Title

Pipeline for using Dictionary Learning for analysis of morphometry differences across populations of MRA data

Author

Mendoza, Steve

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Pipeline for using Dictionary Learning for analysis of morphometry differences across populations of MRA data

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Electrical and Computer Engineering

by

Steve Anthony Mendoza

2021

ABSTRACT OF THE THESIS

Pipeline for using Dictionary Learning for analysis of morphometry differences across populations of MRA data

by

Steve Anthony Mendoza
Master of Science in Electrical and Computer Engineering
University of California, Los Angeles, 2021
Professor Fabien Scalzo, Chair

Identifying population differences can serve as an insightful tool for diagnostic radiology. To do so, a reliable preprocessing framework and data representation are vital. We consider building a machine learning model to visualize gender differences in the circle of Willis, (CoW) an integral part of the brain's vasculature. We start with a dataset of 570 individuals and process them for analysis using 389 for the final analysis. We find statistical differences between male and female patients for one viewpoint and can visualize where they are. In particular, we find that the interior cartoid atery (ICA) is larger in males than in females, and that the CoW is more likely to be complete in females. We also see differences between the right and left-hand sides of the brain confirmed using SVM. This process can be applied to automatically detecting population variations in the vasculature and can serve as a guide to explaining machine learning decisions.

The thesis of Steve Anthony Mendoza is approved.

Aichi Chien

Achuta Kadambi

Anthony Chen

Fabien Scalzo, Committee Chair

University of California, Los Angeles

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

(Acknowledgments omitted for brevity.)

CHAPTER 1

Introduction

1.1 Background

1.1.1 Overview

Stroke is a common cause of death or disability in the developed world, ranking among the top 5 causes of death. Due to its prevalence, extensive research and data exist for stroke prevention, diagnosis, and risk factors. The latest research centers around using data-driven algorithms using machine learning analysis. Among the standard data used is angiography, which means imaging blood vessels. An angiography is a tool that can diagnose strokes and other vascular diseases. We focus on a type of angiography known as Magnetic Resonance Angiography, or MRA. The MRA technique provides high contrast blood vessel imaging in 3 dimensions without ionizing radiation, differentiating it from other angiography techniques. Several MRA imaging datasets for population-level studies exist. Among the questions researchers ask using these datasets include changes in the vasculature during aging or the average size of vessels. The motivation for this work is to develop a pipeline for determining regions that are different between two different populations. These populations could be gender-based, age-based, or disease-based. We will go through the steps of processing the data to optimizing and interpreting the differences between two different groups. We will focus on the anatomical differences between male and female patients since the literature suggests reproducible anatomical differences in vasculature. We conjecture that studying the population differences between them can offer insights into different stroke outcomes. Our main goal is to create a quantitative and visual representation of these vascular differences.

The results can influence both designing and interpreting machine learning networks and disease, highlighting regions for further diagnostics.

1.1.2 Clinical Significance and Basic Anatomy

There is a significant variance in the incidence of stroke patient outcomes; males having a greater chance of strokes than females, while females have a greater chance of more severe stroke when they get them. [23] Given the discrepancy in these outcomes, it is reasonable to expect some clinical or anatomical variance between them, although the extent to these differences is not entirely clear. The clinical differences between males and females serve as motivation to find anatomical differences. The objective is that this same method, tested on healthy patients, could serve as a platform for determining stroke patient severity and differences in future works.

In this work, we will focus on the part of the brain known as the Circle of Willis. This area is where the blood comes in for distribution to other parts of the brain. [17] It sits at roughly eye level at the base of the brain. It is a redundant network of blood vessels, shown in 1.1. These vessels are integral in determining blood flow throughout the brain, and there are several viewpoints to consider when interpreting these images. The projection shown is the axial viewpoint, but when considering anatomical analysis, it is vital to account for the 3-dimensional nature of the imaging so we can observe all angles of the brain. In addition to the axial viewpoint, we also consider the coronal axis to account for the 3-dimensional structure. The Circle of Willis rests on the lower part of the head closely connected to the neck arteries, situated far from the skull. Next, the circle branches out with the largest near the circle center, narrowing farther away from the center as shown in the axial viewpoint. The most prominent are the Internal Carotid Arteries (ICA), from then on are connected to the Anterior Circulation Artery (ACA), Middle Circulation Artery (MCA), and Posterior Communication Artery (PCA). While other arteries, including the Posterior Communicating Artery (PCom) and the Anterior Communicating Artery (ACom), exist, these smaller arteries are harder to see with this imaging modality and absent in some

