

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 MRI image 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 image of the brain regarding the connected regions. It will also dive into fully convolutional approaches, in order to process the said T1 image. 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

ROI region of interest 6

FNN fully connected neural network 8

FCNN fully convolutional neural network 7

NIfTI Neuroimaging Informatics Technology Initiative 7

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 35 control and 45 Huntington records of already registered [3] T1 MRI images with isotropic voxels of 1 millimeter resolution and dMRI 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 [4], 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 [4]; where the connectivity maps are representing the probability of connections with each cortical target, which were thresholded at the 50th percentile to minimize overlapping voxels between each segment of the ROI.

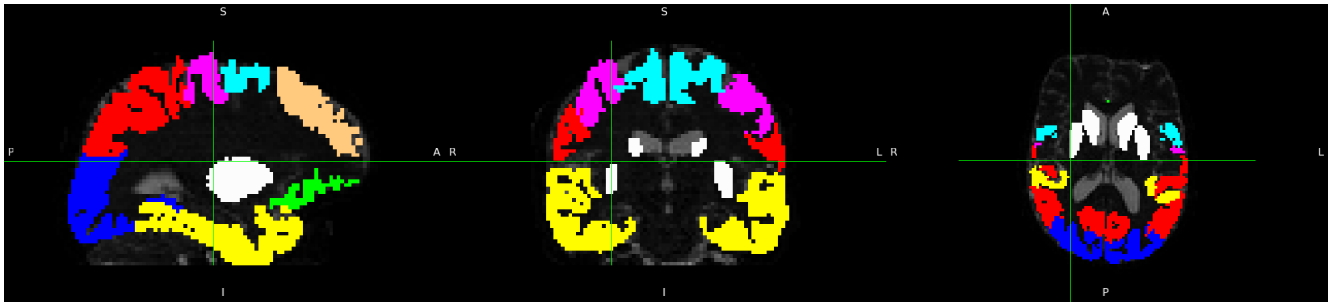

Figure 1.1: Basal Ganglia (ROI) & Cortical Targets

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.

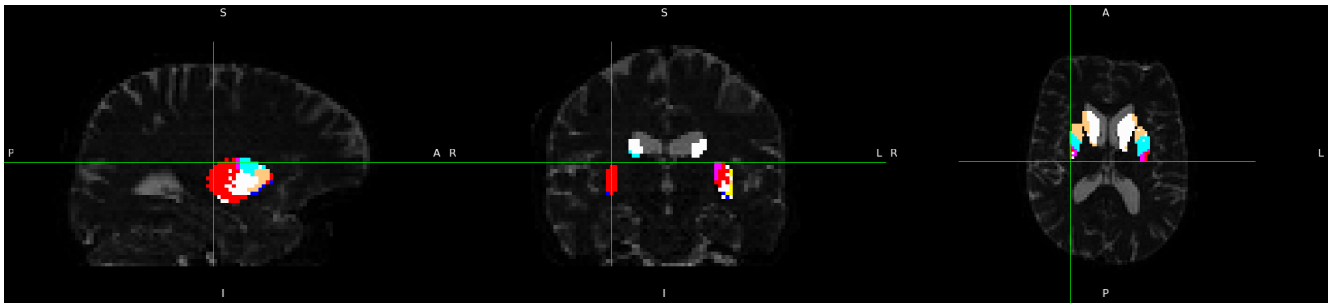

Figure 1.2: Connectivity Maps

The goal is to extract the connectivity maps from the T1 MRI images, skipping the time and resource consuming process performing dMRI and tractorgraphy. Two approaches will be performed in this thesis, trying to extract the connectivity maps from radiomic [5] features, and trying to extract the connectivity maps with a fully convolutional neural network (FCNN).

All provided files in the dataset are in Neuroimaging Informatics Technology Initiative (NIfTI) format. And all of them are registered in the same anatomical space.

Requirements

The following requirements were defined based on the objectives of this thesis:

1. Preprocess the data.
2. Extract radiomics features.
3. Observe direct correlations between extracted features and connectivity maps.
4. Design a decision tree for predicting the connectivity maps from the extracted features.
5. Design an fully connected neural network (FNN) for predicting the connectivity maps from the extracted features.
6. Grid search radiomics feature extraction parameters for the best results.
7. Design an FCNN for predicting the connectivity maps from the T1 image.
8. Modify the FCNN by appending the ROI mask to the input.
9. Modify the FCNN by appending the cortical target masks to the input.
10. Design a generalized FCNN for predicting one connectivity map at a time from the T1 image, ROI & target masks.

