
# PREDICTING BRAIN CONNECTIVITY MAPPING USING RADIOMICS FEATURES IN ANATOMICAL MRI

**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 study explores alternatives to diffusion MRI for mapping brain connectivity, predicting fractional anisotropy and mean diffusivity, using radiomic features derived from T1 and T2 structural MRI images. This approach aims to significantly enhance the cost and time efficiency of data acquisition, eliminating the need for diffusion MRI and tractography. The research is centered on the basal ganglia, a region primarily affected by neurodegeneration in Huntington's disease, comparing its characteristics between control subjects and patients with the condition.*

# 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

**NN** neural network . . . . . 13

**FNN** feedforward neural network . . . . . 20

**CNN** convolutional neural network . . . . . 13

**FCNN** fully convolutional neural network . . . . . 12

**NIfTI** neuroimaging informatics technology initiative . . . . . 9

**FMRIB** functional magnetic resonance imaging of the brain

**FNIRT** FMRIB's nonlinear image registration tool . . . . . 6

**GLCM** gray level co-occurrence matrix . . . . . 14

**GLSZM** gray level size zone matrix . . . . . 14

**GLRLM** gray level run length matrix . . . . . 14

**NGTDM** neighbouring gray tone difference matrix . . . . . 14

**GLDM** gray level dependence matrix . . . . . 14

**cUHDRS** composite Unified Huntington's Disease Rating Scale . . . . . 17

**CAP** CAG Age Product . . . . . 17

# List of Figures

# List of Tables

# Introduction

Basal ganglia is a part of the human brain which is a 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 of 32 control and 38 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. Furthermore this dataset also contains the mask for the basal ganglia, which will also be referenced as the region of interest (ROI). 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 this 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 subcortical segmentation of the Basal Ganglia is also available, for the Caudate, Putamen and Accumbens on the control records. And lastly FMRIB's nonlinear image registration tool (FNIRT) warp fields were also provided for converting the records into normalized space.

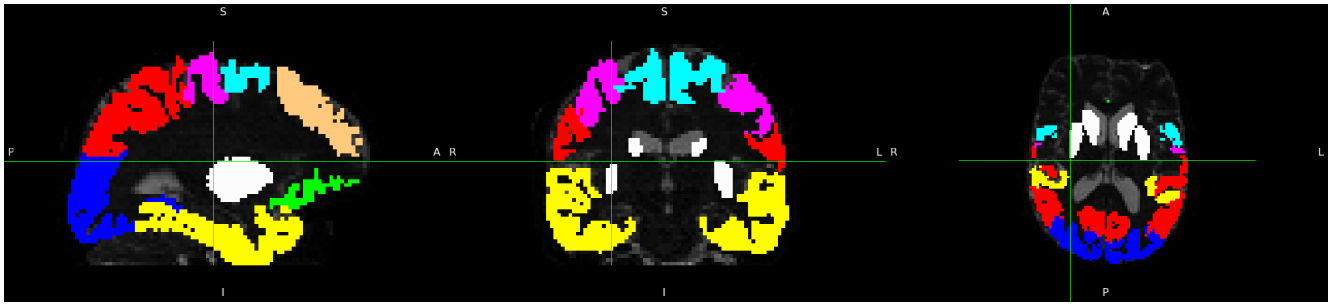


Figure 1.1: Basal Ganglia (ROI) & Cortical Targets

| 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 hemispheres of the brain. Thus there are actually 2 ROIs and $2 \cdot 7 = 14$ target regions.

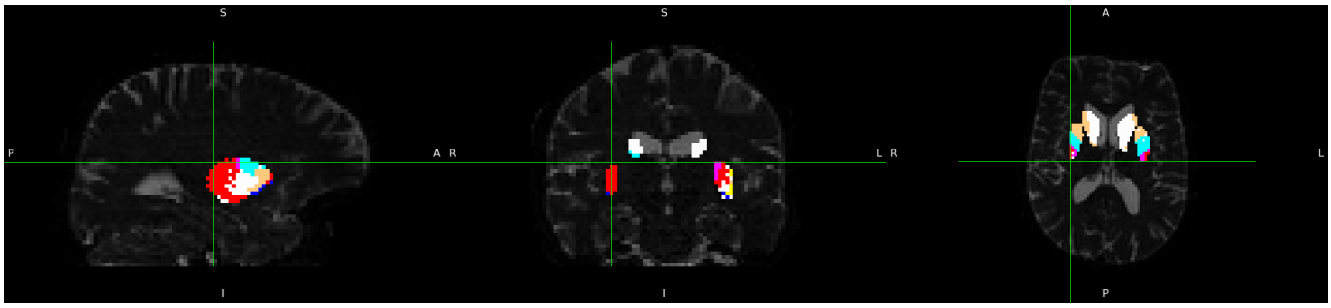


Figure 1.2: Connectivity Maps

## 1.1 Objectives

The end goal is to predict the relative connectivity of the Basal Ganglia to the cortical targets, from the radiomics features of 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, 70 records 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.

A simpler task leading up to the complex end goal, is a model for the simple segmentation of the Basal Ganglia for the subcortical regions Caudate, Putamen and Accumbens. In order to confirm that the radiomics texture of the T1 and T1/T2 images of this dataset are correlated to the 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 subcortical 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.