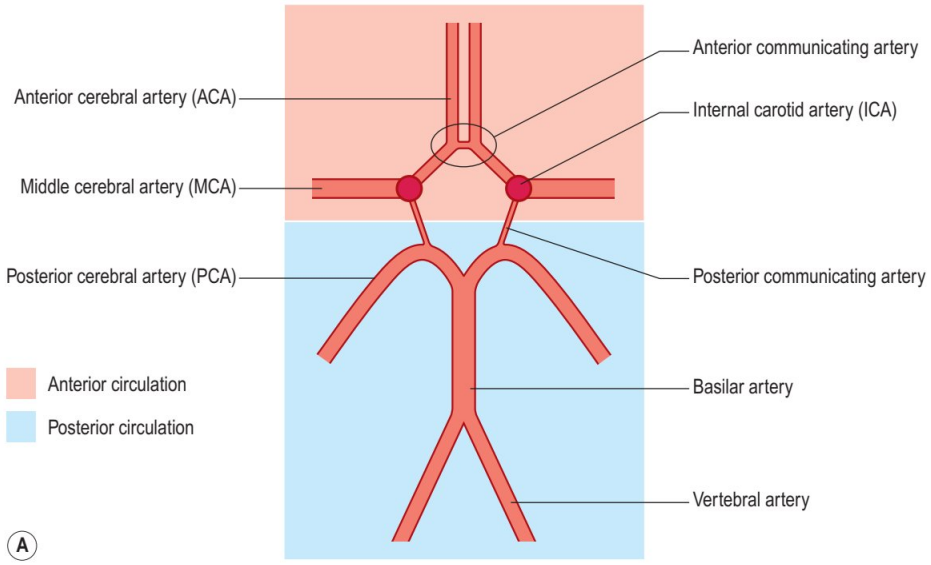

Fig. 10.5 **The anterior and posterior circulations and their origins from the circle of Willis.** **(A)** Schematic, two-dimensional, representation of the circle of Willis, showing the anterior (internal carotid) and posterior (vertebrobasilar) circulations; **(B)** Illustration of the three-dimensional arrangement of the circle of Willis.

Figure 1.1: Circle of Willis. This anatomy will be the reference by which we will describe arteries. The viewpoint given here is the canonical viewpoint as shown in textbooks and the results described later in the work will mainly use this viewpoint.

patients. So we will focus our discussion on the ICA, ACA, MCA, and PCA arteries.

1.1.3 Imaging

In our work, we use a specialization of MRI known as MRA. To explain how MRA works, we will first consider MRI in general. MRI stands for magnetic resonance imaging and uses radio frequency (RF) pulses to excite hydrogen atoms to all have the same spin. MRI involves a high-powered magnet that directly correlates with signal strength necessitating a substantial magnetic force, usually more than 1 tesla (T). Various groups and clinics may use different scan parameters affecting how the images appear. The most common values are 1.5T and 3T; higher fields correlate with the signal-to-noise ratio.

Once we excite the protons in the sample, the proton spin will parallel the magnetic field direction. Although the proton spins are parallel, their orientation, known as spin up or spin down, is random. The net amount of spin up or spin down creates a net magnetic

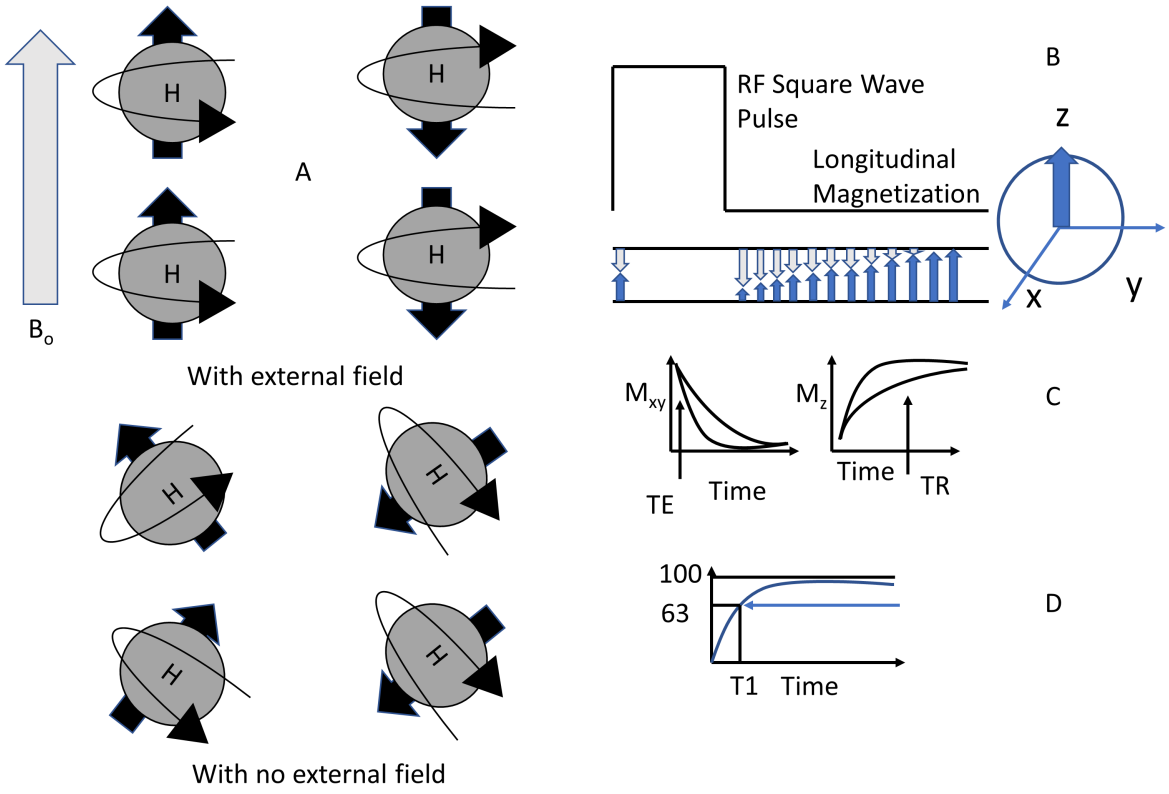

Figure 1.2: Overview of MRI (A). The key concept is that atoms all have a random spin under no magnetic field then a net magnetic field under a strong magnetic field B_o (B.) An RF pulse is applied, flipping the net magnetic field. After the pulse, the protons gradually come to the original magnetic field direction (C.) This relaxation time is known as the T1 time (D). For sequence generation, we manipulate two parameters, TE and TR. TE is echo time, and TR is the repetition time.

