Uncovering the topological features

Characterization of a new biomarker in Parkinson’s Disease

El Missaoui Imane

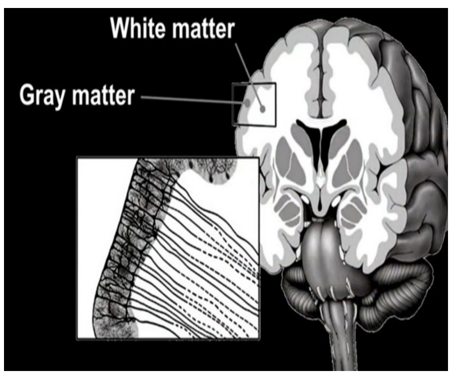
First year Masters Data science in Healthcare

Faculty of Health engineering and management

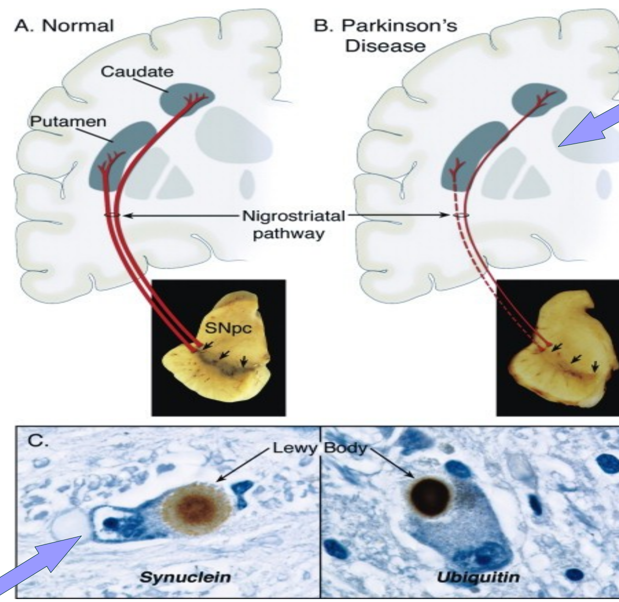
September 2, 2023

Parkinson’s Disease

- **Progressive neurodegenerative disorder** that primarily affects **automacity.**
- Characterized by the **degenerescence of the nigro-striatal pathway**, and presence of **Lewy bodies.**
- **Motor** symptoms of PD : bradykinesia, resting tremor, rigidity, and postural instability.
- **Non-motor** symptoms : cognitive impairment, mood disorders, sleep disturbances, and autonomic dysfunction.
- Both **genetic** (10% of all PD cases) **and environmental factors** contributing to the disease risk.



https://theconversation.com/brain-wrinkles-and-folds-matter-researchers-are-studying-the-mechanics-of-how-they-form-170194



Degenerescence of the nigro-striatal pathway

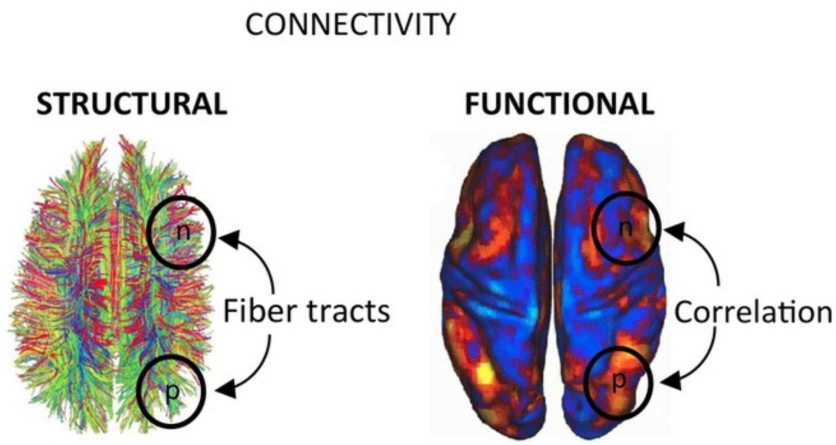
Lewy bodies

Neuropathology of Parkinson’s Disease

William Dauer and Serge Przedborski. “Parkinson’s disease: mechanisms and models”. eng. In:Neuron 39.6 (Sept. 2003), pp. 889–909. issn: 0896-6273. doi: 10.1016/s0896-6273(03)00568-3.

Brain Connectivity in Parkinson's Disease

- Two main types of brain connectivity: **structural** and **functional**.



Anatomical connections between brain regions formed by white matter tracts.

Correlations between neurophysiological events in different brain regions

https://www.semanticscholar.org/paper/Functional-connectivity-dynamically-evolves-on-over-Cabral-Kringelbach/81a81ff448b0911cbb63abecd9c54949ce8d50fd

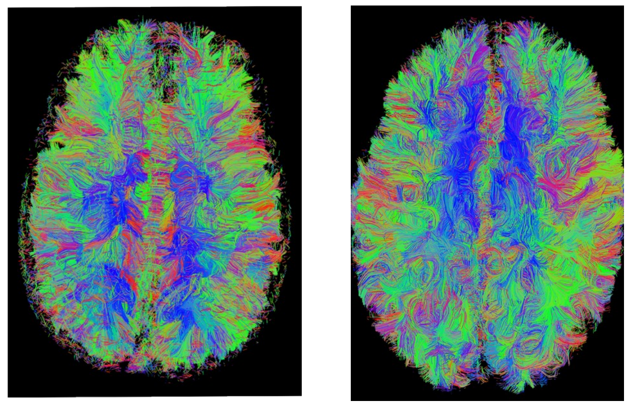
Alterations in the white matter integrity of Parkinsonian patients :

- **Reduced fractional anisotropy (FA)** in **the corpus callosum, corticospinal tracts,** and **the frontal** and **parietal association fibers.** → **axonal degeneration** or **demyelination.**
- **Increased mean diffusivity (MD)** in **substantia nigra**, **corpus callosum**, **cingulum**, etc → **higher diffusion and potential tissue damage**

Context and objective

Parkinson's disease impacts the morphology and connectivity of the brain, making networks a suitable tool for studying and modeling their effects.

PD HEALTHY



- Structural alterations correlate with the severity of motor and cognitive symptoms in PD patients, suggesting their potential as biomarkers for disease progression and cognitive decline

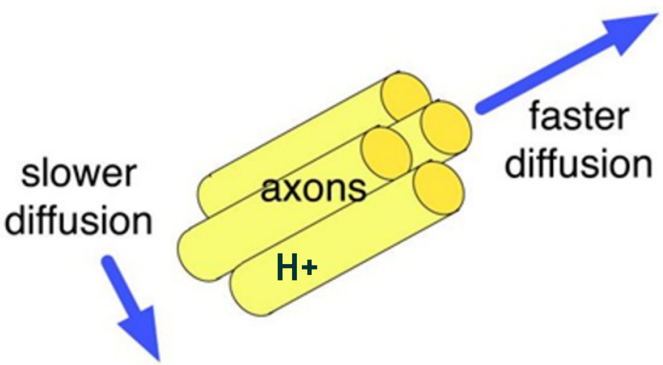
- The development of new biomarkers for Parkinson's disease prognosis is crucial to enhance our predictive capabilities.

Detect a difference between the different stages of Parkinson's disease progression

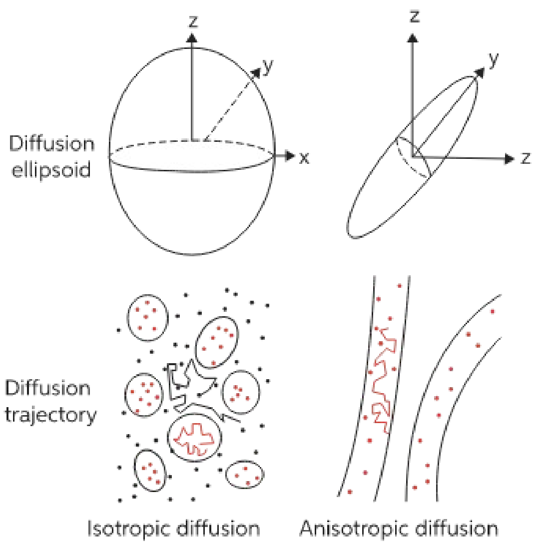
- Measuring the difference in structural connectivity between Parkinson's disease stages is done using **diffusion MRI Data**.

- Non invasive **neuroimaging technique** that studies **the movement of water molecules** within biological tissues in vivo.