Another intermediate task, is a model for predicting FA and MD images. This is also related to the main goal, as these images are computed from the dMRI images, the same image that the relative connectivity is computed from. But it is inherently simpler, not needing to perform complex algorithms like tractography.

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 definitely 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. And pearson correlation will be used as the metric to evaluate the FA and MD regression predictions.

## 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

# Design

In order to understand some of the following design choices, it makes sense to establish it early that the model will be operating on extracted voxel based features and non-voxel based features, and will predict on a voxel by voxel level.

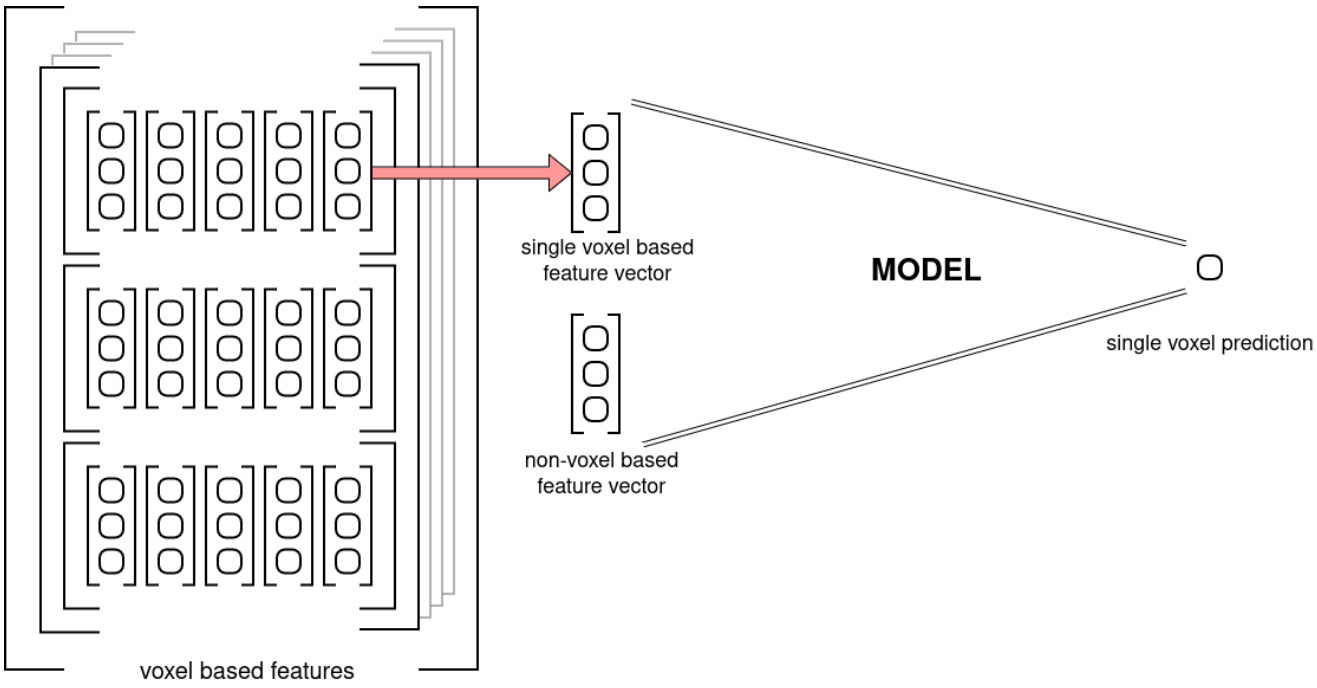


Figure 2.1: Simple Model Overview

**This report will reference to the control/patient spatial data as 'Record'** *("voxel based features" in Figure 2.1)* **and will reference to the individual feature vectors as 'Datapoint'** *("single voxel based feature vector" and "non-voxel based feature vector" in Figure 2.1)***.** This logical differentiation is needed, as the model only operates on Datapoints and has no global context available, while some preprocessing and evaluation logic should happen on a Record level.

## 2.1 Preprocessing

### 2.1.1 Raw Data

All provided records 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 used for aligning different spaces.

#### 2.1.1.a Available Data

The following records 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.

| Data | Shape | Range | Type | Space | Reference |
|---|---|---|---|---|---|
| dMRI | (118, 118, 60, 74) | $[0, 4096]$ | uint | diffusion | diffusion |
| Diffusion FA | (118, 118, 60) | $[0,\ 2]$ | float | diffusion | diffusion_fa |
| Diffusion MD | (118, 118, 60) | $[0,\ 0.01]$ | float | diffusion | diffusion_md |
| Diffusion RD | (118, 118, 60) | $[0,\ 0.01]$ | float | diffusion | diffusion_rd |
| T1 | (208, 256, 256) | $[0,\ 1000]$ | float | t1 | t1 |
| T1/T2 | (208, 256, 256) | $[0, 1]$ | float | d_aligned | t1t2 |
| Cortical Targets | (118, 118, 60, 14) | $\{0, 1\}$ | bool | diffusion | targets |
| Relative Connectivity | (118, 118, 60, 14) | $[0, 1]$ | float | diffusion | connectivity |
| Streamline Image | (118, 118, 60, 14) | $[0, 5000]$ | uint | diffusion | streamline |
| ROI Mask (Basal Ganglia) | (118, 118, 60, 2) | $\{0, 1\}$ | bool | diffusion | mask_basal & roi |
| Brain Mask | (208, 256, 256) | $\{0, 1\}$ | bool | t1 | mask_brain |
| Basal Ganglia Segmentation | (208, 256, 256) | $[0, 58]$ | uint | t1 | basal_seg |