moment. This net magnetic moment is equal to the magnitude of the resultant magnetic field. This strength will depend on the local properties of the protons themselves. Once the magnet is on, a series of RF pulses follows, with one procedure involving flipping the proton orientation 90 degrees, making it perpendicular to the magnetic field. Following the pulse, the protons will relax to their original orientation as dictated by the magnetic field, called the T1 relaxation. This T1 time will depend on the environment of the proton itself. The T1 time represents the time it takes when turning off the RF pulse for the protons to relax down towards the thermal equilibrium, depending on the proton environment. Figure 1.2 shows the T1 process. The images were from [24], which gives an overview of the MRI concept. The relaxation process then creates an electric field which the apparatus then measures, digitized into the MRI signal.

In our case, we focus on MRA, which is Magnetic Resonance Angiography, a subclass of MRI for measuring blood flow. It is specialized to distinguish static tissues from flowing blood. It came about by observing some artifacts when first using MRI wherein normal MRI, places near vessels were dark. Using these artifacts, researchers modified the procedure such that the T1 times for stationary and moving tissue would allow for imaging. The specific technique is TOF MRA, TOF stands for Time of Flight, which relies on the saturation of blood vessels so that after a short time after the demagnetization, we get a saturated signal. Looking back at 1.2, this would correspond to a lower TE time and a lower TR for the optimal contrast. Next, the signal gets digitized into grayscale images, where bright image values correspond to high blood levels allowing for arterial images at high contrast. These values can affect the image contrast and image quality preprocessing algorithms. Machine learning algorithms could be affected by the variation of these different parameters.

1.2 Introduction to Morphometry and Motivation

Determining medical imaging patterns across populations could be useful in diagnosing or making sense of medical imaging data. The process begins with obtaining a significant data sample of at least a few hundred cases and then performing statistical analysis to find high

variance areas. Finding and validating these high variance regions is known as morphometry. The simplest approach to solve this problem would be to consider two populations and see which voxels contribute the highest variance between different subjects. A general platform for doing so is voxel-based morphometry (VBM) [7]. In this methodology, one acquires images from various subjects and registers them to each other using anatomical atlas. VBM has been most often applied to the analysis of MRI images to identify changes in normal patients and those with various psychiatric disorders [7]. However, it is almost impossible to distinguish population differences from differences in registration or other image processing details such as dynamic range without proper normalization. A few papers [10], have mentioned some key facts to keep in mind for the interpretation of the data. Key among them are the registration and the normalization of the brain data. Several groups use different algorithms for normalization and registration that influence the results. Due to its simplicity, it is one of the most popular methods for determining differences among populations. There is also a well-established platform for determining statistical significance for any brain differences found. The main method is the Gaussian linear method to determine statistical significance. In essence, the individual voxels are individual variables, and the Gaussian linear field tries to find corresponding voxels that are different in two respective populations. However, if the data is not well registered, the statistical pattern would reflect misregistration rather than anatomical differences. [10]. An illustration of the technique in practice is shown in 1.3, from [31].

1.2.1 Pattern Based Morphometry and Data Formatting

In addition to VBM, there is also another more fairly recent approach known as pattern-based morphometry or PBM. We will discuss the basic principle behind PBM, which applies dictionary learning of a sparse representation of a basis. The original idea of dictionary learning was to solve the sparse representation of signals problem [2]. Conceptually, the idea is to take a signal, then find a small number of atoms to represent the basis. The key to doing so is finding a basis representation in which the result is sparse. Among the first applications

