

ENHANCING BRAIN CONNECTIVITY MAPPING: A RADIOMICS APPROACH TO DIFFUSION MRI TRACTOGRAPHY

LEVENTE ZSOLT NAGY

Thesis supervisor
ALFREDO VELLIDO ALCACENA (Department of Computer Science)

Thesis co-supervisor
ESTELA CAMARA MANCHA (Hospital Universitari de Bellvitge)

Degree
Master's Degree in Artificial Intelligence

Master's thesis

School of Engineering
Universitat Rovira i Virgili (URV)

Faculty of Mathematics
Universitat de Barcelona (UB)

Barcelona School of Informatics (FIB)
Universitat Politècnica de Catalunya (UPC) - BarcelonaTech

Abstract

This thesis about medical imaging aims to find alternative ways to map brain connectivity, utilizing the T1 and T2 MRI images instead of the diffusion MRI image, vastly improving the cost and time efficiency of the process. As a replacement for tractograpy, the currently used and accepted tool for processing the diffusion images, this thesis will reveal if there are any simple or complex relationship between radiomic features of the T1 and T2 images of the brain regarding the connected regions. The results and conclusions are limited to the connectivity of the basal ganglia to other main cortical regions of the brain. It is in no way a generalized conclusion, but rather a proof of concept from experiments ran on a specific dataset, provided by Hospital Universitari de Bellvitge.

Contents

List of Notations & Abbreviations

MRI magnetic resonance imaging 6

dMRI diffusion magnetic resonance imaging 6

FA fractional anisotropy 6

MD mean diffusivity 6

RD radial diffusivity 6

ROI region of interest 6

FNN fully connected neural network 15

NIfTI Neuroimaging Informatics Technology Initiative 9

GLCM Gray Level Co-occurrence Matrix 11

GLSZM Gray Level Size Zone Matrix 11

GLRLM Gray Level Run Length Matrix 11

NGTDM Neighbouring Gray Tone Difference Matrix 11

GLDM Gray Level Dependence Matrix 11

List of Figures

List of Tables

Introduction

Basal ganglia is a part of the human brain which is group of subcortical nuclei responsible primarily for motor control, as well as other roles such as motor learning, executive functions and behaviors, and emotions. [1] Huntington's disease is a disorder that causes the progressive degeneration of the basal nuclei. [2]

Hospital de Bellvitge provided an excellent dataset of magnetic resonance imaging (MRI) and diffusion magnetic resonance imaging (dMRI) records. This dataset contains 32 control and 37 Huntington patient records of T1 and T1/T2 MRI images with isotropic voxels of 1 millimeter resolution and dMRI fractional anisotropy (FA), mean diffusivity (MD) and radial diffusivity (RD) images with isotropic voxels of 2 millimeter resolution and 1 second temporal resolution. Furthermore this dataset also contains the mask for the basal ganglia, which will also be referenced as the region of interest (ROI). And taking inspiration from this paper [3], masks for the 7 main cortical regions of the brain, which will also be referenced as the target regions: Limbic, Executive, Rostral-Motor, Caudal-Motor, Parietal, Occipital and Temporal are also included in the dataset. Tractography was performed on the dMRI images to figure out which parts of the ROI are connected to which cortical target, in a similar manner to how it was done in said paper [3]; where the relative connectivity maps are representing the ratio of the number of streamlines to each cortical target. Furthermore, the raw streamline images are also available, where there are a maximum of 5000 streamlines from each voxel in the ROI. The anatomical segmentation of the Basal Ganglia is also available, for the Caudate Putamen and Accumbens.

