffusion-tensor-imaging-and-fibre-tractography?lang=us#:~:text=Diffusion%20tensor%20imaging%20(DTI)%20is,collected%20by%20diffusion%20tensor%20imaging.
- http://people.csail.mit.edu/lauren/dtmri.html
- https://en.wikipedia.org/wiki/Diffusion_MRI
- https://www.sciencedirect.com/science/article/pii/S0079656519300032?casa_token=HU1QpbMaYnkAAAAA:6XDzzO-fY6qWo2RFor-bKs5G4GjlUxoPrjTEMYuLxRC7mTsMiT9EA91R8wW57_aF1l4ZVCE
- https://www.jneurosci.org/content/32/22/7418
- https://www.youtube.com/watch?v=jrC8iY6_aZQ&ab_channel=AlfredAnwander