Table 2.1: Raw Data

#### 2.1.1.b 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 of the T1 image and T1 mask.

#### 2.1.1.c Registration

The process of aligning different records into the same native space is called "registration". The provided dataset comes with with 2 (3) different spaces, earlier referenced to as t1 and diffusion (and d_aligned). Most of the data are in diffusion space, thus it is logical to register the rest into the same space. After manual inspection, only 15 records required registration. Out of which 3 only required a tiny translation, and the rest 12 needed a complete affine registration.

The image T1/T2 is the odd one out, as it is inherently in a different space from diffusion (due to them being different resolution). But they are aligned into diffusion space. Although they do not need to be registered, this has to be taken into account later on.

#### 2.1.1.d Normalization

The process of warping each brain into a common space is called "normalization". Applying the FNIRT warp fields are more or less straight forward, as two warp fields are provided, one for the diffusion space and one for the T1 space. Note that this process inherently contains the benefits of registration, as it is warping the different images into a common brain shape and space. This also paves the direction of future experiments, as it opens the door to working in either native and normalized space.

The only encountered obstacle was with the T1/T2 image. As it is aligned in diffusion space, but FNIRT convention ignores the affine transformation of the NIfTI format, thus making it's registration useless as the raw data of the t1t2 has nothing to do with the raw diffusion data

(due to them being different resolution). The solution is to apply an affine matrix to t1t2's raw data which transforms it into t1's raw data space, after which the t1's FNIRT warp field can be applied to the t1t2 image. This affine transformation matrix can be easily calculated from the already given matrices. Let $A$ denote T1/T2's affine matrix and $B$ denote T1's affine matrix (after registration), thus the matrix which transforms the T1/T2 into T1 space is $M = A \cdot B^{-1}$.

#### 2.1.1.e Basal Ganglia Segmentation

As the tractography of the brain is performed on the diffusion image, it inherently means that the connectivity maps and the roi are in diffusion space. But the basal ganglia's subcortical segmentation is in T1 space. This means that even if they are registered in the same space, they will not have a pixel perfect union due to the different resolutions.

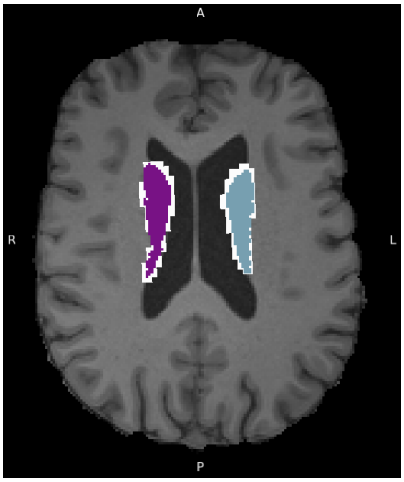


Figure 2.2: Basal Ganglia Subcortical Segmentation

The figure above visualizes the alignment of the Caudate subcortical region, where the white (larger) region is the Basal Ganglia mask from the diffusion space and the colored (smaller) regions are the Basal Ganglia segmentation from T1 space.

In order to keep the data consistent, mapping the segmentation to the Basal Ganglia mask can be done by assigning the same label for each voxel in the basal ganglia as the label of the closest voxel in the subcortical segmentation.

#### 2.1.1.f N-Dim Array

The used NIfTI format stores the raw voxel space and the affine transformation matrix separately, in order to not loose data in the process of interpolating voxels when applying the transformation. But in order to consistently compare voxel data across different spaces (even if they are registered in the same space), the transformation needs to be applied, computing the interpolated voxels in the common space, bringing them into the same raw format of matching X, Y and Z dimensions, and discarding the stored affine matrices.

By default the native anatomical space's origin is near the center of mass of the brain, between the ears. This makes sense for medical professionals, when working with MRI records, but datastructure wise an array is indexed from 0. Meaning after applying the transformation to the voxel space, the yielded array will only contain one quadrant of the record as the rest are clipped in the

negative regions. 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 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 length of the dimensions of the T1 and diffusion records. This can be simply fixed by truncating the excess along each dimension.

#### 2.1.1.g Uniform Shape

After aligning the data into the same space per record, it is still very likely that the individual records 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.

