Yaniv Assaf Lab Assignment – Gal Levy

Pre-Abstract

In this assignment, I am tasked with analyzing neuroimaging data to explore the spatial distribution of mean diffusivity across various brain regions. Utilizing Python, I will process and analyze NIfTI-formatted data, specifically a mean-diffusivity map and a registered atlas of a subject. My objective includes the removal of outliers and calculation of average mean diffusivity for each indexed region in the atlas. This comprehensive analysis aims to visualize and interpret the variation of mean diffusivity, enhancing our understanding of its significance in different brain areas.
Assignment source: https://yanivassaflab.com/apply/

Abstract

This study investigates the spatial distribution of mean diffusivity (MD) across brain regions by processing neuroimaging data in NIfTI format using Python, which includes a mean-diffusivity map and a registered atlas of a subject. Focusing on outlier removal and the calculation of average mean diffusivity for each of the 274 indexed regions, the analysis reveals that the distribution of MD values for all voxels approximates a Gaussian distribution with a slight right tail. This pattern holds consistently across individual regions. In subsequent analyses, I go deeper into the characteristics of each region and sub-region, aiming to uncover detailed insights into the microstructural properties that differentiate various parts of the brain.

Introduction

Diffusion Tensor Imaging (DTI) is an advanced MRI technique that enables the measurement of the directional diffusion of water molecules in biological tissues, primarily used to study white matter tracts in the brain. This technique provides insights into the microstructural organization of brain tissue by capturing the diffusion process, which can be anisotropic (directionally dependent) in structured tissues like neural fibers and isotropic (uniform in all directions) in less structured tissues. DTI quantifies this diffusion process using a mathematical model represented by a tensor, from which several scalar metrics can be derived to describe tissue characteristics.

One of the key metrics derived from DTI is Mean Diffusivity (MD), which represents the average rate at which water molecules spread out from their initial location over time. MD is a scalar value that gives an overall indication of the cellular density and integrity of brain tissues. High MD values can indicate a loss of tissue integrity or reduced cellular density. On the other hand, low MD values suggest tightly packed cellular environment.

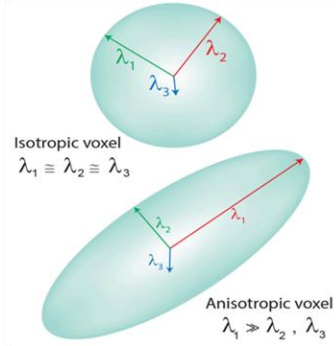

Figure 1: *The image show two voxels illustrating diffusion tensor imaging (DTI) principles: an isotropic voxel, where water diffusion (λ1, λ2, λ3) is equal in all directions, and an anisotropic voxel, with greater diffusion along one axis (λ1) than the others*

The data for this study comprises two NIfTI (Neuroimaging Informatics Technology Initiative) format files: a mean-diffusivity (MD) file (Sub2_MD.nii) and a registered brain atlas (Sub2_AtlasBNA.nii). The MD file contains voxel-wise MD values calculated from DTI scans of a subject, providing a detailed representation of water molecule diffusion throughout the brain. The registered atlas file contains 274 distinct brain regions, with indexes ranging from 0 to 274, where region 0 represents non-brain areas, regions 1-246 correspond to the cerebrum, and regions 247-274 are associated with the cerebellum. The size of the dataset is specified by the dimensions of the imaging data, which are 131 x 155 x 126 voxels. This three-dimensional describes MD values and regions across the entire brain.

In the provided dataset, brain regions **we**re identified solely by numerical labels, necessitating an extensive search online to determine the corresponding brain areas for each number. This effort was critical to accurately associate the mean diffusivity values calculated from the DTI data with specific anatomical regions, enabling a detailed analysis of the spatial distribution of mean diffusivity throughout the brain. Eventually I located the necessary information delineating the numerical identifiers to their respective brain regions at the following URL: https://academic.oup.com/view-large/86181045

Analysis Results

In this study, I began My exploration by rigorously assessing the shape, type, and overarching statistical metrics of the neuroimaging data at hand. Such an initial analysis is imperative for several reasons. Firstly, it confirms that the mean diffusivity (MD) and atlas data are precisely aligned in terms of their dimensions, a critical factor for ensuring accurate mapping of MD values to specific brain regions. Secondly, understanding the data types and their overall distribution provides a preliminary sense of the data's quality and complexity, guiding the selection of appropriate analytical techniques. Lastly, this early engagement with the dataset helps in identifying any anomalies that may require specialized handling, laying a solid foundation for the intricate analyses that follow. This step is crucial for setting the stage for a detailed examination of the spatial distribution of mean diffusivity across the brain, enhancing our understanding of its microstructural variations.