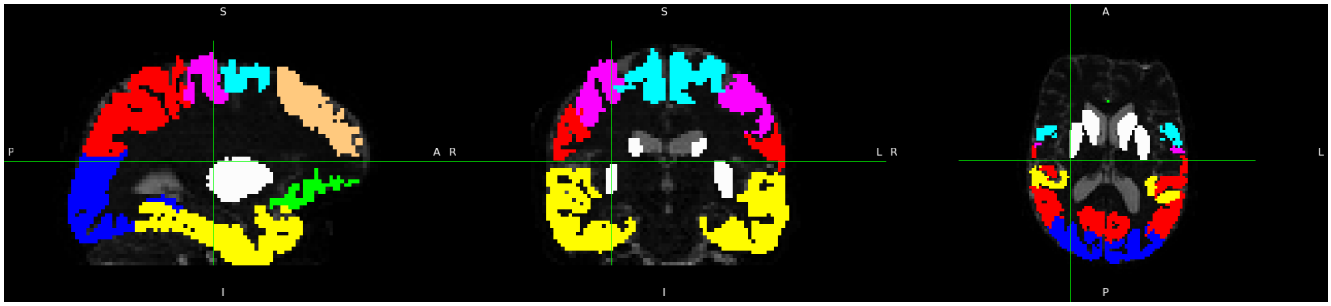

Figure 1.1: Basal Ganglia (ROI) & Cortical Targets

Color	Region
White	Basal Ganglia (ROI)
Green	Limbic
Brown	Executive
Light Blue	Rostral-Motor
Purple	Caudal-Motor
Red	Parietal
Blue	Occipital
Yellow	Temporal

Table 1.1: Regions Legend

Furthermore, for both the ROI and cortical targets, the dataset distinguishes between the right and left halves of the brain. Thus there are actually 2 ROIs and $2 \cdot 7 = 14$ target regions.

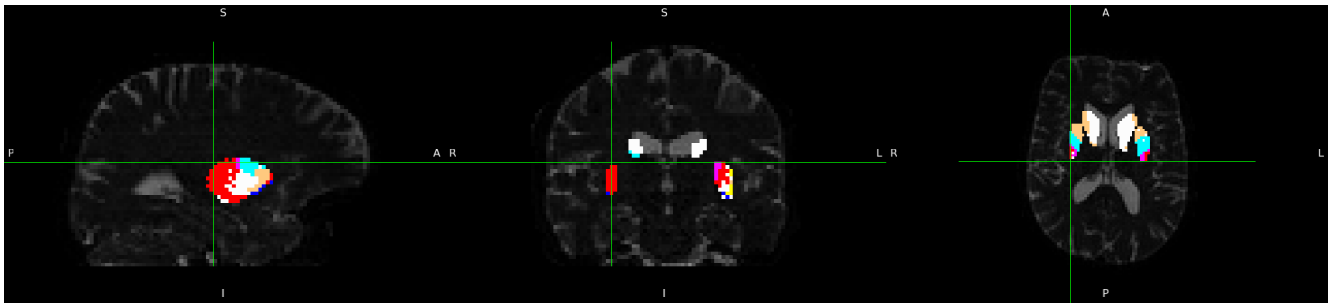

Figure 1.2: Connectivity Maps

1.1 Objectives

The end goal is to predict the relative connectivity of the Basal Ganglia to the cortical targets, from the T1 and T1/T2 images.

This being a very complex problem, there is the possibility that the correlation between the connectivity of the brain and the T1, T1/T2 images are too weak to be mapped on this dataset. As from a datascience perspective, 69 datapoints are not much. But from a medical perspective it is substantial as it is very hard to collect uniform, clean data, with permissions to use it for research.

As a simpler task, leading up to the complex end goal, is a model for the simple segmentation of the Basal Ganglia for the regions Caudate Putamen and Accumbens. In order to confirm that the radiomics texture of the T1 and T1/T2 images of this dataset are are correlated to the anatomical segmentation of the Basal Ganglia. This problem is inherently connected to the main goal, as the relative connectivity does obey certain anatomical restrictions, and the anatomical segmentation of the Basal Ganglia is confirmed to be related to the relative connectivity. Thus if this simpler prediction fails, there is a good chance that the complex end goal will fail as well.

The biggest obstacle of this project is the preprocessing of the data, as there are many variations and hyperparameters that can be tuned. An exhaustive search definetly will not be viable, thus the preprocessing and model will needed to be tuned in a waterfall like manner, making educated guesses and comparing model performances across different tries. The main metric to measure

model performance, will be the accuracy of the label prediction across voxels, as it should be comparable between all approaches.

1.2 Motivation

The motivation for predicting the connectivity maps from the T1 and T1/T2 MRI images, is skipping the time and resource consuming process performing dMRI and tractorgrapy.