- May indicate **(early) pathologic change :** Water molecule diffusion patterns can reveal microscopic details about tissue architecture, either normal or in a diseased state.



https://www.semanticscholar.org/paper/Implementation-of-an-algorithm-for-estimating-in-Golub/5c798fdb67178848cfe81460365fef91ee5d064f

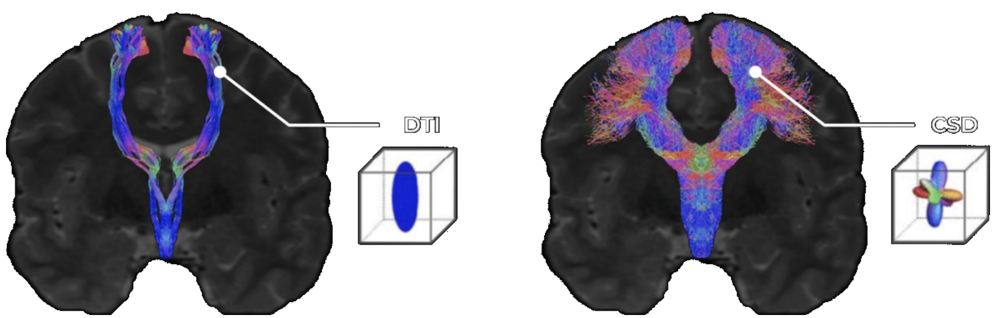


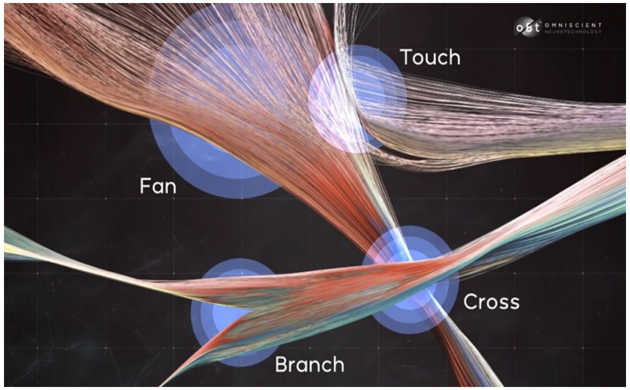
https://www.chegg.com/learn/topic/diffusion-tensor-imaging

- **Combining dMRI data with modeling algorithms, such as DTI or CSD, allows for the inference of diffusion direction(s) in each cubic region, or voxel, of the brain.**

- **Tractography** : the process of mapping the brain's white matter connections

-**DTI** : conventional tractography technique that is limited when modelling more complex structures of the brain





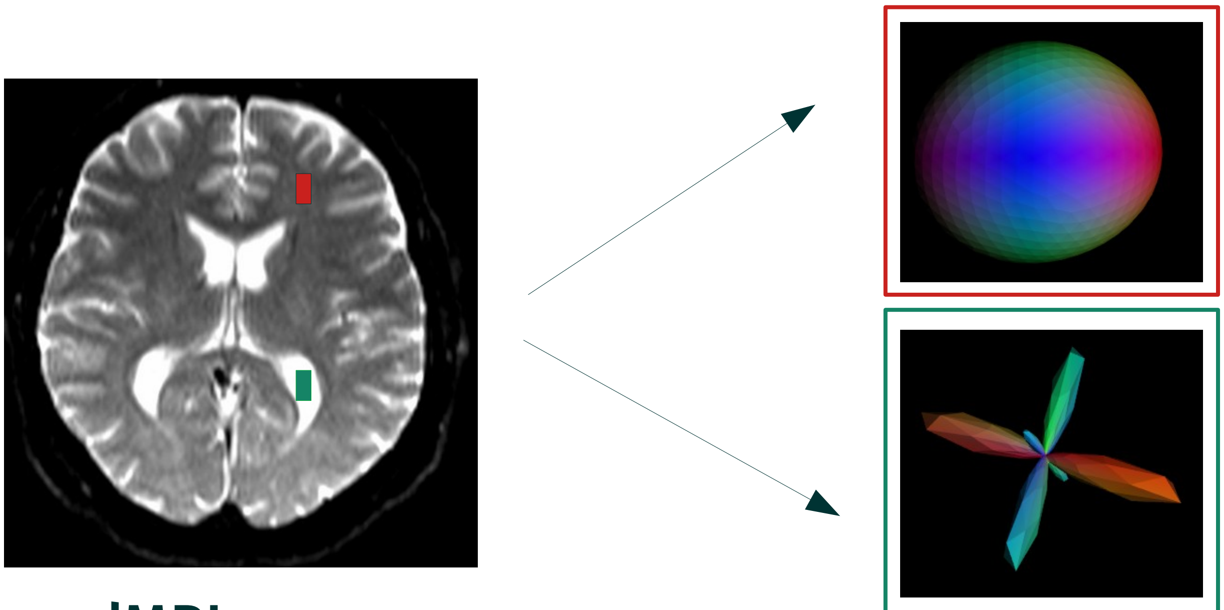
https://www.o8t.com/blog/tractography

Instead of an **'average water diffusivity'** value, CSD tractography generates a **'fiber orientation distribution' (fODF)** value for each voxel of neural tissue.

fiber Orientation Distribution Function to represent the stuctural connectivity

- An **fODF** is a 3D function that describes the distribution of fiber orientations within a voxel.

- **fODFs** provide **a more comprehensive representation** of the complex **intra-voxel fiber architectures**, offering a significant step forward in modeling and visualizing brain connectivity.



Isotropic

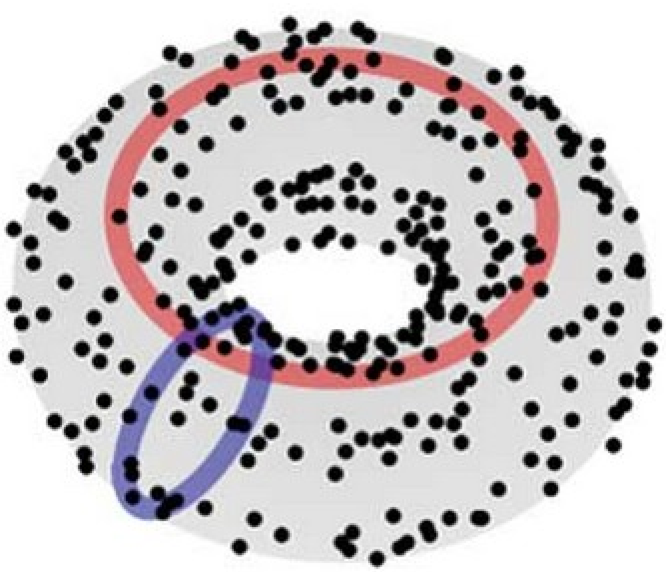
Anisotropic

dMRI scan

fODFs encode a lot of features at each voxel and can be thought of as data points in a high-dimensional space. The **goal** now is to **uncover the "shape" of this data**, **which represents the topological structure of the brain's white matter connectivity.**

Hypothesis in geometric data analysis :

Data has a shape: points are drawn from a geometric object that exists in a higher-dimensional space. Obtaining the properties of these shapes provides valuable information about the data



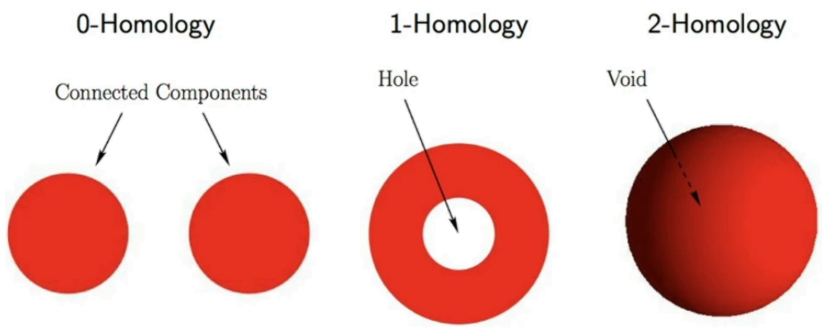
https://chance.amstat.org/2021/04/topological-data-analysis/

- **Data-Driven** Approach
- Studying **complex high dimensional data** without any assumptions
- **Shape has Meaning** ; extracting shapes(patterns) of data
- **Invariant to smooth deformations** (stretching, bending, scaling)

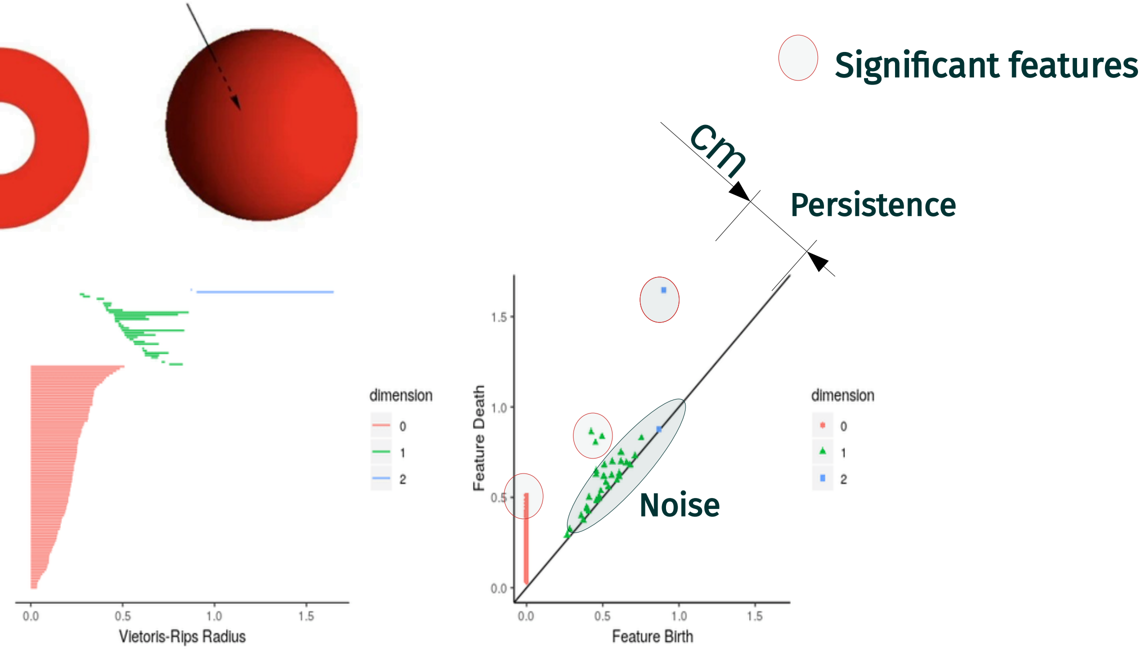
-Persistent Homology is a **mathematical method in TDA** that **identifies and quantifies topological features of a dataset throughout various scales**.