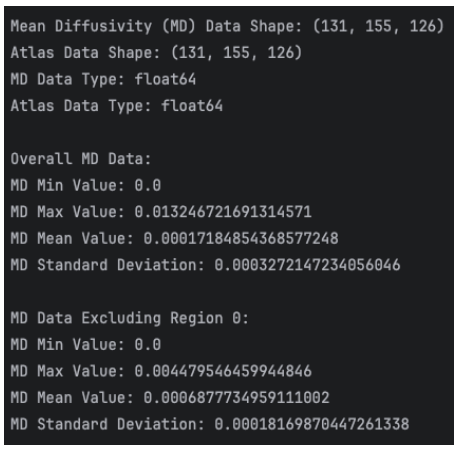

Figure 2: *Upon evaluating the mean diffusivity (MD) data, it was confirmed that the dimensions of the MD and atlas data perfectly match, ensuring accurate alignment between MD values and brain regions. Analysis revealed that excluding non-brain areas (region 0) from the dataset resulted in a notable increase in the mean MD value. This observation indicates that the non-brain area, largely contributing zero values, significantly affects the overall distribution of MD values. Furthermore, the standard deviation of MD values decreased upon excluding region 0, suggesting a more uniform distribution of MD within the brain tissue. The high maximum MD value observed when including region 0 points towards the presence of outliers, likely attributed to non-brain areas with different diffusion properties. This streamlined analysis enhances our understanding of the brain's microstructural integrity by providing a clearer picture of mean diffusivity distribution, emphasizing the importance of distinguishing between brain and non-brain regions.*