1.3 State of the Art

Me :)

Design

2.1 Preprocessing

2.1.1 Raw Data

All provided data are in the Neuroimaging Informatics Technology Initiative (NIfTI) format, first these are need to be understood and parsed. This format stores the raw output of the MRI record, and additionally an affine transformation matrix which can transform this raw space in order to align them with each other.

2.1.1.a Available Data

The following data will be preprocessed and read, even if not all of them are going to be used later on it helps providing the largest possible flexibility.

TODO squigly line

Data	**Shape**	**Range**	**Type**	**Space**	**Reference**
dMRI	(118, 118, 60, 74)	$[0, 3000]$	float	diffusion_1	diffusion
Diffusion FA	(118, 118, 60)	$[0,\ 2]$	float	diffusion_2	diffusion_fa
Diffusion MD	(118, 118, 60)	$[0,\ 0.01]$	float	diffusion_2	diffusion_md
Diffusion RD	(118, 118, 60)	$[0,\ 0.01]$	float	diffusion_2	diffusion_rd
T1	(208, 256, 256)	$[0,\ 1000]$	float	t1_1	t1
T1/T2	(208, 256, 256)	$[0, 1]$	float	t1_2	t1t2
Cortical Targets	(118, 118, 60, 14)	$\{0, 1\}$	bool	diffusion_1	targets
Relative Connectivity	(118, 118, 60, 14)	$[0, 1]$	float	diffusion_1	connectivity
Streamline Image	(118, 118, 60, 14)	$[0, 5000]$	uint	diffusion_1	streamline
ROI Mask (Basal Ganglia)	(118, 118, 60, 2)	$\{0, 1\}$	bool	diffusion_1	mask_basal & roi
Brain Mask	(208, 256, 256)	$\{0, 1\}$	bool	t1_1	mask_brain
Basal Ganglia Segmentation	(208, 256, 256)	$[0, 58]$	uint	t1_1	mask_basal_seg

Table 2.1: Raw Datapoint

2.1.1.b Registration

The process of aligning different records into the same native space is called "registration". The provided dataset comes with with 4 different spaces, earlier referenced to as t1_1, t1_2, diffusion_1 and diffusion_2. Most of the data are in diffusion_1 space, thus it is logical to register the rest into the same space. The registration is done with Dipy TODO REFERENCE bla bla.

2.1.1.c Brain Mask

The provided dataset did not apply the brain masks for the T1 images out of the box so it can be done with a simple element wise multiplication.

2.1.1.d Native Space

Transforming two different readings into the same native space requires a bit of math. As NIfTI format stores the origin of the data at an arbitrary location in the voxel space. After applying the extracted transformation matrices, the records will line up, but the origin of the voxel space will be at somewhere inside it, instead of at $(0, 0, 0)$. Meaning that only the first quadrant will be visible of the record, thus the space is also needed to be translated with the negative vector of the transformed space's bounding box's lower end.

The translation value can be calculated by calculating the boundaries of the transformed space's bounding box. Get all 8 corners of the voxel space and apply the transformation matrix to all of them. Then get the min-max coordinates along X, Y and Z from the 8 transformed vectors, yielding the lower and upper bounds of the transformed space's bounding box.

It is very important to use the same translation value across different raw spaces to properly align them in the native space. For example let D and T denote a diffusion and t1 records and M_D and M_T denote their respective transformation matrices. Let T_D and T_T denote their respective translation values. In order to properly align them we need to apply $A_D = (M_D \cdot T_D)$ matrix and $A_T = (M_T \cdot T_D)$ matrix to D and T respectively, with matching T_D translation values.

The last issue is the missaligned new shapes of the T1 and Diffusion records. This can be simply fixed by truncating the excess along each dimension.

2.1.1.e Uniform Shape

After aligning the data into the same space per datapoint, it is still very likely that the individual datapoints do not have a uniform shape. This is due to them being in native space, some records will contain a smaller volume brain, some will contain a larger, they will not be the same.

Fixing this can be done by figuring out the min-max boundaries along each axis that the brain masks take up in the voxel space. Then the range of the masks along each axis can be calculated from the lower and upper boundaries per datapoint. And then the max range can be selected per axis, across all datapoints, yielding the new uniform shape. Finally, the voxel spaces can be sliced down to the new uniform shape, which can fit all brains of all data points (with some padding for most of them).