-The underlying principle of this method is the notion of "persistence", which refers to the lifespan of topological features as one varies the scale of observation.

-In a dataset, Persistent homology encodes topological features such as connected components (0-dimensional holes), loops (1-dimensional holes), voids (2-dimensional holes), and their higher-dimensional analogs can be detected.

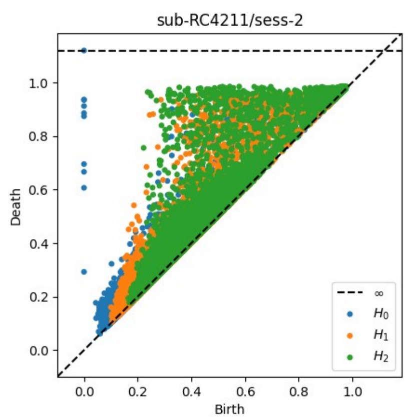


Represented as:

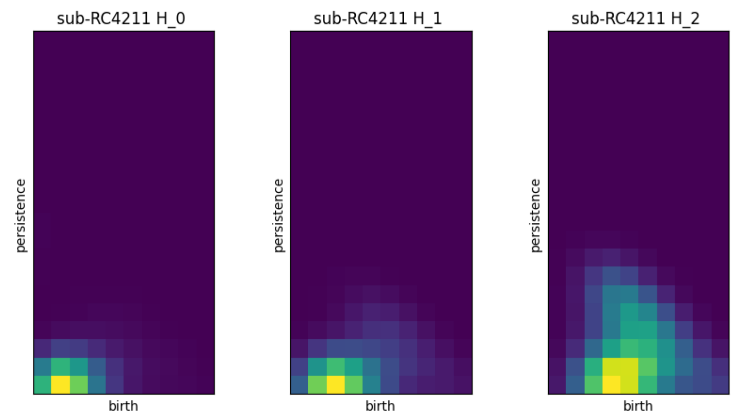


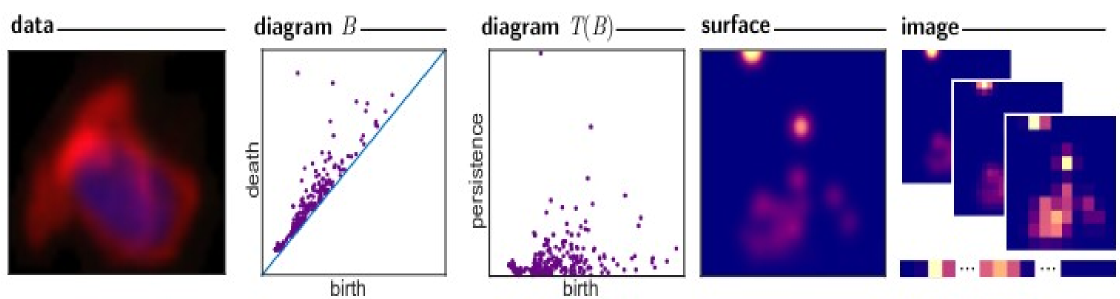
https://www.researchgate.net/figure/Topological-barcode-left-and-persistence-diagram-right-of-the-sphere3d-sample-dataset_fig1_326911777

- **Persistence Image** : regular grid of pixels obtained through "heat map" discretization
- The **pixel values** in the persistent image gives a **density estimate** of features in the persistence diagram
- Persistence images are a **stable** vector representation of Persistent Homology, **suitable for Machine learning algorithms**.



Transformation into an euclidian space



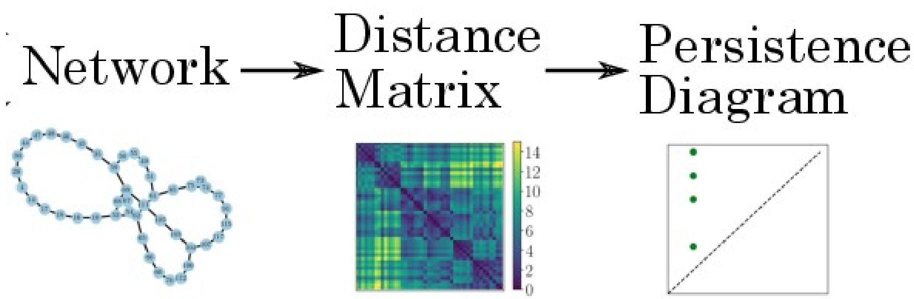


Pipeline to transform data into a persistence image

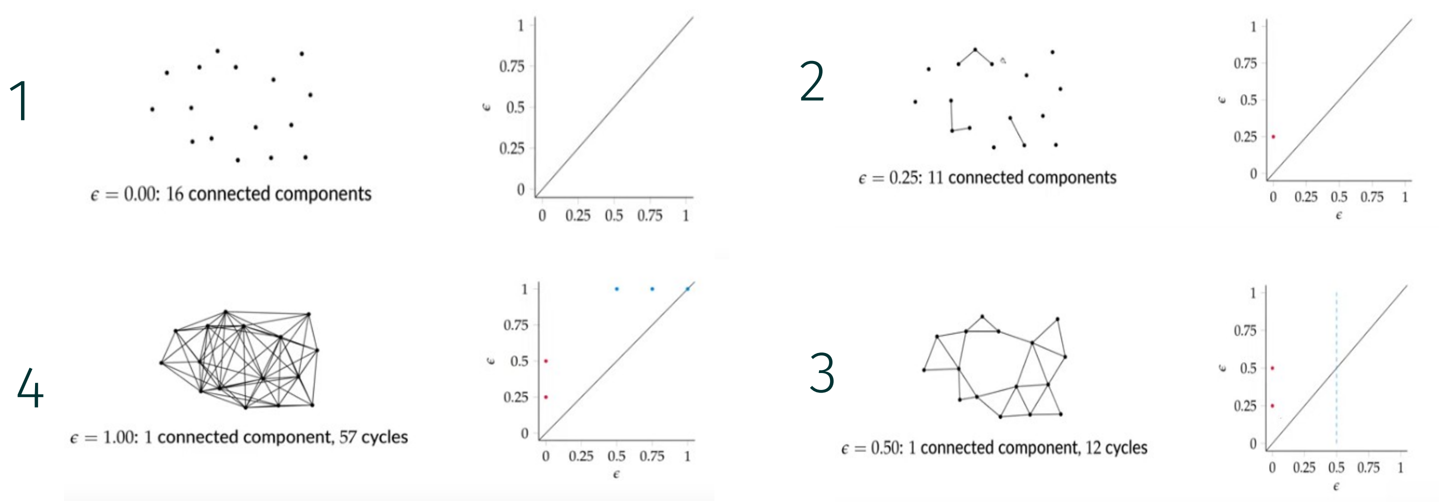
https://jmlr.org/papers/volume18/16-337/16-337.pdf

Distance matrices

- Essential in **computing Persistent Homology**
- Provide a **quantitative representation** of the **data's pairwise relationships**
- Guide **the construction of simplicial complexes** at **different scales**, leading to the **identification of topological features** that **persist over a range of parameter values**



http://firaskhasawneh.com/assets/repo_docs/PH_and_networks/index.html



https://www.youtube.com/watch?v=UKXIyC7l16g&t=2661s

First approach : Persistent Homology using fODFs Data

The pipeline:

dMRI Dataset

OpenNeuro Database Dataset "ds001907" 3-tesla MRI data

Model selection

dMRI scans

RUMBA-SD

fODFs Dilation

fODFs Construction

Growth Distance

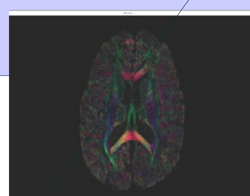
Persistent Homology

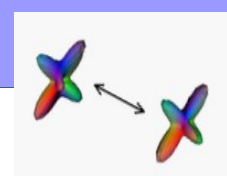
Growth distance Matrix

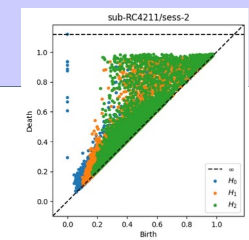
Persistence diagrams

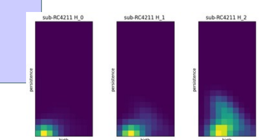
Persistence Images

Machine Learning classification







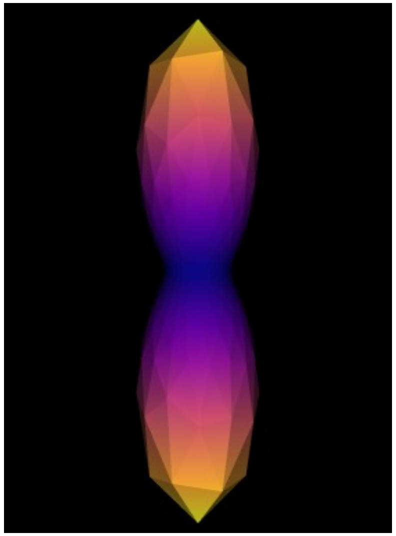


A variant of CSD methods from the open-source software library of Diffusion Imaging in Python **(DIPY)** to estimate fODFs from the dMRI signal