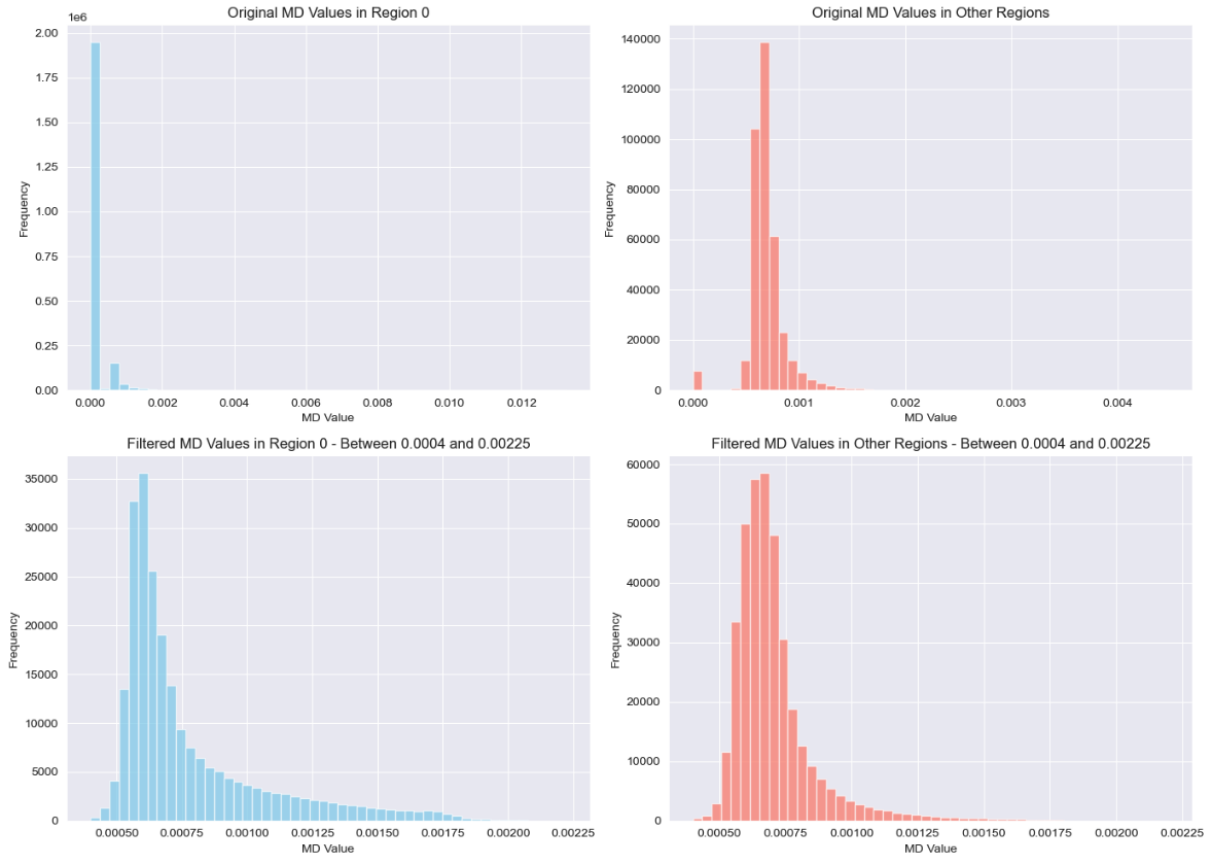

Figure 3: *The figure presents histograms of mean diffusivity (MD) values in brain and non-brain regions (region 0). Region 0's histogram predominantly shows zero values, aligning with its classification as a non-brain area, hence its exclusion from further MD analysis. The original MD values in other brain regions exhibit a broad range with a right-skewed distribution, suggesting the presence of outliers. To address these outliers and focus on physiologically relevant MD values, the data range has been narrowed to 0.0004 to 0.00225 for all regions. Post-filtering, the histograms for both sets display a more Gaussian-like distribution with a right tail, indicating that most MD values are concentrated around a central tendency with some higher values. In summary, region 0 will be disregarded due to its non-brain status, and the filtered MD distribution in brain regions, showing an approximately normal distribution with a rightward skew, will be used for further analysis. This approach minimizes the impact of extreme values and provides a more accurate representation of MD in brain tissue.*

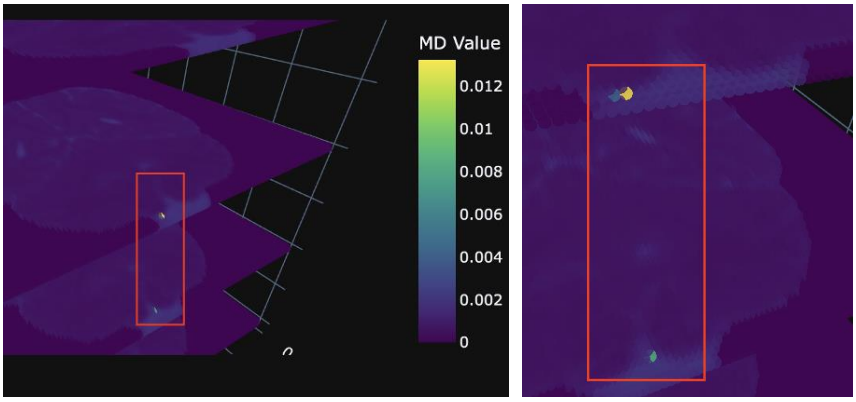

Figure 4: *Examples of Outliers Detected and Removed from the Data. Here, we see an axial cut of slices 51-54. The top 50 highest values and their corresponding voxel numbers are listed in the eda.ipynb file.*

Upon closely examining the data, I must confess that the process of outlier identification and removal presented itself as an ambiguous task. Traditionally, outliers are perceived as data points that stand apart from the general trend or distribution, often due to measurement discrepancies or atypical sample behaviors. As depicted in Figure 5, a mere trio of voxels manifested what could be considered outlier characteristics, distinctly separating themselves from the pack. This deviation led me to categorize them as outliers. However, the rest of the data, especially the extremes, seemed to comfortably align with the anticipated Gaussian distribution with right tail, suggesting a natural variation rather than an error.