Due to the per-voxel based prediction model architecture this is not a problem, but fixing this for being able to use the data in a spatial model like a fully convolutional neural network (FCNN) can be simply solved by adding padding to the records in order to match their shapes.

| **Data** | **Volumes** | **Range** | **Type** |
|---|---|---|---|
| diffusion | 74 | $[0, 4096]$ | float16 |
| diffusion_fa | 1 | $[0, 2]$ | float16 |
| diffusion_md | 1 | $[0, 0.01]$ | float16 |
| diffusion_rd | 1 | $[0, 0.01]$ | float16 |
| t1 | 1 | $[0, 1000]$ | float16 |
| t1t1 | 1 | $[0, 1]$ | float16 |
| targets | 14 | $\{0, 1\}$ | bool |
| connectivity | 14 | $[0, 1]$ | float16 |
| streamline | 14 | $[0, 5000]$ | float16 |
| mask_basal | 2 | $\{0, 1\}$ | bool |
| mask_brain | 1 | $\{0, 1\}$ | bool |
| basal_seg | 6 | $\{0, 1\}$ | bool |

Table 2.2: Uniform Data

### 2.1.2 Quality Control

Having a low count of records means that if there are even just a few outliers, it can heavily affect the end result. Thus all data were manually inspected to make sure they are as clean as possible.

#### 2.1.2.a Mismatched Data

Looking through the diffusion, diffusion_fa, diffusion_md and diffusion_rd images, 2 records' FA, MD and RD images were seemingly from completely different patients. Thus the FA, MD and RD images were omitted for 2 records.

#### 2.1.2.b Garbled Data

Looking through the subcortical segmentation of the Basal Ganglia revealed that 1 record had a garbled segmentation. Thus, said basal_seg image was omitted for 1 record.

And one record had a garbled T1 FNIRT warp field. Said record was entirely omitted from the normalized set of records.

#### 2.1.2.c Missing Data

Looking through the relative connectivity and streamline images, 3 records were missing these images, said 3 records were completely omitted, as these records are effectively missing the labels.

And the t1t2 images were missing for 10 records, but these were not omitted completely as the t1 images were present for these records, thus experiments only concerning the t1 can have a bit more available data.

### 2.1.3 Radiomics Features

Although the term is not strictly defined, radiomics generally aims to extract quantitative, and ideally reproducible, information from diagnostic images, including complex patterns that are difficult to recognize or quantify by the human eye. [4] Using these features is key, as there are not nearly enough data for neural network (NN) based features extraction such as a convolutional neural network (CNN).

Extracting the voxel based radiomic features has two main parameters to tune, the bin width and the kernel width. Where the binning parameter(s) influence how the intensity values of the image are binned, and the kernel size influences the size of the 'sliding window' similar to a convolution.

The two approaches for binning are absolute discretization and relative discretization. Where in the prior one, a fixed bin width is chosen 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." [5] 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

#### 2.1.3.a Voxel Based

The 92 features in Table **??** will be calculated voxel based. Shape features do not makes sense to calculate voxel based as it would just describe the shape of the used kernel, which is constant and independent from the input image.

#### 2.1.3.b Non-Voxel Based

However, the additional shape features in Table **??** do make sense for the non-voxel based features. As it can be computed for each target region, both hemispheres of the ROI and the entire brain.

### 2.1.4 Coordinates

One additional input that can be included in the experiments is the coordinates. Although this approach only makes sense in normalized space, where the images from different records are aligned. This theoretically would allow the model to learn certain anatomical markers based on the location of the voxel, adding a type of global context to the input of the model.

Furthermore, this approach can be adopted to the native space, by constructing the normalized coordinate map and then 'de-normalizing' them with an inverse FNIRT warp field.

### 2.1.5 Data Augmentation

The only data augmentation that makes sense involves applying small rotation values to the input images in their native space before calculating radiomic features. Applying transformations to the already extracted features is illogical, as interpolating between voxels in feature space is unlikely to yield the same results as computing features after transforming the input images. In summary, any spatial data transformations should be performed upstream. Furthermore, data augmentation only makes sense in native space, as by definition such transformations would make the normalized image pointless.

### 2.1.6 Scaling and Normalization

As the extracted features have very different ranges, it makes sense to follow the standard practice of scaling the data to a fixed range. Inspecting the histogram of some of the radiomic features reveals that most of them follow a bell curve with moderate standard deviation, such as Figure 2.3 (Firstorder Energy).

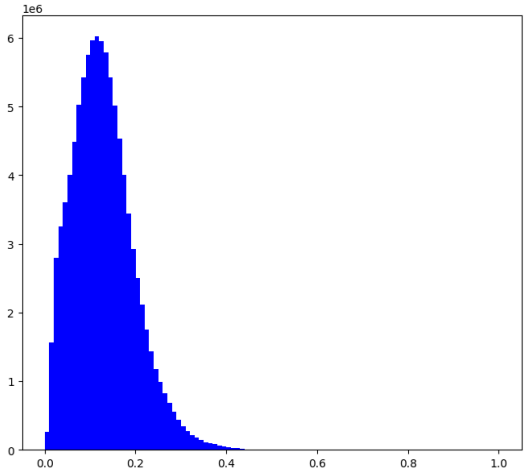


Figure 2.3: Histogram: Firstorder Energy

However, some other features like Figure 2.4 (GLDM Small Dependence High Gray Level Emphasis) and Figure 2.5 (NGTDM Busyness) have a very skewed distribution, the latter one being the most extreme case. This skewing can be eliminated by applying logarithm to the offending features.