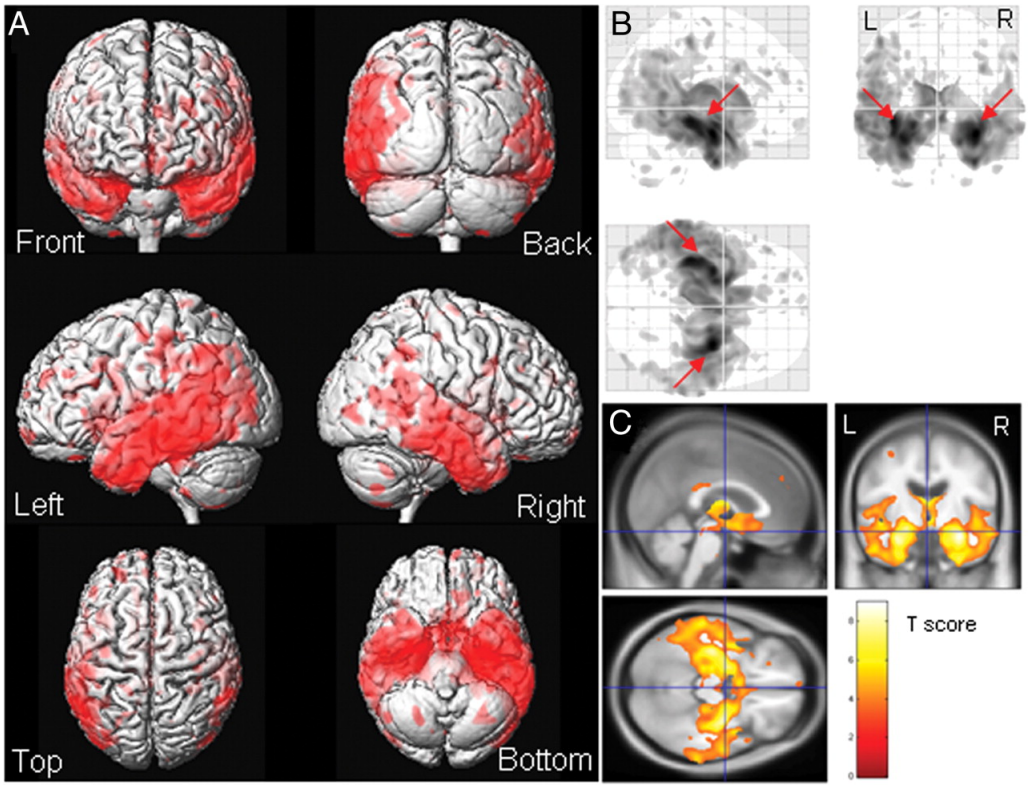

Figure 1.3: Voxel Morphometry approach on MRI imaging of the brain. These images show the brain regions that show the greatest variance on brain imaging. There are various projections of the brain to show the 3D volumes.

using dictionary learning is image compression [2]. The dictionary method implementation uses a wavelet transform of the image data; this wavelet transforms are the dictionary that would allow a compressed representation of the image, allowing for faithful preservation while at the same time allowing for compression. The general mathematical intuition is that it is a generalization of the k-means clustering algorithm, which for a certain k, the idea is to split the data so that k different centers are found such that any one point in the dataset is no further than a value *epsilon* from the closest k. In that sense, we can think of the dictionary learning problem as having a set of basis vectors k such that these k basis vectors can reconstruct any training data example with an error less than a value epsilon.

Pattern-Based Morphometry (PBM) tries to answer the same questions as VBM, with a more global data representation. [14] The main difference in the PBM approach is in the sparsity constraint. This sparsity constraint encourages the algorithm to generate global samples present in several samples rather than voxel-based representations that are highly local and possibly noisy. Specifically, we will search for a set of 6 subtraction images to represent the differences between two populations. Our definition of subtraction images is: we take one case minus all the population members of another population. The user then selects a parameter k, the number of images to compare each image in the other population. The key metric is to take the euclidean distance between images, and the algorithm considers the closest k images for each image. These difference images then represent differences among different populations via image subtraction. The goal of PBM is to compress the representation of all these images into a sparse dictionary showing the most salient differences between the populations. The number of different images to consider is a hyperparameter, known as the number of atoms. The user specifies the number of atoms for the algorithm to find with the algorithm converging to the resultant atoms that best represent the image basis by minimizing the error between the set of difference images and the closest dictionary atom to each case. It is similar to how clustering algorithm convergence. We formalize the above steps by constructing two different groups, S defined as $S = \{S_1, ..., S_n\}$ and group 2 with $\{Z_1, ...Z_r\} \subset Z$, where r is the number of neighbors in our case six. The definition of

the difference vectors is:

$$D_{ij} = S_i - Z_j \quad \forall i \in \{1, ..., n\} \quad and \quad \forall j \in \{1, ..., r\} \tag{1.1}$$

From the equation above, the order matters; the definition of group 1 or group 2 will influence the patterns. In the following subsection, we then use Equation 1.1 to describe our specific dictionary learning approach.

1.2.2 Introduction to Dictionary Learning

For a brief overview of the theory and motivations of dictionary learning, we review principal mathematical equations and visualize the results. In the first paper describing PBM, [14] the authors described differences in brains between healthy patients and Alzheimer's patients. We will discuss the mathematical details discussed in [14]. The main set of equations in the paper are as follows.

$$\begin{aligned} &\underset{B,C}{\text{minimize}} \quad \|X - BC\|_F^2 \\ &\text{subject to} \quad \forall i, \|c_i\|0 \leq \ T \end{aligned} \tag{1.2}$$

The above is a minimization problem to take a basis matrix B, which in our case will be the dictionary atoms, and C, which is a sparse matrix used to rank the dictionary atoms. X is the original dataset.

$$\begin{aligned} &\forall i \in \{1, 2, ...n\} \quad \underset{c_i}{\text{minimize}} \quad \|x_i - Bc_i\|_2^2 \\ &\text{subject to} \quad \forall i, \|c_i\|0 \leq \ T \end{aligned} \tag{1.3}$$

This second equation is the algorithm used to solve the first problem. The algorithm initializes basis vectors, then searches for the sparest representation.