Diving into research, I uncovered that data often labeled with outliers may be treated with a variety of corrective measures, including mathematical transformations and linear algebra techniques. I tried to do some transformation which described later. Despite these efforts, the transformation applied was not significantly correct. Consequently, I decided to adjust the mean values for the three anomalous voxels, which seemed the most practical approach in this context.

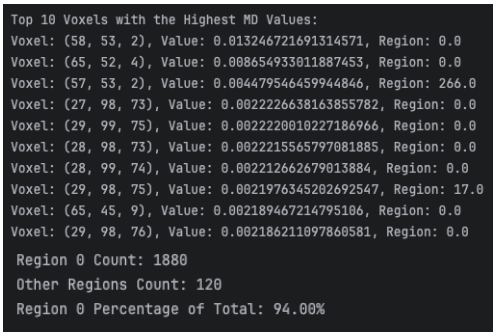

Figure 5: *Displayed here are the top 10 highest mean diffusivity (MD) values. The three highest MD values for the corresponding voxels appear abnormal, and I regard them as outliers. As observed in the previous distributions, the MD values decrease monotonically and at a uniform pace, leading me not to classify the remaining samples as outliers. Additionally, it is noteworthy that the majority of the highest values (94%) are attributed to region 0, which is considered a non-brain region.*

The attempt of applying transformation to the Mean Diffusivity (MD) data in the study was to address a couple of key issues: reducing noise and handling outliers. By averaging the MD value of each voxel with those of its immediate neighbors, excluding any neighbors in non-relevant areas (specifically, those in region 0), I aimed to smooth out the data. In this technique each voxel can have up to 26 neighboring voxels, and I thought it will help for reducing random fluctuations inherent in raw imaging data.

However, this process, aimed at both noise reduction and outlier management, didn't yield the expected outcomes. Despite the theoretical benefits of such smoothing—whereby the influence of outliers would be minimized through averaging with neighboring values—the results did not align with initial hopes and the operation was not significantly and I decided to continue to work on the data without any changes besides the removal of the 3 significant outliers for the highest values and for the lowest values, all the voxels that held MD values of 0 I changed them to 0.004 which is the lower limit of the distribution. Maybe in some time in the future another transformation will be apply on this data.

The transformation was applied based on the following equation:

$$MD\_VALUE_{\{adj\}}(x,y,z) = \left\{ \begin{array}{c} \frac{1}{N} \sum_{i,j,k \in neighbors} MD\_VALUE(i,j,k) \;\; if\ N > 0 \\ MD\_VALUE(x,y,z) \;\; otherwise \end{array} \right\}$$

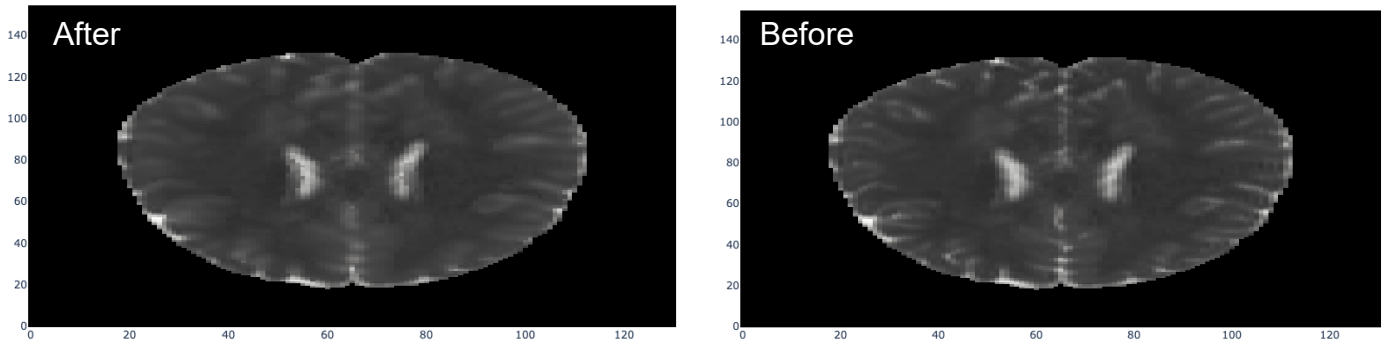

Figure 6: *After performing AB test I obtained the following results: the mean original MD value was 0.00017184, and the mean transformed MD value was slightly higher at 0.00017189. The statistical analysis, specifically a two-sample t-test, gave us a T-statistic of -0.1749414 and a P-value of 0.86112.*

The null hypothesis was that the transformation—averaging voxel values with their neighbors—would not significantly change the overall statistical properties of the MD dataset, essentially maintaining the mean MD value. Given the P-value of 0.86112, which is substantially higher than the conventional threshold of 0.05 used to determine statistical significance, the null hypothesis was not rejected. This indicates that the differences observed between the original and transformed datasets were not statistically significant.