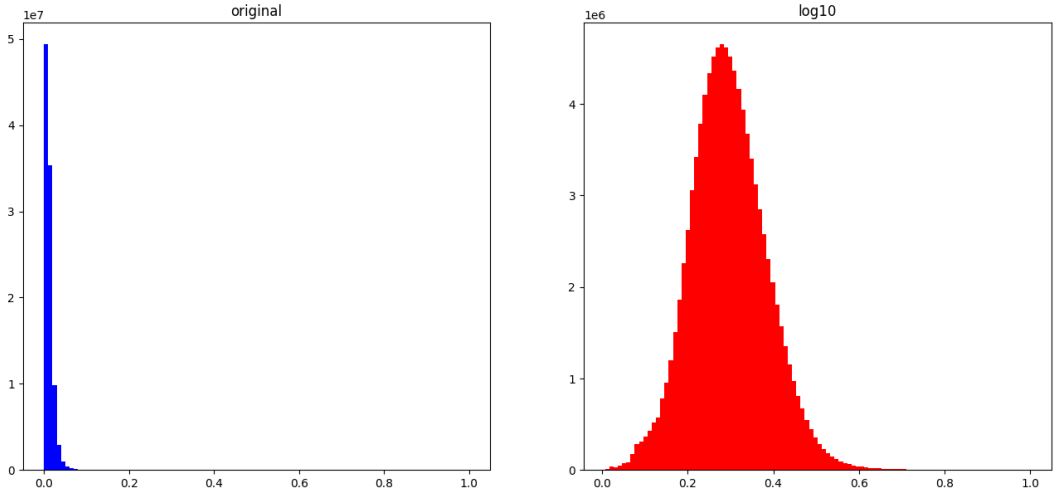


Figure 2.4: Histogram: GLDM Small Dependence High Gray Level Emphasis

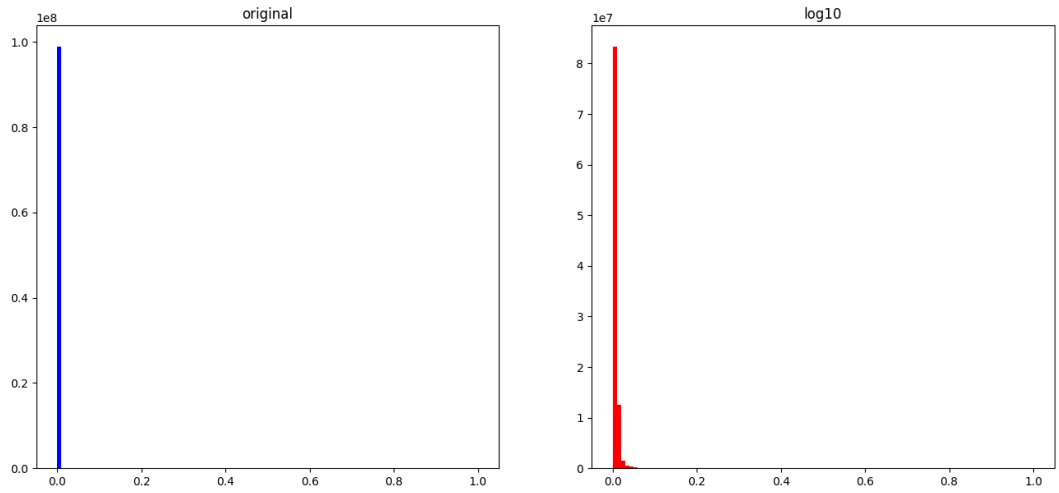


Figure 2.5: Histogram: NGTDM Busyness

Besides the standard benefits of making the optimization process more stable and efficient, and reducing the sensitivity to outliers. It also have some less evident benefits.

Although it is very subtle, but storing these records in float16 inherently looses some information. This loss is not a problem for the features that have a healthy distribution, but in the more extreme cases it can cause compression artifacts visible even to the naked eye, such as the very subtle loss of detail in Figure **??**. And in the most extreme case it can even render the entire feature useless like in Figure **??**. While the normalized features have no problem storing this fine detail in float16.

This makes the system much more robust from a practical perspective; as depending on the hardware, some GPUs are much more efficient at computing in float16. And it also halves the memory and storage requirements, as in float32 a sinlge MRI image of 92 volumes (for the 92 features) takes up around 1GB of space.

### 2.1.7 Data Balancing

Working with highly unbalanced data can be challenging, and balancing it does not necessarily going to help the model's generalization capability. Thus, a method for partially balancing the data will be used, where the bins of the unbalanced data will be up-sampled by a ratio of the difference of the number of datapoints in the bin (compared to the bin with the maximum number of datapoints). Figure 2.6 demonstrates how a ratio 1 means perfectly balanced data, 0 means unbalanced data. And how the ratios in between are approximately preserving the shape of the distribution and partially balance the data.