Measure **the distance between these fODFs** to capture the topological structure of the brain's white matter connectivity

The growth distance matrix can be used as a **distance matrix** to **compute persistence homology diagrams**

- Open-access database **OpenNeuro**
- Dataset ID **ds001907**
- Rich multi-modal neuroimaging dataset dedicated to the study of PD
- 46 participants : **healthy participants** (n = 25, RC41*) and **participants with Parkinson's disease** (n = 21, RC42*) at **two sessions each**



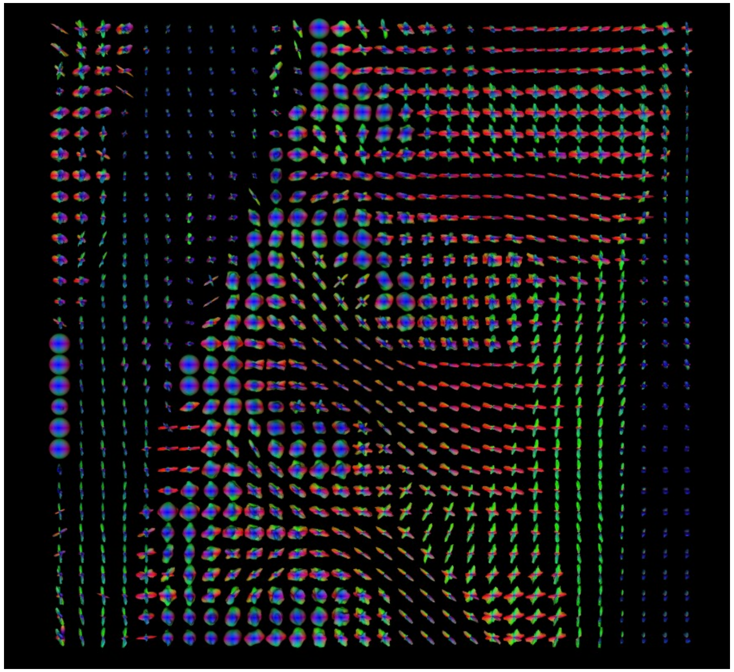
Step 1 : Estimation of the fiber response function

Step2 : Reconstruction of the fODFs voxel-wise

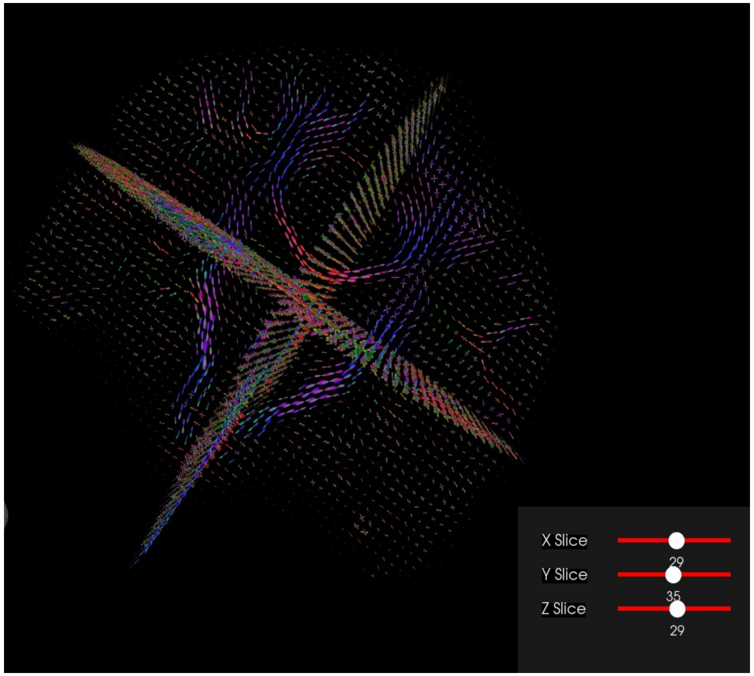
```
rumba = RumbaSDModel(
    gtab, wm_response=response[0], gm_response=None, sphere=sphere)

data_small = data[20:50, 55:85, 38:39]

rumba_fit = rumba.fit(data_small)
odf = rumba_fit.odf()
```

DIPY

Step3 : Interactive Visualization of the fODFs using Fury library



fODF map in the region of interest



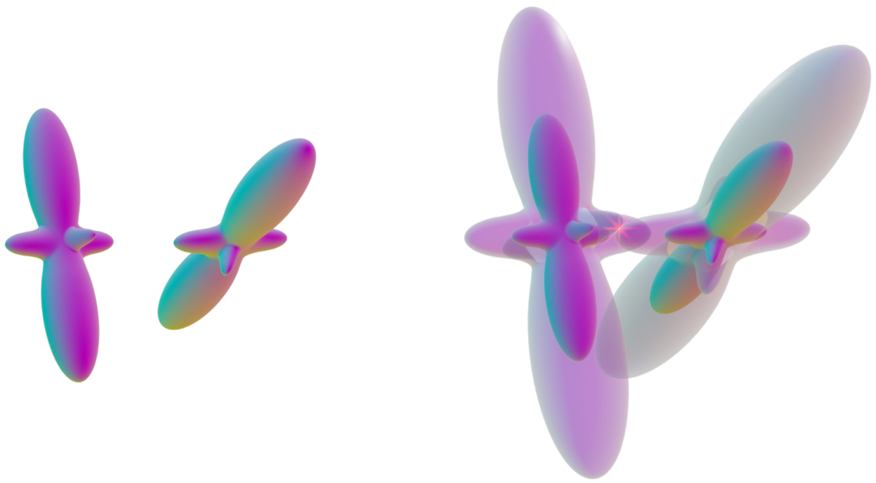
fODF slices whole brain

- In order to construct the distance matrix for computing persistent homology we need to measure the **distance between fODFs**

- This distance **should take into account the specific geometric properties** of the fODFs and **capture the inherent spatial nature of their orientations**

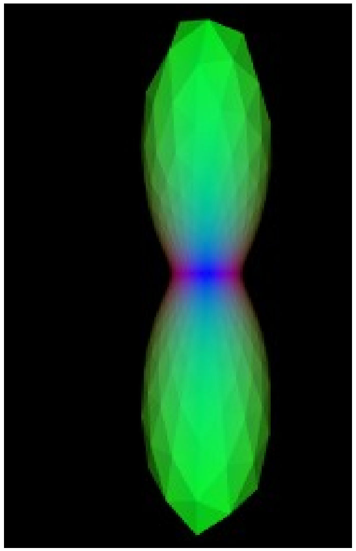
> **“Growth distance” as introduced by Ong and Gilbert is defined for a pair of convex objects as a measure of how much each of the objects must be grown, outward from fixed seed points in their interiors, so that they just touch**

[3]C. J. Ong and E. G. Gilbert, “IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, VOL. 12, NO. 6, DECEMBER 1996”.

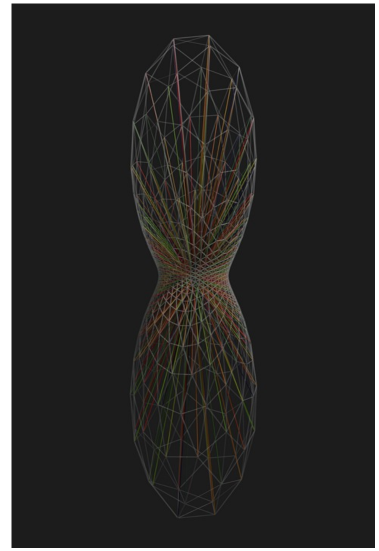


Growth distance between two fODFs

The growth distance concept between two objects requires that the objects of study should be convex



Tetrahedral subdivision



```
# Retrieve the tetrahedrons for ODF centers (0, 0, 0) and (0, 0, 1)
tetrahedrons_A = odf_cartesian_faces_dict[(0, 0, 0)]
tetrahedrons_B = odf_cartesian_faces_dict[(0, 1,0)]

# Calculate the optimized sigma_star
optimized_sigma_star = calculate_optimized_sigma_star(tetrahedrons_A, tetrahedrons_B)

# Print or use the optimized_sigma_star as desired
print("Optimized sigma_star:", optimized_sigma_star)

#run time 8:56
```

```
Optimized sigma_star: 1.982345336731888
```