For the next step, I proceeded with data identical to the original set, except for the modifications made to certain samples to remove outliers. For all samples with MD values of zero, I adjusted their values to 0.0004, which represents the lower bound of the distribution. For any samples with MD values exceeding 0.00225, I capped their values at this number, establishing it as the upper limit of the distribution.

After collecting all the necessary data, I computed the mean for each region with the aim of understanding the variations in mean MD values both between different regions and within subregions. Ultimately, my analysis extends to major regions. To illustrate these calculations and findings, I will describe them in detail, followed by visual representations in Figures 7, 8, 9, and 10.

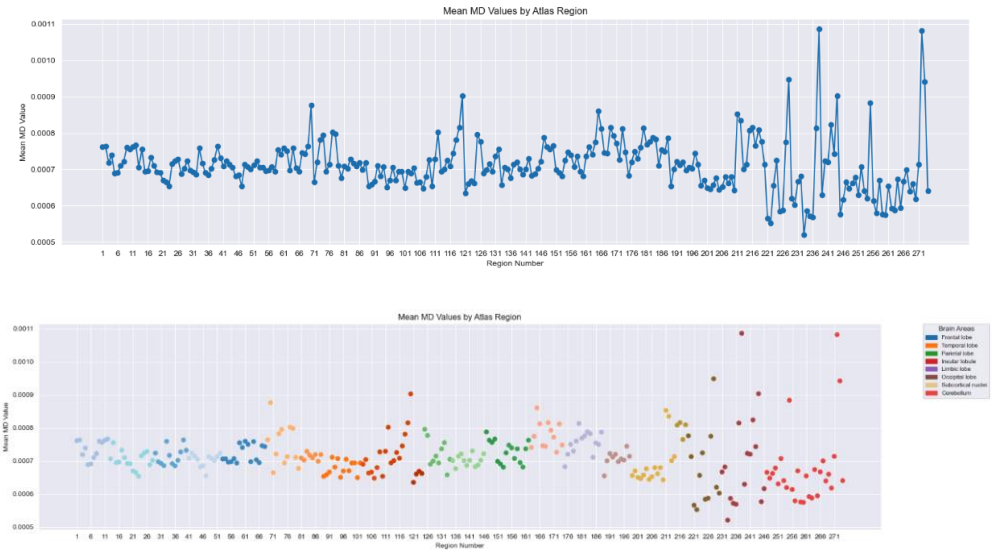

Figure 7: *As observed in the plots, most major regions display similar values for regional MD, approximately ranging from 0.0007 to 0.0008. Furthermore, it can be observed that the subcortical nuclei and cerebellum tend to exhibit slightly higher variability in terms of average MD values across their regions. In contrast, for the other major areas, the variance appears to be relatively small. In Figure X, we will see that some regions of the cerebellum could be considered outliers; therefore, the variance for the cerebellum may appear smaller than it actually is. Without delving into the biological aspects, it will be challenging to determine whether these regions are indeed outliers or if they serve an important function, resulting in higher MD values.*