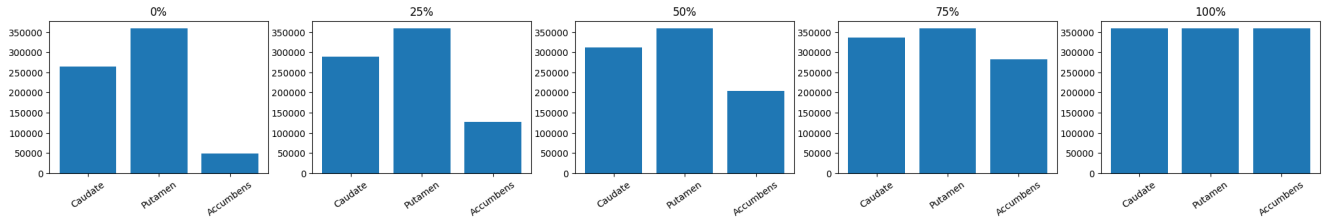


Figure 2.6: Balance: Subcortical

For the diffusion_md and diffusion_fa, which are regression problems and have continuous labels, binning can be used to create artificial groups which can be balanced.

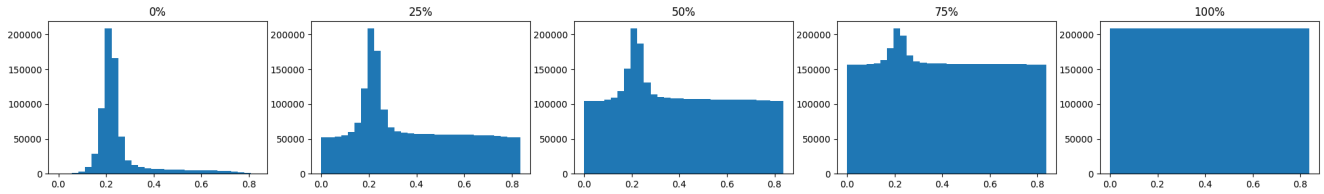


Figure 2.7: Balance: Diffusion MD

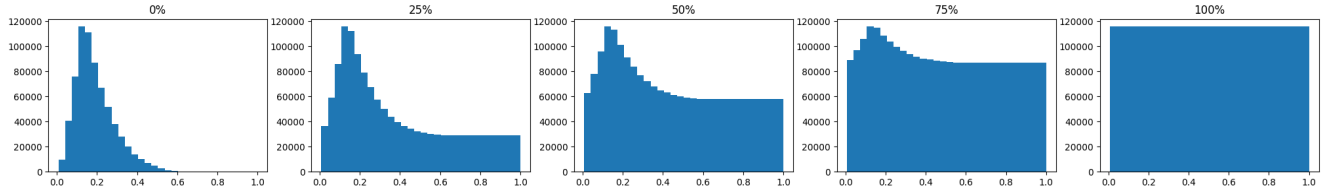


Figure 2.8: Balance: Diffusion FA

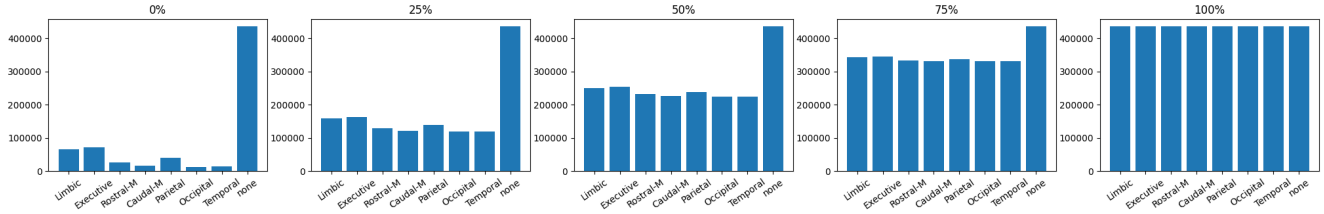


Figure 2.9: Balance: Relative Connectivity (thresholded at 0.6 & binarized)

### 2.1.8 Clinical Data

There are additional clinical data available for the Patient records. Disease severity can be characterized in terms of CAG Age Product (CAP) score. Providing a measure of cumulative exposure to the mutant HTT gene. [6] This widely accepted and used CAP score is available for all patients.

Another, newer metric for characterizing disease severity is the composite Unified Huntington's Disease Rating Scale (cUHDRS) [7]. This is calculated from 4 other basic metrics: Total Functional Capacity, Total Motor Score, Symbol Digit Modalities Test and Stroop Word Reading. These are available for most patients, with a handful exceptions.

And there are a total of 91 available clinical features, with relatively a lot of missing data on some of these features. There are 8 additional patients available in the clinical data. These can be used to aid the data imputation for the missing values, and can be omitted afterwards, as these have no corresponding MRI records.

All clinical features were scaled the range of 0-1 with min-max scaling (per feature). And euclidean distance was used for the following imputation process. The imputation strategy itself consisted of 2 steps, first the few missing cUHDRS values were imputed from the CAP score. And then the remaining features were imputed from the combined CAP and cUHDRS values.

### 2.1.9 Relative Connectivity

asdf

## 2.2 Evaluation

### 2.2.1 Train, Validation and Test Splits

There are 2 important aspects when splitting the data into Train/Validation/Test groups. In order to truly validate the model's generalization capability, the split must happen on a record level and not on a datapoint level. This means that our model can only learn on certain records, and it can be validated on records that it never seen before, not even partially. This has the consequence of that the split will not follow the defined ratio on a datapoint level, as it could happen that by pure chance the train split contains records with larger volumes, resulting in having a bit more datapoints than the validation split.