$$E_k = X - \sum_{j \neq k}^{k} b_j c_j \tag{1.4}$$

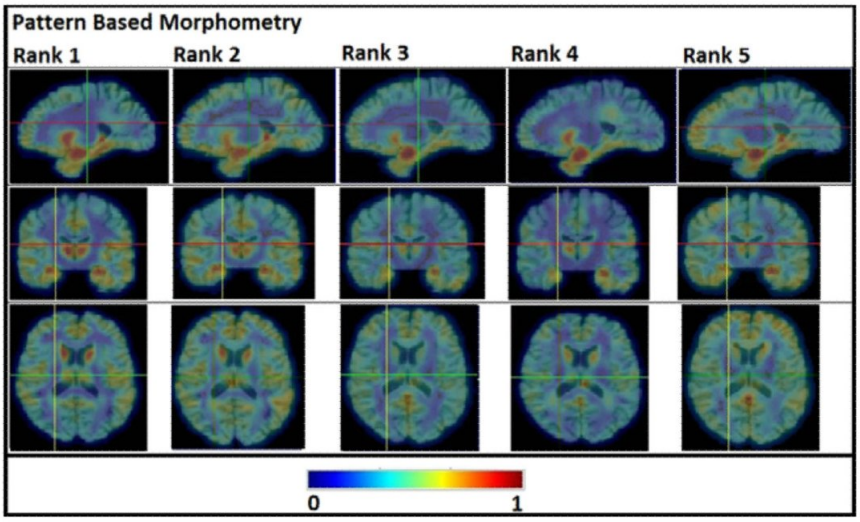

Fig. 2.
The 5 top ranked basis/patterns obtained from PBM in WM from ADNI data

Figure 1.4: From [14], showing the salient features in the three different viewpoints taken from pattern-based morphometry. They compare Alzheimer's Disease patients with normal ones; the patterns shown are the most significant regions compared to normal cases in the WM. The red areas are the most salient; the darker regions have the least variance.

The equation below updates the basis vectors B_k one by one using the error metric as shown in the previous equation.

$$\underset{B_k}{\text{minimize}} \quad \|E_k - b_k c_k\| \tag{1.5}$$

Figure 1.4 shows the pattern-based morphometry images from [2], we see that the most significant areas, those closer to 1, are found in places where there are meaningful differences between populations, and these areas link to actual changes based on anatomical studies.

CHAPTER 2

Methods

2.1 Registration and Data Processing

We start with a background on the dataset used and the data selection process. Then a detailed analysis of the mathematical tools follows. These mathematical tools include pattern-based morphometry and registration and segmentation preprocessing for machine learning. The second part of the methods section focuses on the SVM classification and data selection for finding differences between populations.

For most of the analysis, we used Matlab 2018 [21], as well as python calls to software known as ITK written in c++ for the preprocessing and segmentation. The bulk of the preprocessing follows a GitHub repository called pypbm; python pattern-based morphometry. [19].

2.1.1 Dataset Description

The dataset is known as the IXI dataset, with data from 3 different institutions in the United Kingdom. [12] There are 570 MRA datasets, with attributes such as age, gender, educational level, and ethnicity. The three different institutions allow for better generalization controlling for bias due to diverse scanner parameters or contrast. Owing to the data variety, robust data preprocessing is necessary. Of particular note was one taken for the institute of psychiatry (IHOP). For this study, we have considered cases from this site though another paper [22] did not include this dataset since the MRI acquisition patterns were not given but included

the two other institutions. We will incorporate this in our dataset since we want the most data points. We then go through a series of data filtering and preprocessing steps described in the Methods section to go from 570 cases to 389 cases. Four hundred subjects met our filtering criteria, but out of these 400, 11 subjects had no gender listed.

2.1.2 Registration Algorithm

For each registration algorithm, there are three modules: a transform method, a similarity measure, and an optimization model. The transform is the mathematical representation that can make the registration work more efficiently. A similarity model is a metric for measuring the distance between target and template. Finally, an optimization model, such as gradient descent, provides a minimization method to reduce the error and converge to the target. Two broad approaches to brain registration are rigid and non-rigid registration. Rigid registration includes rotations and scaling; these are simple linear transformations. The brain shape is preserved and not deformed to fit a target. For rigid registration, one can select a mathematical transform to parametrize the image in an affine space, use cross-correlation as a similarity measure, and use gradient descent to minimize the cross-correlation error. However, the second method, non-rigid registration, is necessary for detailed analysis of the brain since the vessels themselves have some deformation and need to have some local deformation to better fit a template. For the non-rigid registration, it is subtler to create a mathematical model and is an active research area. We will detail the specific mathematical model for our data analysis.

. Mathematical methods behind registration. The mathematical theory behind our registration method is the symmetric image normalization method (SyN), [8] the general concept of which is that if you consider two different brains, for instance, you want to use an image space that minimizes the differences between them. The main way is to model the two different brains as two distinct optical flows, then casting that optical flow into a differentiable map. One then sets that differentiable map as fixing one map while taking another brain as

a moving optical flow and mapping that optical flow to the static template. The notion of a differentiable map allows for a gradient-based method to minimize the geometric distance between these two brains. The symmetric normalization stems from the insight that the differentiable map can go from the template to the target or vice versa. The convergence criteria for our system is the mutual information between the template and the target. For our registration technique, the main theory is the diffeomorphic operator for which we describe below: A diffeomorphism in general, takes a diffeomorphism ϕ, on the image domain Ω such that the diffeomorphism applied to the image domain and some multiplication δ yields an affine transformation on the identity transform. The symbols in the equation are as follows: u, a displacement field, v, a velocity field, and L, a linear operator. This mapping ϕ comes from integrating the ordinary differential equation described below:

$$\frac{d\phi(x,t)}{dt} = v(\phi(x,t),t), \phi(x,0) = x \tag{2.1}$$

The integrated field generated through the diffeomorphism is $u(x) = \phi(x,1) - x$.