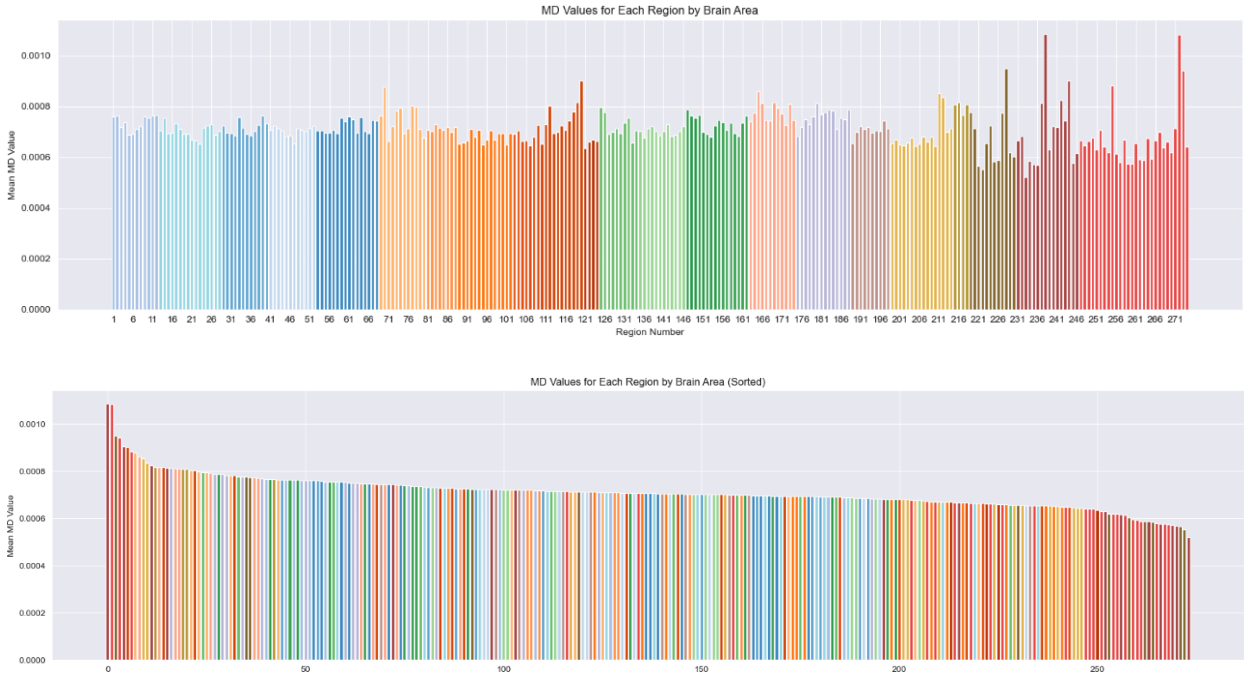

Figure 8: *Here we can observe essentially the same patterns as seen in Figure 7, but depicted in a histogram plot, with each bar representing a region. The primary reason for presenting this plot is to showcase the sorted arrangement of the regions, implying that there is no inherent order for the regions, subregions, or major regions. There is no established law governing the MD values of the regions based on this sorted plot.*

Figure 9: *In this figure, I computed the average for all the regions belonging to a specific subregion. Following the principles of bootstrapping, the law of large numbers, and probability theory, it appears that as I average the regions, the variance among the samples is reduced. Additionally, in this sorted plot, there is no governing law for the subregion areas.*