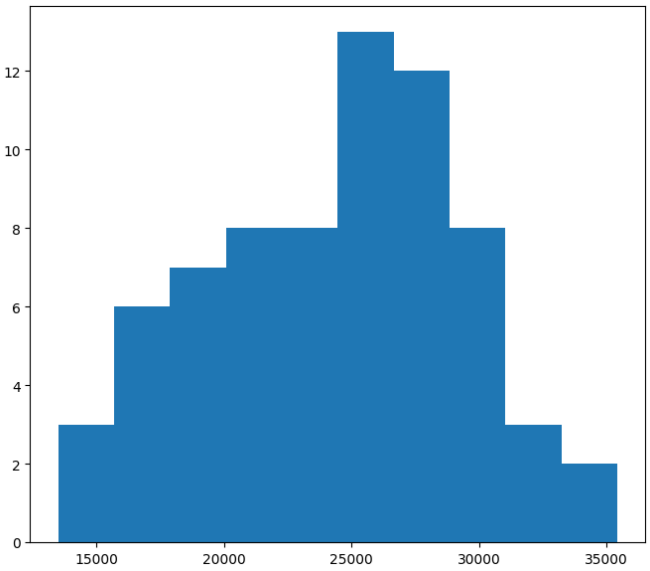


Figure 2.10: Distribution of Records in relation to Datapoints

In practice the lower end of datapoint count per record is around half of the higher end. Figure 2.10 shows the distribution of both Control and Patient records and both Left and Right datapoints. During experimentation, with 0.8 Train split and 0.5 Validation/Test split, the datapoint ratios stayed in the range of $0.8 \pm 0.02$ and $0.5 \pm 0.06$ for the two split ratios.

Furthermore to avoid introducing bias in the case of experiments with mixed Control and Patient records, the ratio of Controls/Patients must be constant across the different splits. This is required as Controls and Patients can have vast differences due to neurodegeneration. An extra caveat is having another ratio that must also be kept constant for the same reason, which is the symptomatic and asymptomatic patients, as they can also have vast differences due to different stages of neurodegeneration.

### 2.2.2 Accuracy and Pearson Correlation

There will be 3 groups of metrics for evaluating each model. First is the 'raw' (will also be referenced as 'train') metric group (train/val/test), which is computed on the datapoints that were extracted with the same hyperparameters as the datapoints during the training process. Meaning that this metric group best reflects how the model performs on the different splits, such as if the model was trained with a balancing of 0.5, both splits will be balanced the same way and the metrics will be computed on a datapoint level (the same way as how it is naturally computed in the loss function).

The second and third groups are for comparing model performances in between models and are for practical evaluation. The difference is that the metrics in this case are computed for each record, and then averaged out inbetween records. This means that it is computed on a record level instead of a datapoint level, resulting in the elimination of potential bias coming from the deviation from the number of datapoints per records. It also means that these metric groups will inherently ignore data balancing, as it operates on a record level.

And the 2nd metric group is computed in native space, while the 3rd is computed in normalized space. This means that if the model operates in native space, the normalized metrics will be computed by predicting the datapoints for each record, then the spaital record is reconstucted from the datapoints and warped to normalized space, and then the datapoints are extracted from the normalized spaital prediction, and compared against the normalized labels (This process would be computationally quite expensive, so the implementation does not follows this exact logic, but numerically it is doing the same; more information on this in Appendix **??**). This way the models can have comparable metrics even if they operate in different spaces.

# Experiments

The following hyperparameters were constant during all of the experiments:

| **Hyperparameter** | **Value** |
|---|---|
| Train Split | 0.8 |
| Validation/Test Split | 0.5 |
| Model Type | feedforward neural network (FNN) |
| Optimizer | Adam |

Table 3.1: Hyperparameters: Common

## 3.1 Subcortical Segmentation

This simple problem did not need a lot of tuning, as it was working very well almost from the start. The following set of hyperparameters were constant during these experiments:

| **Hyperparameter** | **Value** |
|---|---|
| Control/Huntington Datapoints | Control Only |
| Left/Right Hemisphere Datapoints | Both |
| Space | Native |
| Image | T1 |
| Scaling/Normalization | Normalized Voxel Based Features |
| Hidden Layers | $1024 \rightarrow 512 \rightarrow 256 \rightarrow 128$ |
| Loss | Categorical Crossentropy |
| Activation | Sigmoid (softmax for the output layer) |
| Learning Rate | 0.001 |
| Batch Size | 10000 |
| Early Stopping Patience | 7 |

Table 3.2: Hyperparameters: Subcortical

The reasoning behind the initial choices of these parameters are straight forward. The T1 image and native space were chosen, because those are the simplest to acquire in practice. Thus if the model is doing great on those, there is no need for more complicated inputs. Including both hemispheres would hopefully result in a model which can generalize better. Only using control datapoints should translate into less variance between the general characteristics of the datapoints, as it does not contain patients with neurodegeneration. The number and sizes of the hidden layers were choosen based on the potential size of the input layer, which should range from 92 (single set of voxel based features) up to $\sim 1000 - 2000$ including many different kernel sizes and non-voxel based features as well. Categorical crossentropy loss function and the output layer's softmax activation function are standard practice for a classification problem. The Sigmoid

activation function should work fine without having to deal with exploding gradients and dieing relu problems. Learning rate is the default learning rate of the Adam optimizer in TensorFlow. And a batch size of $10,000$ seems appropriate for a train split size of $1,000,000$ datapoints. And the early stopping patience of 7 epochs should also be good enough to prevent overfitting and stop the training in time, but it will be evaluated based on the learning curves and the accuracy of the model.