To obtain the integrated field as the differential equation implies, the SyN protocol allows for a Large Deformation Diffeomorphic Metric Matching (LDDMM) expression. A LDDMM representation allows for a representation of the data in the diffeomorphic space, so that the registration can take place within that feature space. This expression is necessary since individual anatomical cases can vary widely in their intensity profiles so the standard mean squared error metric is not optimal for certain cases. The formulation is as follows:

$$v^* = \underset{v}{\operatorname{argmin}}\{\int_0^1 \|Lv\|^2 dt + \lambda \int_\Omega \Pi\ (j, \phi(x,1), J) d\Omega\ \} \tag{2.2}$$

The Π above is the similarity metric with the image domain as the integral domain. The λ parameter is a regularization parameter that controls how exact the matching is. The above formulation can also be broken into two different integral limits, reflecting the symmetry between the target and the template.

$$\{v_1^*, v_2^*\} = \underset{v}{\operatorname{argmin}}\{\int_0^{0.5} \|Lv_1(x,t)\|^2 dt + \int_0^{0.5} \|Lv_2(x,t)\|^2 dt + \lambda \int_\Omega \Pi\ (j, \phi(x,1), J) d\Omega\ \} \tag{2.3}$$

Minimizing the respective integrated fields for v_1 and v_2, yields a representation so that regardless of the initial conditions, the algorithm converges. The second half of the above equation is the same and has the same regularization purpose as the original equation.

The specific algorithm used in our work is the SyNRA, which stands for Rigid + Affine + deformable transformation, with mutual information as an optimization metric. The mutual information metric applied to medical imaging considers the image histogram in image subregions. Well registered images should have corresponding image histograms for image subregions. If there is misregistration, then the image histograms would not have corresponding histograms.

To complete image registration, preprocessing steps are sometimes necessary to account for individual patients with low contrast or low signal-to-noise ratio. The image processing pipeline described does incorporate histogram normalization, though, for some cases, we used the non-local means filter implemented in MatLab for additional denoising. While the large vessels, namely, the Internal carotid artery (ICA) or the Middle cerebral artery (MCA), are unaffected, the smaller Posterior Cerebral artery is harder to observe. Thus for some cases, it was best to do some contrast enhancement before the registration to enhance the lower contrast small vessels. Since the image quality varies from one patient to another, some cases required manual fine-tuning where the registration was of noticeably low quality using the registration quality metrics described.

2.1.3 Radon Transform

The radon transform is similar to the Fourier transform and is used to perform rigid registration analogous to Fourier domain phase shifts are. The theory is that when transforming images into the Fourier domain, linear translations in the Fourier domain correspond to

rotations in the original image domain. Knowing this, one can create an algorithm for registration using this relationship. However, we will use a similar concept by computing the radon transform of the template and the individual cases. After computing both transforms, we then calculate the maximum cross-correlation. The higher the cross-correlation, the better the correspondence. A paper [11] used the radon concept to register images in the 2D axial projection for MRI images similar to our data. Their metric relied on the tuning of the registration angle to maximize the cross-correlation in the radon transform basis; we can use similar logic for a quantitative measure of the registration quality since most registration algorithms, including the non-rigid registration algorithm, described beforehand, incorporate cross-correlation as a convergence metric.

The mathematics behind the radon transform and the algorithm described in the paper is described below:

$$Rg(x,y) = \int_L g(x,y)dl \tag{2.4}$$

where L is the integration over the polar coordinate space:

$$p = x\cos\theta + y\sin\theta \tag{2.5}$$

by replacing x and y by the line integral relationship, we get:

$$x = p\cos\theta - q\sin\theta \tag{2.6}$$

and

$$y = p\sin\theta + q\cos\theta \tag{2.7}$$

for x and y respectively

After the transformation, we then take the integral using the x and y transformations

detailed above to arrive at the integral shown below:

$$g(p,\theta) = \int_{-\inf}^{+\inf} g(p\cos\theta - q\sin\theta, p\sin\theta + q\cos\theta)dq \tag{2.8}$$

The result of the integral is then shown below

$$g(x\cos\emptyset - y\sin\emptyset, x\sin\emptyset + y\cos\emptyset) \tag{2.9}$$

Which represents the radon transform representation.

$$Rg(p,\theta+\emptyset) \tag{2.10}$$

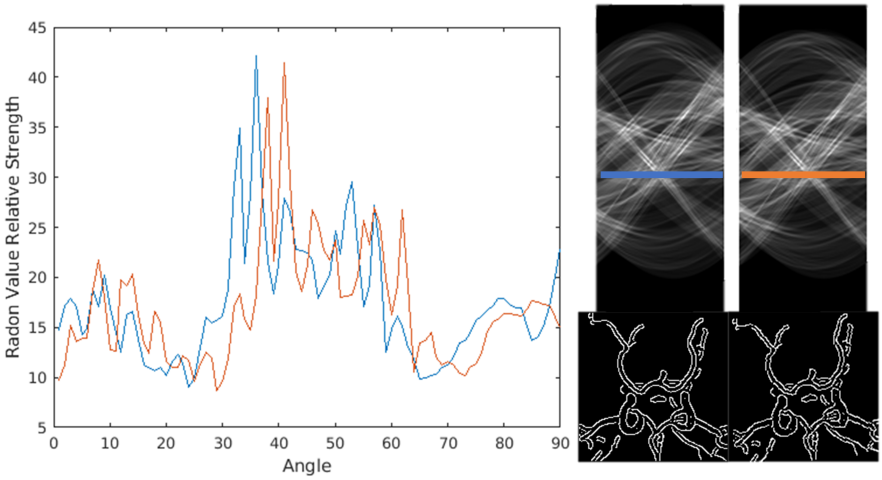

Figure 2.1: Radon Transform from [11]. The blue line is the original image, the orange line is the same image but rotated 5 degrees. We can see a clear offset in the radon transform, which we could then measure using the maximum normalized cross-correlation. The higher the value, the more closely aligned the images.