Note that this fix also greatly improves space efficiency, as it cuts out the unused voxels. This will be beneficial for storage and computational demands of future experiments.

Data	Shape	Range	Type
diffusion	(70, 153, 218, 157, 74)	$[0, 4096]$	float16
t1	(70, 153, 218, 157, 1)	$[0, 947]$	float16
roi	(70, 153, 218, 157, 2)	$\{0, 1\}$	bool
targets	(70, 153, 218, 157, 14)	$\{0, 1\}$	bool
connectivity	(70, 153, 218, 157, 14)	$[0, 1]$	float16
diffusion_mask	(70, 153, 218, 157, 1)	$\{0, 1\}$	bool
t1_mask	(70, 153, 218, 157, 1)	$\{0, 1\}$	bool

Table 2.2: Uniform Data

Note that the new shapes are all 5 dimensional, where the first dim is for the datapoint index. The next 3 is for the coordinates of the voxels. And the last is for any additional information, like the temporal dimension of the dMRI or the target masks or the connectivity labels.

2.1.2 Radiomics Features

Extracting the voxel based radiomic features has two main parameters to tune, the bin width and the kernel width.

The two main approaches for binning are absolute discretization and relative discretization. Where in the prior one, a fixed bin width is choosen and in the latter one, a fixed number of bins are chosen and the bin width scales relatively according to the min-max voxel values. This study found that "The absolute discretization consistently provided statistically significantly more reproducible features than the relative discretization." [4] Relying on this information, the obvious choice to start with is the absolute discretization.

The bin width and the kernel width will be tuned in later experiments. And possibly features calculated with different setting will be concatenated and used simultaneously for better results. The used default values will be 25 and 5 for the bin and kernel widths respectively.

The following types of radiomic features will be used:

Feature Type	Number of Features
First Order	18
Gray Level Co-occurrence Matrix (GLCM)	23
Gray Level Size Zone Matrix (GLSZM)	16
Gray Level Run Length Matrix (GLRLM)	16
Neighbouring Gray Tone Difference Matrix (NGTDM)	5
Gray Level Dependence Matrix (GLDM)	14
3D Shape	17

Table 2.3: Radiomic Feature Types

2.1.2.a Voxel Based

The following 92 features will be calculated voxel based:

First Order	**GLCM**	**GLSZM**
Energy	Autocorrelation	SmallAreaEmphasis
TotalEnergy	JointAverage	LargeAreaEmphasis
Entropy	ClusterProminence	GrayLevelNonUniformity
Minimum	ClusterShade	GrayLevelNonUniformityNormalized
10Percentile	ClusterTendency	SizeZoneNonUniformity
90Percentile	Contrast	SizeZoneNonUniformityNormalized
Maximum	Correlation	ZonePercentage
Mean	DifferenceAverage	GrayLevelVariance
Median	DifferenceEntropy	ZoneVariance
InterquartileRange	DifferenceVariance	ZoneEntropy
Range	JointEnergy	LowGrayLevelZoneEmphasis
MeanAbsoluteDeviation	JointEntropy	HighGrayLevelZoneEmphasis
RobustMeanAbsoluteDeviation	Imc1	SmallAreaLowGrayLevelEmphasis
RootMeanSquared	Imc2	SmallAreaHighGrayLevelEmphasis
Skewness	Idm	LargeAreaLowGrayLevelEmphasis
Kurtosis	MCC	LargeAreaHighGrayLevelEmphasis
Variance	Idmn	
Uniformity	Id	
	Idn	
	InverseVariance	
	MaximumProbability	
	SumEntropy	
	SumSquares	