The used metric for evaluating model performance on the Train/Validation/Test splits is Accuracy. The 'k' and 'b' notations stand for kernel and bin, where k5 means a kernel width of 5mm, b25 means an absolute bin size of 25, and b10r means relative binning with 10 bins. And in the case of multiple kernel sizes denoted by a dash, it naturally only means odd kernel sizes. Tuning the rest of the hyperparameters were done in the following experiments:

| | **Experiment** | **Train** | **Val** | **Test** | **Input Layer** |
|---|---|---|---|---|---|
| 1. | **Voxel Features k5_b25** | 68.9 | 69.1 | 72.4 | 92 |
| 2. | *Voxel Features k5_b25*<br>**Non-Voxel Features of Target Regions b25** | 73.2 | 68.9 | 72.5 | 1576 |
| 3. | *Voxel Features k5_b25*<br>**Non-Voxel Features of ROI b25** | 75 | 74.3 | 78.5 | 304 |
| 4. | *Voxel Features k5_b25*<br>*Non-Voxel Features of ROI b25*<br>**Non-Voxel Features of Brain b25** | 74.5 | 70.4 | 70.3 | 410 |
| 5. | *Voxel Features k5_b25*<br>**Non-Voxel Features of ROI b10 b25 b50 b75** | 71.2 | 70.6 | 74.3 | 856 |
| 6. | **Voxel Features k5_b25 - k21_b25**<br>*Non-Voxel Features of ROI b25* | 94.5 | 94.1 | 95.1 | 1040 |
| 7. | **Voxel Features k5_b25 - k21_b25** | 95 | 94.6 | 93.7 | 828 |
| 8. | *Voxel Features k5_b25 - k21_b25*<br>*Non-Voxel Features of ROI b25*<br>**Balance Ratio 0.5** | 94.9 | 94.4 | 94.9 | 1040 |
| 9. | *Voxel Features k5_b25 - k21_b25*<br>*Non-Voxel Features of ROI b25*<br>**Balance Ratio 1** | 95.9 | 95.5 | 95.9 | 1040 |

Table 3.3: Hyperparameter Tuning: Subcortical

The best performing model was in experiment number 9, where it achieved a 96% accuracy with practically no overfitting. The biggest improvement during the experiments was to include many different kernel sizes for the voxel based features. The additional non-voxel based features of the ROI yielded a small improvement. And balancing the data yielded a marginal improvement, by reducing overfitting.

Examples of the true/predicted records can be found in Figures **??** **??** **??**. And the loss training curves can be found in Figure **??**.

## 3.2 Methodology

The experimentation from this point on, will be divided into 4 main groups:

- Native - T1

- Native - T1/T2
- Normalized - T1
- Normalized - T1/T2

The same set of core experiments will be run for all 4 groups, and some additional experiments will be run per group, depending on how they perform. The experiments will concern the following aspects:

- Single/Many Different Kernel Sizes for Voxel Based Features
- Additional Non-Voxel Based Features
  - Single/Many Different Bin Sizes
- Control/Patient/Both Records
- Left/Right/Both Hemisphere Datapoints
- Additional Clinical Features for Patient Records
- Additional Coordinate Map Features
- Scaled Voxel Based Features (not normalized)
- Different Bin Sizes for Voxel Based Features
- Different Balance Ratios

### 3.2.1 Missing Datapoints

In order to be completely fair when comparing model performances, only records should be used which are available for all 4 groups of experiments. In practice the following records were missing:

| **Record** | **Missing Amount** |
|---|---|
| Normalized | 1 |
| T1/T2 | 10 |
| Diffusion FA & MD | 2 |

Table 3.4: Missing Records

This meant that for the Diffusion FA & MD experiments there were a total of 13 records omitted, yielding 57 records in total, out of which 29 are Control and 28 are Patient records. And for the Relative Connectivity experiment, 11 records were omitted, yielding 59 records in total, out of which 30 are Control and 29 are Patient records.

As additional experiments for the groups with more available data (such as T1, where 10 more records could be included), these records can be appended to the train split on the best performing model, feasibly increasing model performance.

### 3.2.2 Arhitecture Tuning

For the best performing model, the architecture will be further tuned, concering the following aspects:

- Number of Layers and Layer Sizes
- Activation Function
- Batch Size
- Learning Rate
- Dropout Normalization
- Early Stopping Patience

## 3.3 Mean Diffusivity Regression

# 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] 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.

[5] 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.

[6] *Special Characteristics of HD Data*. 2022. URL: https://enroll-hd.org/for-researchers/analyzing-data/special-characteristics-of-hd-data-2.

[7] Dylan Trundell et al. "Defining Clinically Meaningful Change on the Composite Unified Huntington's Disease Rating Scale". In: *Neurology* 92.15_supplement (2019), P1.8–043. URL: https://doi.org/10.1212/WNL.92.15_supplement.P1.8-043.