The atlas is vertically flipped relative to the standard viewpoint of the Circle of Willis. Thus we show the orientation we use to complete the machine learning and analysis, but for representation of the results, we will show them in the standard Circle of Willis view used in anatomy books since it is the conventional view and makes it easier to discuss.

2.1.4 Segmentation Strategies

After the registration, we then seek to extract the main vessels from the registered brains. The normalization as part of the registration makes comparisons between different samples simpler than if the brains were analyzed separately. The principle variable for the utility of this segmentation is the contrast between the brain and the vessels. However, only using image contrast is not always ideal as the imaging field can be non-uniform, and imaging conditions could vary from patient to patient. To solve the problem, a 2nd order image derivative, [16] is needed since the second derivative indicates concavity and has a robust response in vessels.

The key equation is shown below:

$$H_{i,j}(x,s) = s^2 I(x) * \frac{\partial^2}{\partial x_i \partial x_j} G(x,s) \quad \text{for} \quad i,j = 1,...D \quad where$$
$$G(x,s) = (2\pi s^2)^{\frac{-D}{2}} \exp \frac{-x^T x}{2s^2} \tag{2.11}$$

Equation 2.11 is the general format of the hessian matrix, denoting 2nd order partial derivative found in older vessel segmentation papers. The main problem is generalizing this measure for multiple scales. The solution is to select for certain eigenvalues of the matrix. While the above equation is the template for most studies into segmentation, the authors have their innovation that they describe below, based on an adjustment on the volume ratio of the eigenvalues:

$$VR = |\lambda_1 \lambda_2 \lambda_3| \left[\frac{3}{|\lambda_1| + |\lambda_2| + |\lambda_3|} \right]^3 \tag{2.12}$$

and

$$|(\lambda_2 - \lambda_1)\lambda_2 \lambda_3| \left[\frac{3}{|2\lambda_2 - \lambda_1| + |\lambda_3|} \right]^3 \tag{2.13}$$

Equation 2.12 is the Volume Ratio; it describes vessel thickness. The different ratios correspond to various spherical or rod-type structures. Different thicknesses correspond to different eigenvalues. Another way to think about it is to imagine these different eigenvalues

as identifying three distinct axes. The significant axis will have the largest eigenvalue, distinguishing between thick and thin vessels.

The characteristic response varies as a function of the vessel size. In practice, a user thresholds the vessel response to select vessel size. A higher value shows smaller vessels at the cost of increased noise. The thresholding algorithm is necessary due to the inhomogeneous MRI images, necessitating second-order intensity approaches. The algorithm has extended the concept of multiple scales to analyze several anatomical features found in brain vasculature, such as aneurysms, large bulges in brain arteries, whose ruptures can be life-threatening. So it would serve as a general analysis method for the segmentation of the brain vasculature. We fine-tune the parameters to delineate the principal arteries. Fine-tuning is necessary since different parameter values can affect the apparent thickness of the vessels; high or low values would make them too thick or too thin.

2.1.4.1 Dice Coefficient

Following segmentation and registration, the next step is to calculate the dice coefficient. The dice coefficient is a statistical measure quantifying the similarity between two samples. In the imaging context, the Dice coefficient uses binary masks to calculate the measure. It is also known as the F1 score. The equation is:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{2.14}$$

The highest value is 1.0, meaning perfect overlap, and the lowest is 0.0, meaning the two samples are disjoint. For our purposes, the paper describing the registration process ANTs [8], states that for large structures, the dice coefficient should be around 0.8, while for smaller ones, the value should be roughly 0.6. These values are in the context of whole-brain MRI imaging. For appropriate values in our case, we will base them on a paper using the dice coefficient in 3-dimensional structures [33]. They describe a novel method to obtain a dice coefficient of 0.8-0.9 for whole-brain vessel segmentation, while the other methods were closer to 0.6-0.7. The comparison was between automatic segmentation algorithms and

manual expert segmentation. Our dice coefficient comparison will be between the registration template and the individual cases. While it may not be the most optimal segmentation algorithm comparison, we want to have a quantitative measure of segmentation quality that would control for segmentation errors in dictionary learning since preprocessing errors could propagate and lead to false results. Our cut-off threshold will be 0.65 showing 2/3 overlap between the template and the case, representing cases that are segmented and registered well.

2.2 Registration Verification

For registration, it is necessary to have a proper template. Our template is a statistical atlas optimized for the non-linear nature of registration described in detail in [22], small vessels need non-linear transformations applied such as the SyN method described in [8], necessitating an atlas with high resolution containing fine details.

Registration Template Our registration template [22], [12] includes data from 3 different institutions and has a relatively high resolution of 0.5mmx0.5mmx0.8mm. By comparison, earlier works that have done a similar analysis have conducted the study based on a 1mm x 1mm x 1.75mm space. [20] Essentially, the earlier technique [20] was based on transforming the MRA images into the MRI space, which for the MRI template they have is of lower resolution. The multiple scanners control for multiple acquisition conditions, making our method applicable to many scanners. Table 2.1 shows the MRA acquisition parameters for the institutions used in generating the atlas. The multi-center analysis is important since an algorithm well-tuned to the parameters of one institution may not generalize to a different institution's data. [3] Therefore, it is better to have a robust data processing and normalization pipeline for better inference. The dataset used to create the registration template is also the same dataset we use for our analysis. The parameters include echo time (TE), relaxation time (TR), and field strength (T).