Preprocessing

3.1 Raw Data

All provided data are in the NIfTI format, first these are need to be understood and parsed. This format stores the raw output of the MRI record, and additionally a transformation matrix which can translate this raw space into anatomical space. The process of aligning different records into the same anatomical space is called "registration". The provided dataset's T1 MRI and dMRI records are already registered in the same space.

3.1.1 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 later on.

Data	**Shape**	**Range**	**Type**	**Space**	**Reference**
dMRI	(118, 118, 60, 74)	$[0, 2492]$	float16	diffusion	diffusion
T1	(208, 256, 256)	$[0, 629]$	float16	t1	t1
ROI Mask (Basal Ganglia)	(118, 118, 60, 2)	$\{0, 1\}$	bool	diffusion	roi
Cortical Targets	(118, 118, 60, 14)	$\{0, 1\}$	bool	diffusion	targets
Relative Connectivity	(118, 118, 60, 14)	$[0, 1]$	float16	diffusion	connectivity
dMRI Brain Mask	(118, 118, 60)	$\{0, 1\}$	bool	diffusion	diffusion_mask
T1 Brain Mask	(208, 256, 256)	$\{0, 1\}$	bool	t1	t1_mask

Table 3.1: Raw Datapoint

3.1.2 Brain Mask

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

3.1.3 Anatomical Space

Transforming two different readings into the same anatomical space still requires a bit of math. As after applying the extracted transformation matrices, the records will line up, but the center of the voxel space will be at $(0, 0, 0)$. Meaning that technically 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 anatomical 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.

3.1.4 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 registered into the same space datapoint wise, but can have differrent registrations across multiple datapoints.

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.

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	(80, 150, 186, 124, 74)	$[0, 2492]$	float16
t1	(80, 150, 186, 124, 1)	$[0, 629]$	float16
roi	(80, 150, 186, 124, 2)	$\{0, 1\}$	bool
targets	(80, 150, 186, 124, 14)	$\{0, 1\}$	bool
connectivity	(80, 150, 186, 124, 14)	$[0, 1]$	float16
diffusion_mask	(80, 150, 186, 124, 1)	$\{0, 1\}$	bool
t1_mask	(80, 150, 186, 124, 1)	$\{0, 1\}$	bool

Table 3.2: 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.

3.2 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." [6] Relying on this information, the obvious choice 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 3.3: Radiomic Feature Types

3.2.1 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 3.4: Voxel Based Radiomic Features

3.2.1.a Normalization

3.2.2 Non-Voxel Based

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

Table 3.5: Shape Based Radiomic Features

3.3 All Data

Data	Shape	Range	Type
diffusion	(80, 150, 186, 124, 74)	$[0, 2492]$	float16
t1	(80, 150, 186, 124, 1)	$[0, 629]$	float16
roi	(80, 150, 186, 124, 2)	$\{0, 1\}$	bool
targets	(80, 150, 186, 124, 14)	$\{0, 1\}$	bool
connectivity	(80, 150, 186, 124, 14)	$[0, 1]$	float16
diffusion_mask	(80, 150, 186, 124, 1)	$\{0, 1\}$	bool
t1_mask	(80, 150, 186, 124, 1)	$\{0, 1\}$	bool
radiomics_raw	(80, 150, 186, 124, 92)	$[-363, 13979179]$	float32
radiomics_scaled	(80, 150, 186, 124, 92)	$[0, 1]$	float16

Table 3.6: Voxel Data

Data	Shape	Range	Type
radiomics_brain_raw	(80, 106, 1)	$[-63, 79581700189]$	float64
radiomics_roi_raw	(80, 106, 2)	$[-70, 1078755564]$	float64
radiomics_targets_raw	(80, 106, 14)	$[-124, 4799594982]$	float64
radiomics_brain	(80, 106, 1)	$[0, 1]$	float16
radiomics_roi	(80, 106, 2)	$[0, 1]$	float16
radiomics_targets	(80, 106, 14)	$[0, 1]$	float16

Table 3.7: Non-Voxel Radiomics Data

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

Table 3.8: Features Names Data

Models

Results

Implementation

Conclusions

Future Imporvements

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] Peter J Kostelec and Senthil Periaswamy. "Image registration for MRI". In: *Modern signal processing* 46 (2003), pp. 161–184. URL: https://citeseerx.ist.psu.edu/document?doi=bbe2c10f61acdb4d13b9f6df3f0f3167ddd097a0.

[4] 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.

[5] Marius E Mayerhoefer et al. "Introduction to radiomics". In: *Journal of Nuclear Medicine* 61.4 (2020), pp. 488–495. URL: https://jnm.snmjournals.org/content/jnumed/61/4/488.full.pdf.

[6] 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.