GLRLM	**NGTDM**	**GLDM**
ShortRunEmphasis	Coarseness	SmallDependenceEmphasis
LongRunEmphasis	Contrast	LargeDependenceEmphasis
GrayLevelNonUniformity	Busyness	GrayLevelNonUniformity
GrayLevelNonUniformityNormalized	Complexity	DependenceNonUniformity
RunLengthNonUniformity	Strength	DependenceNonUniformityNormalized
RunLengthNonUniformityNormalized		GrayLevelVariance
RunPercentage		DependenceVariance
GrayLevelVariance		DependenceEntropy
RunVariance		LowGrayLevelEmphasis
RunEntropy		HighGrayLevelEmphasis
LowGrayLevelRunEmphasis		SmallDependenceLowGrayLevelEmphasis
HighGrayLevelRunEmphasis		SmallDependenceHighGrayLevelEmphasis
ShortRunLowGrayLevelEmphasis		LargeDependenceLowGrayLevelEmphasis
ShortRunHighGrayLevelEmphasis		LargeDependenceHighGrayLevelEmphasis
LongRunLowGrayLevelEmphasis		
LongRunHighGrayLevelEmphasis		

Table 2.4: Voxel Based Radiomic Features

Normalization

2.1.2.b Non-Voxel Based

3D Shape
MeshVolume
VoxelVolume
SurfaceArea
SurfaceVolumeRatio
Sphericity
Maximum3DDiameter
Maximum2DDiameterSlice
Maximum2DDiameterColumn
Maximum2DDiameterRow
MajorAxisLength
MinorAxisLength
LeastAxisLength
Elongation
Flatness

Table 2.5: Shape Based Radiomic Features

2.1.3 All Data

Data	**Shape**	**Range**	**Type**
diffusion	(70, 153, 218, 157, 74)	$[0, 4096]$	float16
t1	(70, 153, 218, 157, 1)	$[0, 947]$	float16
roi	(70, 153, 218, 157, 2)	$\{0, 1\}$	bool
targets	(70, 153, 218, 157, 14)	$\{0, 1\}$	bool
connectivity	(70, 153, 218, 157, 14)	$[0, 1]$	float16
diffusion_mask	(70, 153, 218, 157, 1)	$\{0, 1\}$	bool
t1_mask	(70, 153, 218, 157, 1)	$\{0, 1\}$	bool
radiomics_raw	(70, 153, 218, 157, 92)	$[-363, 13979179]$	float32
radiomics_scaled	(70, 153, 218, 157, 92)	$[0, 1]$	float16

Table 2.6: Voxel Data

Data	**Shape**	**Range**	**Type**
radiomics_brain_raw	(70, 106, 1)	$[-63, 79581700189]$	float64
radiomics_roi_raw	(70, 106, 2)	$[-70, 1078755564]$	float64
radiomics_targets_raw	(70, 106, 14)	$[-124, 4799594982]$	float64
radiomics_brain	(70, 106, 1)	$[0, 1]$	float16
radiomics_roi	(70, 106, 2)	$[0, 1]$	float16
radiomics_targets	(70, 106, 14)	$[0, 1]$	float16

Table 2.7: Non-Voxel Radiomics Data

Data	**Length**	**Type**
radiomics_features_vox	92	string
radiomics_log10_features_vox	57	string
radiomics_features	106	string

Table 2.8: Features Names Data

Experiments

First the initial generic hyper parameters were established for all models, which can be shared between different architectures.

Parameter	Value
radiomics_brain_raw	(70, 106, 1)

Table 3.1: Generic Initial Hyperparameters

3.1 Classification FNN

First the starter hyper parameters were established for the classification fully connected neural network (FNN) models.

3.1.1 Simple 1

Sources of Information

[1] José L Lanciego, Natasha Luquin, and José Obeso. "Functional neuroanatomy of the basal ganglia". In: *Cold Spring Harbor perspectives in medicine* (2012). URL: https://doi.org/10.1101/cshperspect.a009621.

[2] Olivia C Matz and Muhammad Spocter. "The Effect of Huntington's Disease on the Basal Nuclei". In: *Cureus* (2022). URL: https://doi.org/10.7759/cureus.24473.

[3] Hyungyou Park et al. "Aberrant cortico-striatal white matter connectivity and associated subregional microstructure of the striatum in obsessive-compulsive disorder". In: *Molecular Psychiatry* (2022). URL: https://doi.org/10.1038/s41380-022-01588-6.

[4] Loïc Duron et al. "Gray-level discretization impacts reproducible MRI radiomics texture features". In: *PLoS One* (2019). URL: https://doi.org/10.1371/journal.pone.0213459.