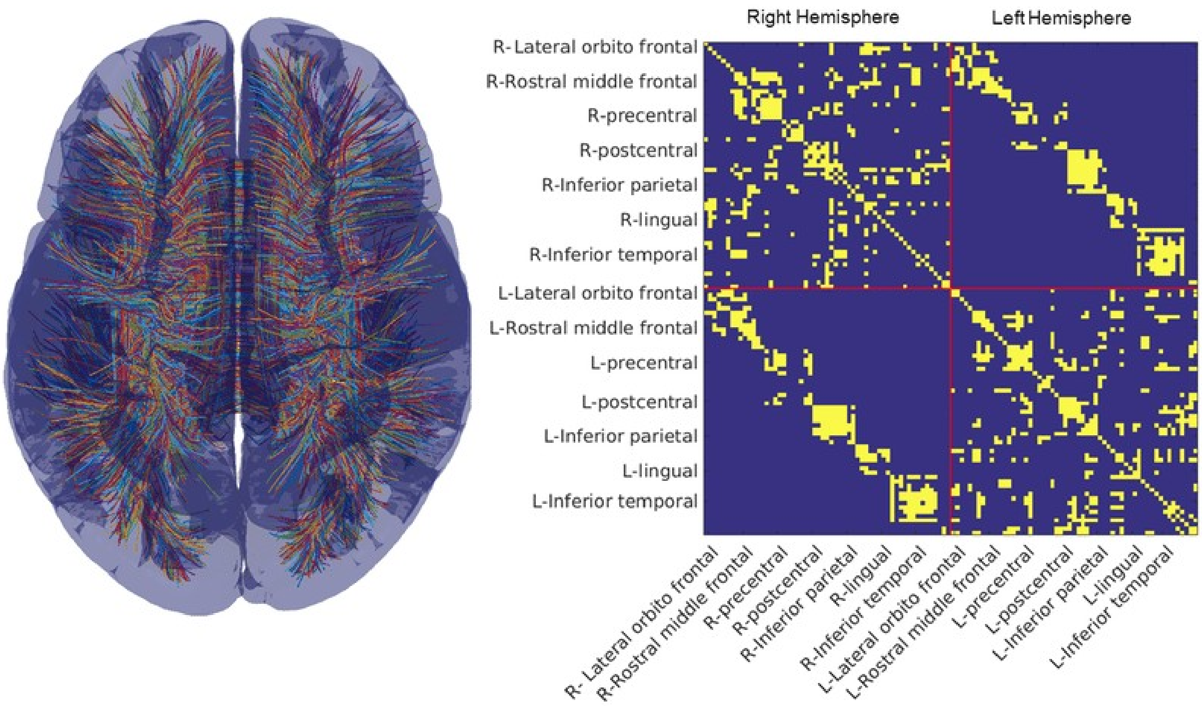
Growth distance

Computational complexity

1444 tetrahedrons per fODF
128*128*32 = 524288 voxels
524288 * 1444 = 757071872 tetrahedrons in total

757071872^2 calculations of distances between 2 fODFs

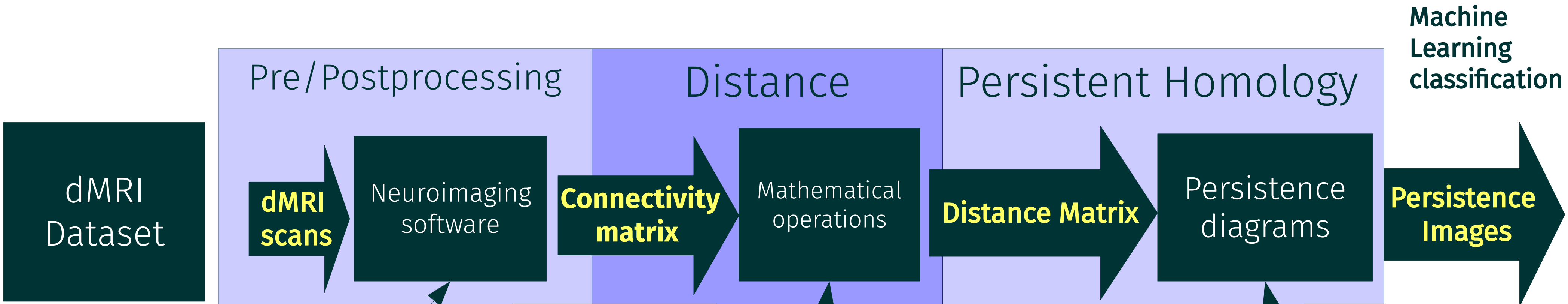
- **Structural connectome / Connectivity matrix** : the comprehensive map of these neural connections that shows the strength of connections between different brain regions

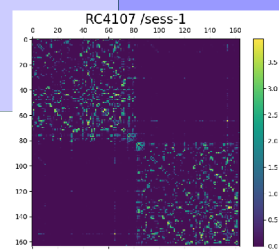


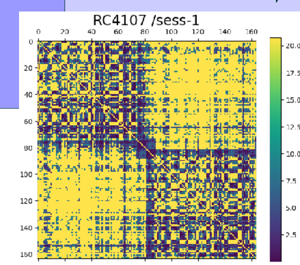
https://www.researchgate.net/publication/336162725_Contribution_of_structural_connectivity_to_MEG_functional_connectivity

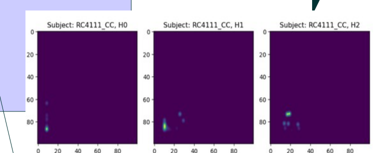
Persistent Homology using the structural Connectome

The pipeline:







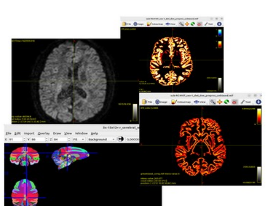


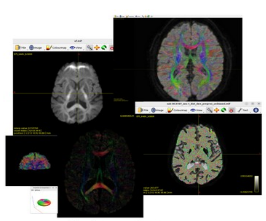
OpenNeuro Database
Dataset “ds001907”
3-tesla MRI data

+

Dataset FAIRPARKII

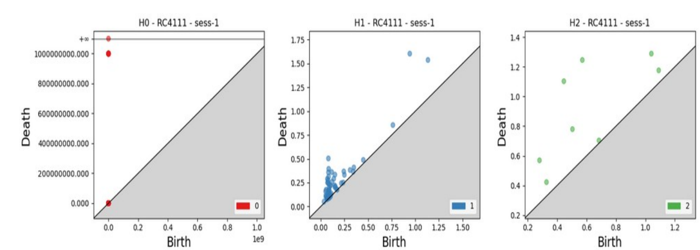
Preprocessing : noise reduction, data normalization, alignment, and brain extraction
Postprocessing : tractography, network construction, Connectivity matrix generation


Preprocessing


Postprocessing

Conversion of the **connectivity matrix into a distance matrix**, a format suitable for persistent homology

The **distance matrices** are used to **compute persistence homology diagrams**



1- Data acquisition

OpenNEURO

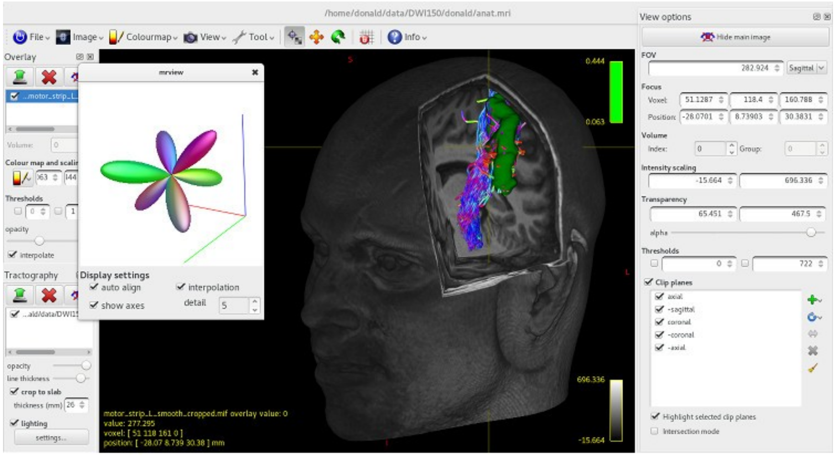
- Open access neuroimaging database
- Dataset ds001907
- Data of **healthy (n = 25, RC41*) participants at two sessions**
- Data of **participants with Parkinson's disease (n = 21, RC42*) at two sessions .**

FAIR PARK II

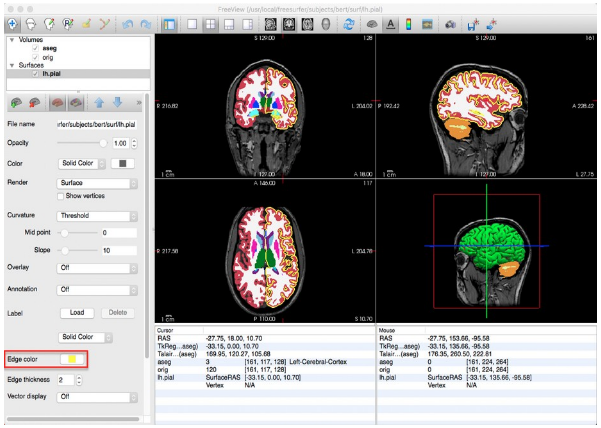
- Multicenter, phase 2, randomized, double blind international trial (Protocol 2015_22)
- Longitudinal study sponsored by the University Hospital of Lille
- Data of **newly diagnosed Pds(session W00)** who never received levodopa, some treated deferiprone
- Data of **matched placebo for 36 weeks(session W36)**