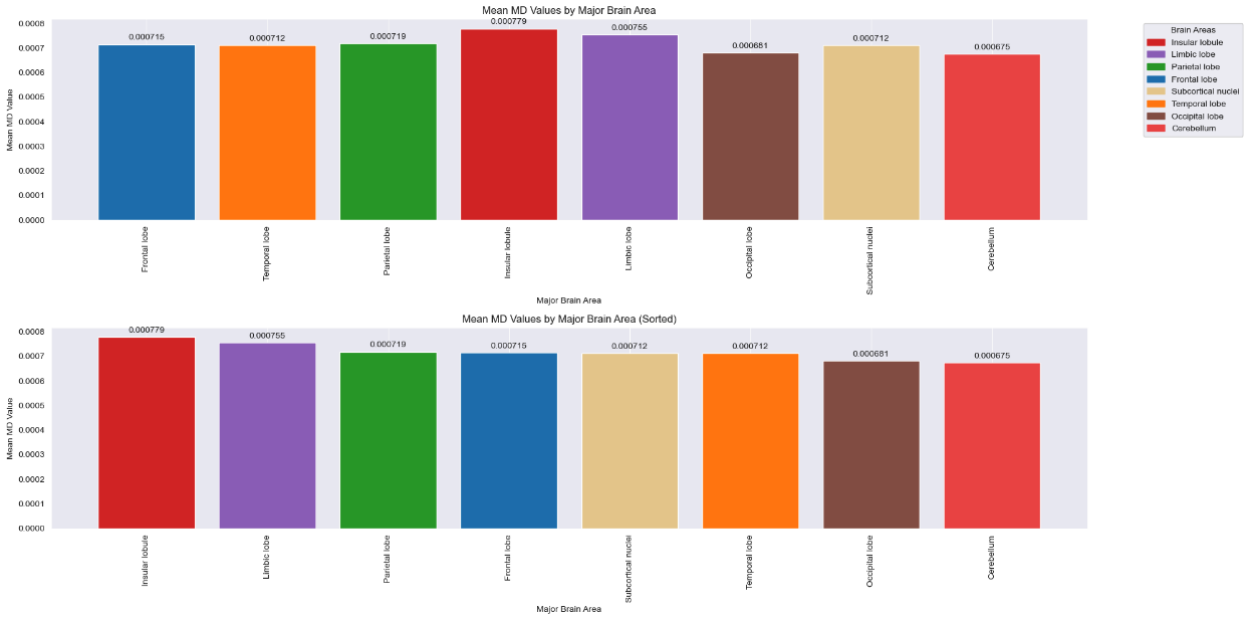

Figure 10: *Here we can observe the total average for each major brain area. An interesting observation from the sorted plot of the MD values for these major brain areas is a 15.4% gap between the average MD value of the Insular lobe and the Cerebellum. This discrepancy may indicate biological differences within these brain tissues, although further investigation is warranted.*

In this box plot, the samples represent the regions themselves rather than the subareas. This means that each variance and standard deviation value for each major region was calculated based on the MD values of the regions, not the MD values of the subareas.

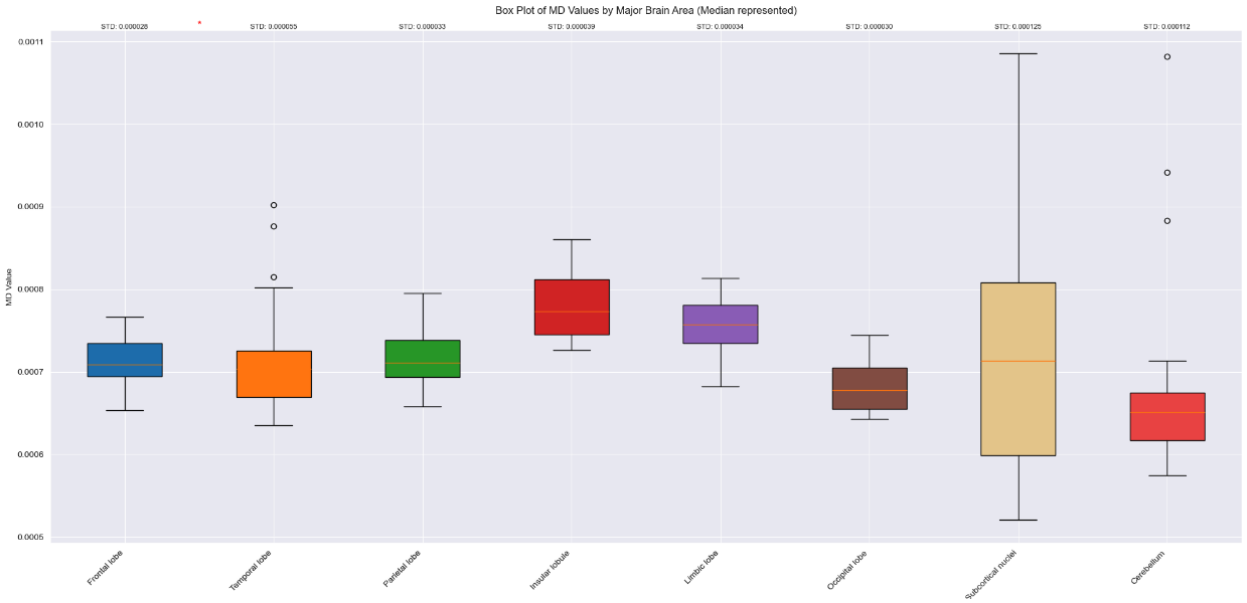

Figure 11: *When examining this box plot, we can observe relatively small variances and standard deviations among most of the major regions. The only exception is the subcortical nuclei, which exhibit relatively large variances among the samples. Additionally, significant outliers are apparent in the Cerebellum, leading to the observation that the standard deviation in the Cerebellum is relatively small, similar to the other major brain areas.*