5 Controls from Openneuro + 5 Parkinsonian patients (placebo) from FPII

2- Preprocessing and post processing of dMRI data



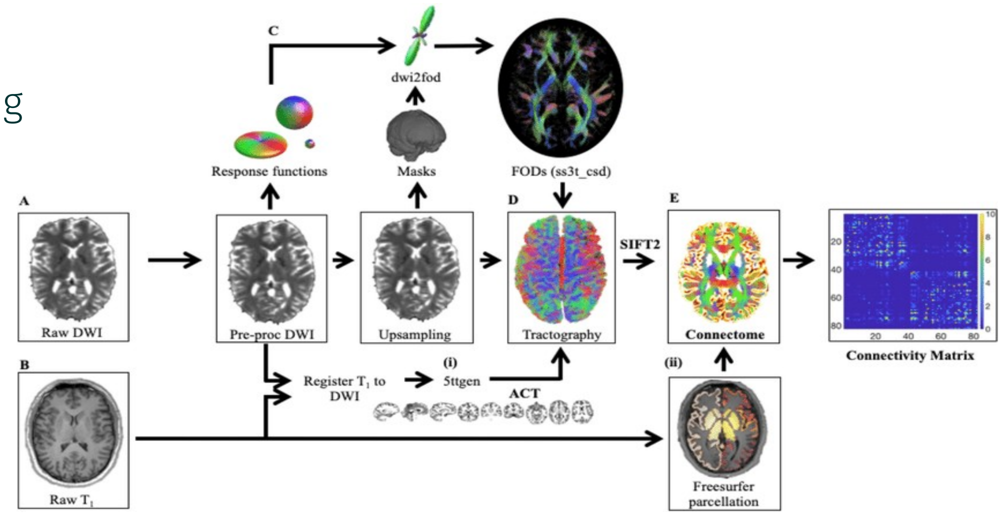
https://www.mrtrix.org/



https://surfer.nmr.mgh.harvard.edu/

- Software package for analyzing diffusion data
- Freely available under an open-source license.
- The preprocessing and post processing are done using commands executed in bash

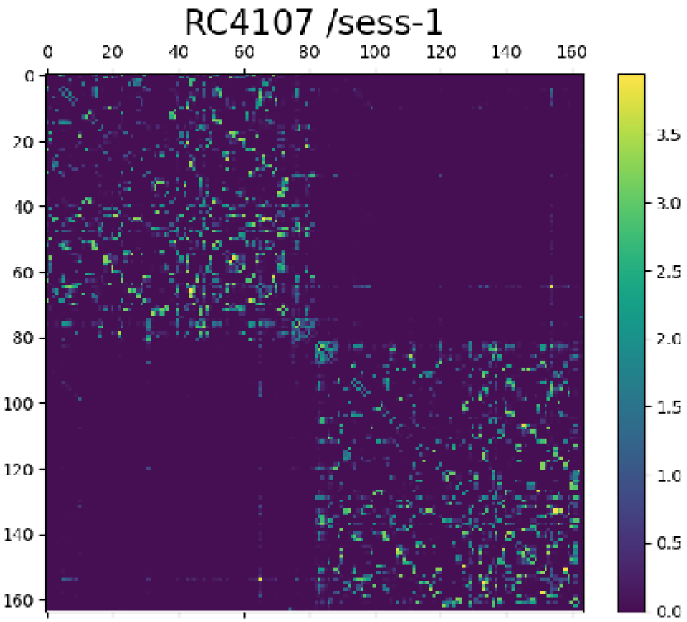
```
#Preprocessing
# Convert NIFTI to .mif and include gradient information
mrconvert  $fileNiiGZ "dwi.mif" -fslgrad $fileBvec $fileBval

# Denoising
dwidenoise dwi.mif dwi_den.mif -noise noise.mif

#Post processing
# Generate mask
dwi2mask dwi_den_preproc_unbiased.mif mask.mif

# Estimating the basis function using Tournier algorithm
dwi2response tournier dwi_den_preproc_unbiased.mif response.txt -voxels voxels.mif
```



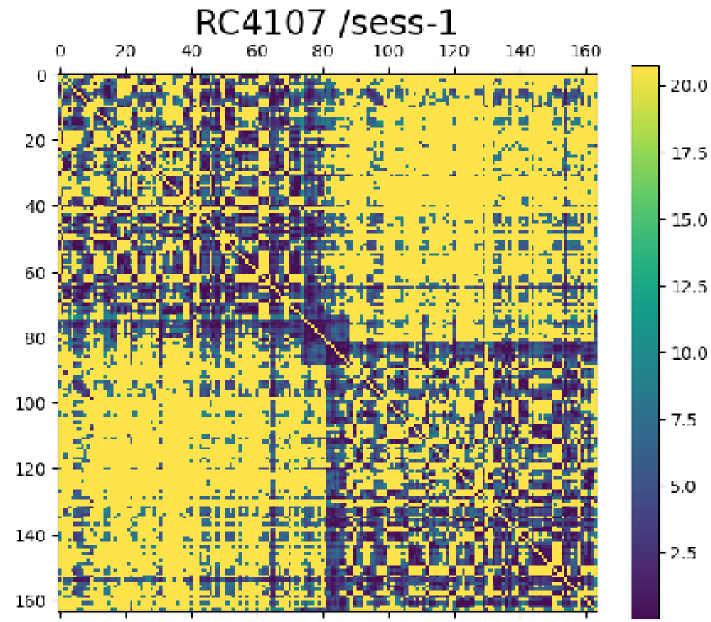
Phoebe Imms et al. "Navigating the link between processing speed and network communication in the human brain". In: Brain Structure and Function 226 (May 2021). doi: 10.1007/s00429-021-02241-8.

3- Conversion of Connectivity Matrices to Distance Matrices

For a format suitable for persistent homology, we consider each entry as a distance, with **stronger connections corresponding to shorter distances.**



Element-wise Inversion



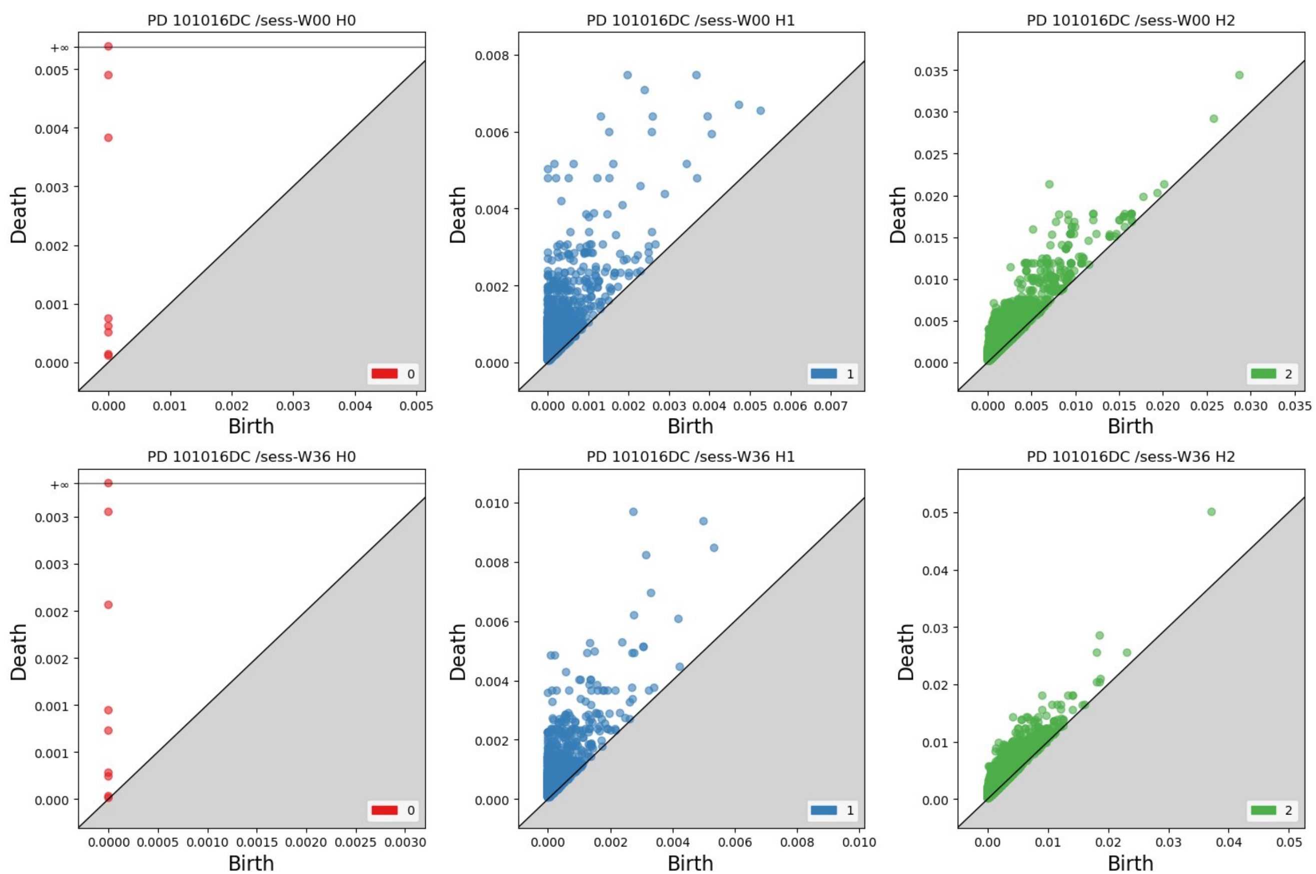
Connectivity Matrix

Each entry in the matrix represents the strength of the connection between two regions in the brain.

Distance Matrix

Each entry represents the distance between two regions in the brain.

Persistence diagrams for each homology degree for patient 101016DC

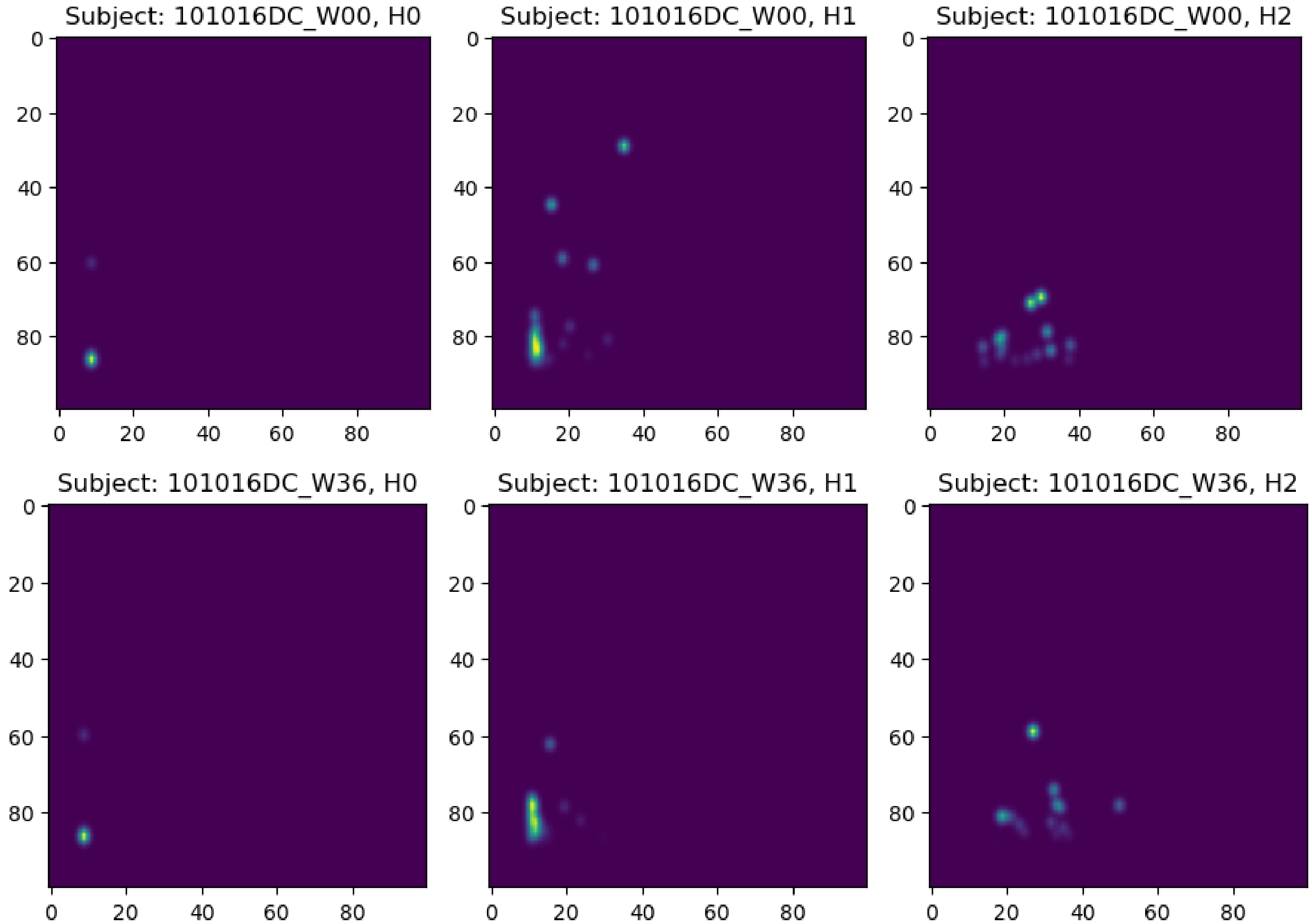


Sess W00 : the baseline

Sess W36 : after 36 weeks

- The persistence diagrams were computed using the **gudhi** library in Python
- 6 persistence diagrams for each patient (3 per session)
- 3 persistence diagrams for each control

Persistence images for each homology degree for patient 101016DC



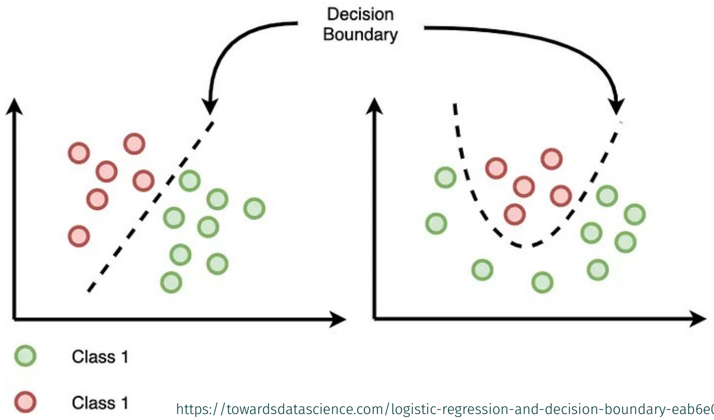
Sess W00 : the baseline

Sess W36 : after 36 weeks

- The persistence images were computed using the **persim** library in Python
- 6 persistence images for each patient (3 per session)
- 3 persistence images for each control

- 2 algorithms were used for this purpose

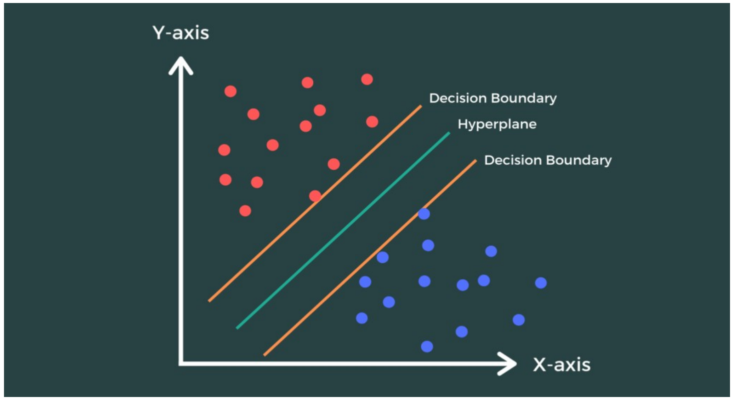
Logistic Regression



https://towardsdatascience.com/logistic-regression-and-decision-boundary-eab6e00c1e8

- Logistic regression can have **different decision boundaries** with different weights that are near the optimal point.
- Works with already **identified independent variables.**
- Based on **statistical approaches**.

Support Vector machine



https://www.theclickreader.com/support-vector-machine-svm-classifier/

- SVM tries to find the **"best" margin** (distance between the line and the support vectors) that separates the classes.
- Works well with **unstructured and semi-structured data** like text and images.
- Based on **geometrical properties of the data**.

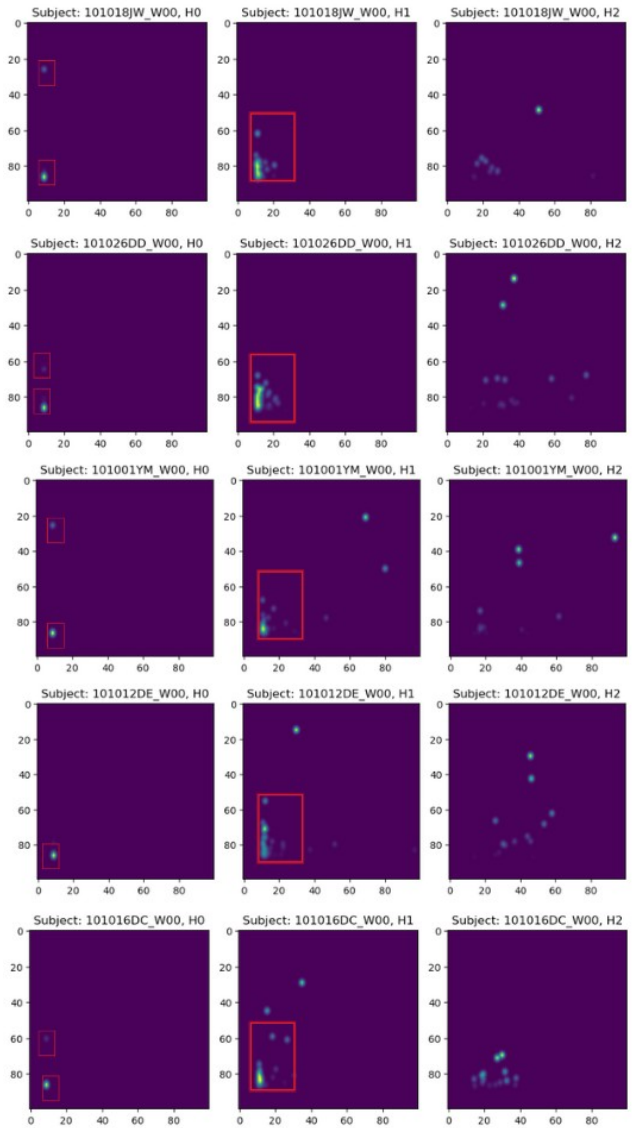
Goal : Track the variations of topological features between Parkinson patients and healthy Controls

Results :