To further enhance the project, I created a Flask web server that enables interactive visualization of the brain. This server utilizes the information and code developed throughout this assignment. Within the web server's user interface (UI), users are presented with two primary options for visualizing the brain:

1. Atlas-Based Visualization: This option allows users to view the brain according to spatial information derived from a brain atlas. It segments the brain into different regions or areas, each represented by distinct colors. This mode is particularly useful for understanding the structural organization and regional distinctions within the brain.
2. MD Value-Based Visualization: This mode focuses on plotting the brain based on Mean Diffusivity (MD) values. It offers a more quantitative view, highlighting variations in MD values across different brain regions. This visualization can be instrumental in identifying areas of interest based on their diffusivity properties, which can be indicative of underlying brain tissue characteristics or pathology.

The web server is designed to be user-friendly, allowing for easy switching between the two visualization modes. Whether users are interested in the anatomical layout of the brain or in exploring its diffusivity profile.

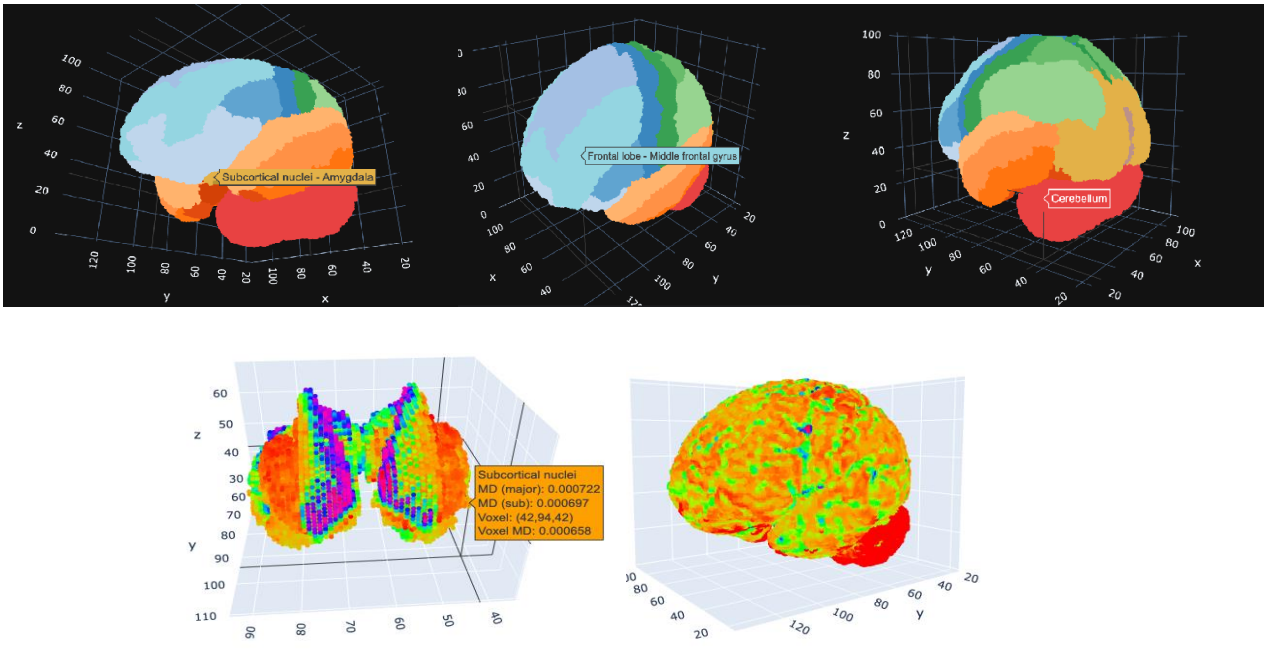

Figure 12: *The image presents 2 types of 3D brain visualization: one based on saptial devision into regions with distinct colors and another, displaying variation in the Mean Diffusivity (MD) value across the brain's anatomy using color gradient.*

Webserver can be found at: http://brain.freetruthworld.com/

GitHub Repo: https://github.com/galevy88/TAU_Assignment