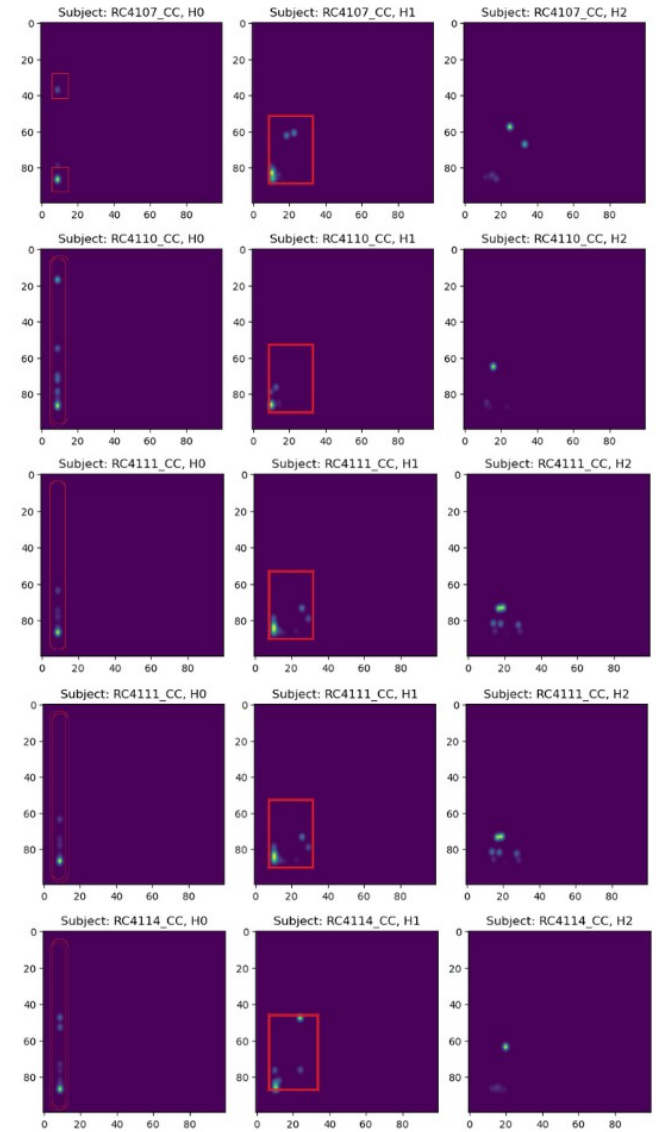
```
X_train, X_test, y_train, y_test = create_dataset(images_cc, images_pd)
train_logistic_regression(X_train, y_train, X_test, y_test)
train_svm(X_train, y_train, X_test, y_test)
```

Model evaluation

```
Training classifiers for H0
Logistic Regression Accuracy: 1.0
SVM Accuracy: 1.0
Training classifiers for H1
Logistic Regression Accuracy: 1.0
SVM Accuracy: 1.0
Training classifiers for H2
Logistic Regression Accuracy: 0.5
SVM Accuracy: 0.5
```



PDs / sess-W00



Controls

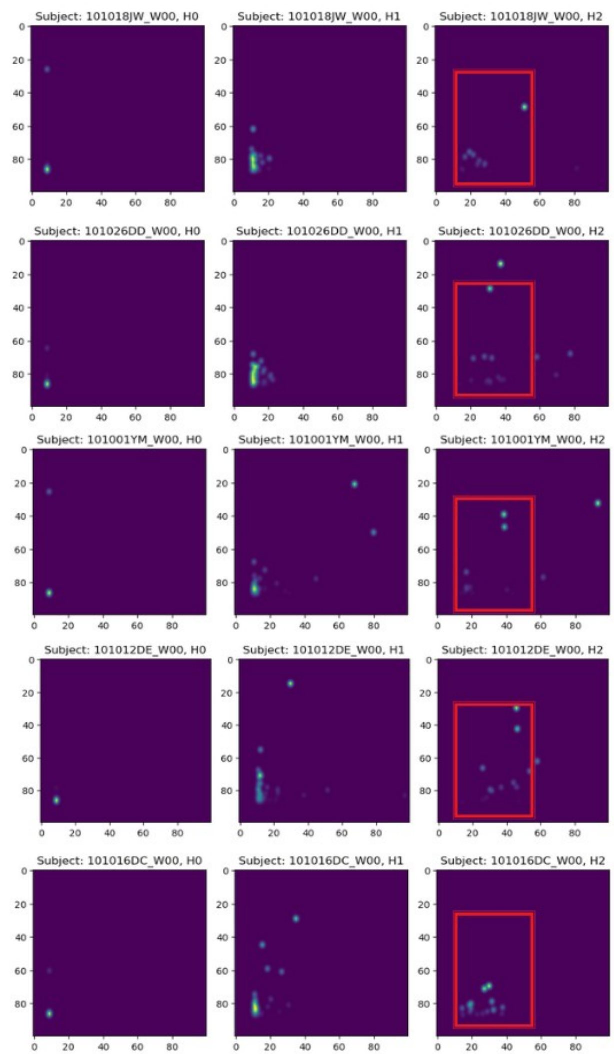
Goal : **Track the evolution of topological features over time and in relation to the disease's progression.**

Results :

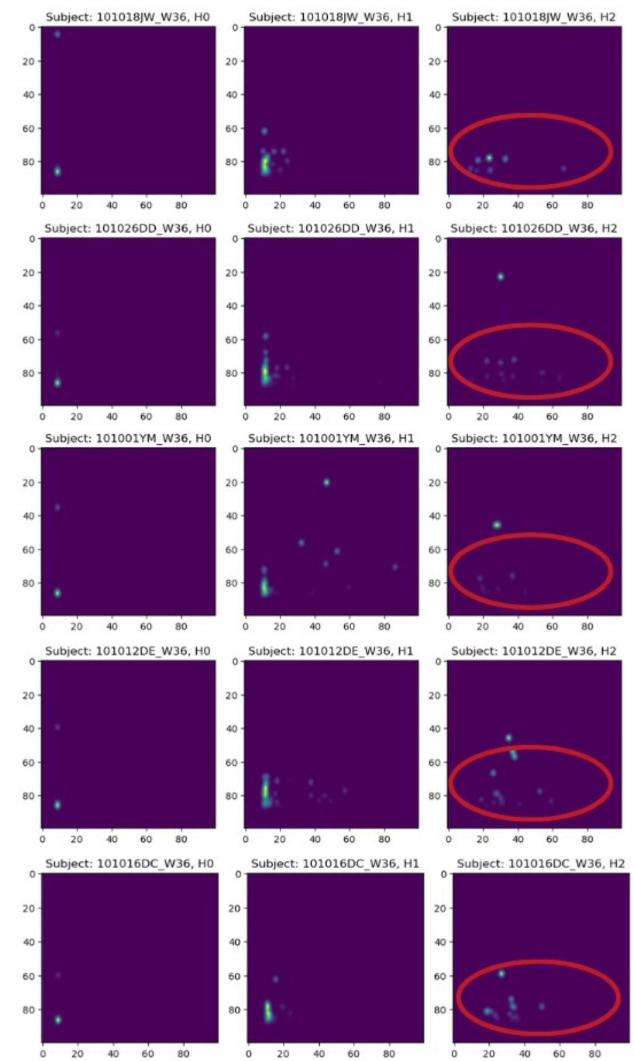
```
train_classifiers(H0_pd_images_W00, H0_pd_images_W36, "H0")
train_classifiers(H1_pd_images_W00, H1_pd_images_W36, "H1")
train_classifiers(H2_pd_images_W00, H2_pd_images_W36, "H2")
```

Model evaluation

```
Training classifiers for H0
Logistic Regression Accuracy: 0.5
SVM Accuracy: 0.5
Training classifiers for H1
Logistic Regression Accuracy: 0.5
SVM Accuracy: 0.5
Training classifiers for H2
Logistic Regression Accuracy: 1.0
SVM Accuracy: 1.0
```



PDs / sess-W00



PDs/ sess-W36

Classification into three different classes

Goal : Discriminate the persistence images into three distinct classes: Controls (CC), PD_W00(just diagnosed), and PD_W36(after 9 months), across different homology degrees in order to track the natural progression of the disease.

```
# Training the classifiers for each homology degree
accuracy_H0 = train_SVM(H0_cc_images, H0_pd_images_W00, H0_pd_images_W36)
accuracy_H1 = train_SVM(H1_cc_images, H1_pd_images_W00, H1_pd_images_W36)
accuracy_H2 = train_SVM(H2_cc_images, H2_pd_images_W00, H2_pd_images_W36)
```

```
SVM Accuracy for H0: 0.3333333333333333
SVM Accuracy for H1: 0.6666666666666666
SVM Accuracy for H2: 0.0
```

Paramater tuning :

Selecting Hyperparameters which are parameters of the model that are not learned from the data but set prior to training, in order to improve the model accuracy

Kernel Type : based on the characteristics of the data(Linear, RBF, etc).
C Parameter : controls the trade-off between maximizing the margin and minimizing the classification error.
Gamma (for RBF Kernel) : controls the shape of the decision boundary. Higher values of gamma result in more complex decision boundaries.

```
SVM Accuracy for H0: 0.3333333333333333
SVM Accuracy for H1: 0.6666666666666666
SVM Accuracy for H2: 0.3333333333333333
```

Conclusion

- The computational complexity associated with fiber Orientation Distribution Functions could be addressed using Code Optimization or/and optimizing the fODFs data
- A larger, more varied dataset will ensure a more comprehensive and generalizable analysis
- Improved classification models or more sophisticated feature extraction methods may be required to enhance accuracy.

Topological data analysis on 3-tesla diffusion MRI images combined with Machine learning classification algorithms have enabled the differentiation of newly diagnosed subjects from those with nine months of disease progression, suggesting a promising imaging biomarker for the prognosis of Parkinson's disease.

Thank you for your attention :)