Table 2.1: MRA acquisition parameters for the centers used in the statistical atlas generation. TE stands for echo time, TR for repetition time. Field strength corresponds to the MRA strength.

Dataset	**Field Strength**	**TE**	**TR**
	T	ms	ms
Hammersmith Hospital	3	5.7	16.7
Guys Hospital	1.5	6.9	20
NC Dataset	3	3.5	35

To generate the statistical atlas, the group used the standard protocol for registering MRA images. Their protocol consists of 4 main steps. The first step is to normalize the MRA images using a bright and high contrast one from the dataset. The selected case serves to normalize the images to control for the different institutions' data. Within this first step, they also do a form of histogram matching. This histogram matching algorithm splits the 3d image into 100 regions and tries to normalize the histogram. Histogram matching is a standard normalization algorithm used in many registration pipelines. The second and third steps consist of segmenting the cases. The first step segments the vessel in a centerline representation. The second uses a distance transform, based on a centerline, to obtain a radius representation. The final step is to register these processed images to each of the reference MRI images using rigid registration, and finally non-rigid registration to an MRI atlas.

The registration template is used to verify the registration results. It is also important since the atlas components can help us with the results. It is shown on 2.2. The figure shows the raw 32 bit image grayscale values. We use a brain subsection for the analysis since the brain periphery has smaller and harder to analyze vessels. Analyzing and processing a smaller brain region speeds up a computationally expensive process, with one full brain taking 200 minutes to register using one CPU, but our subsection takes 20 minutes.

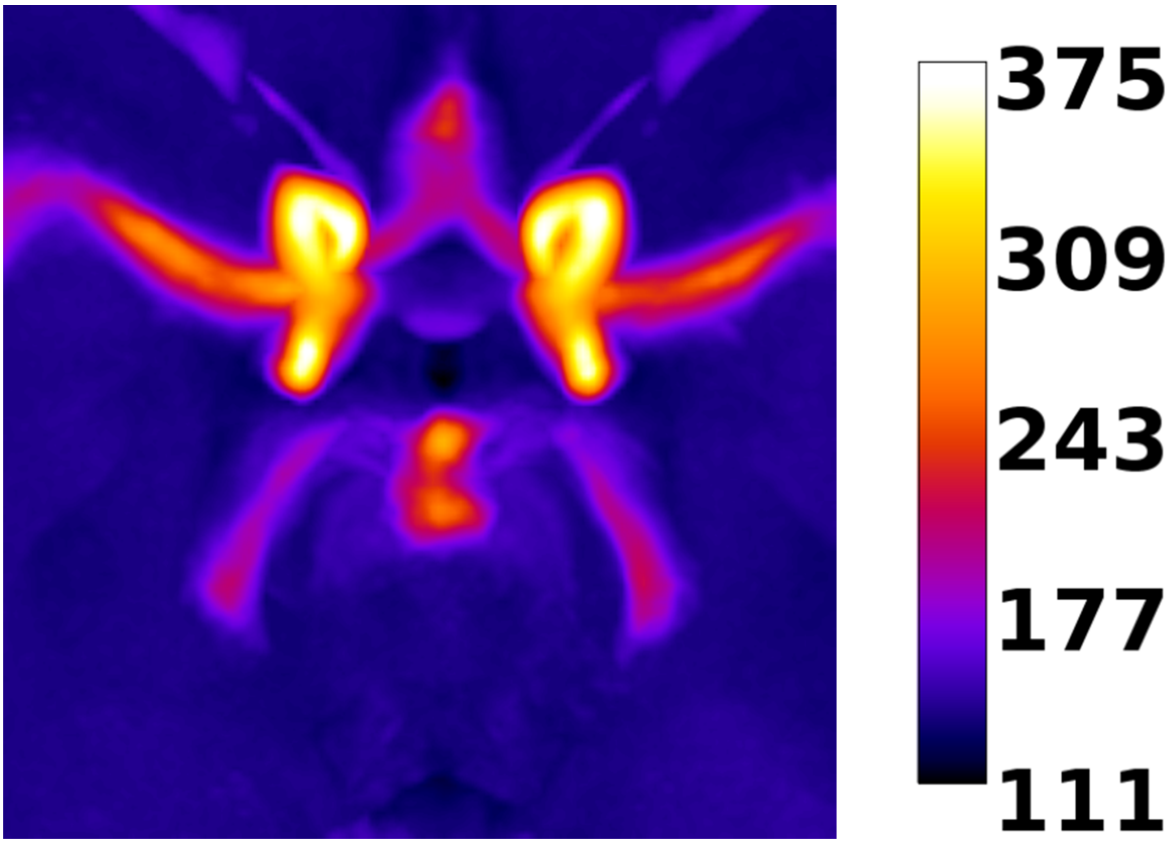

Figure 2.2: Statistical Atlas used in the registration. Shown is the original 32-bit image used in registration. The grayscale scale is the raw scale unaltered from the publication. The area represents the region of interest we are interested in analyzing. This section of the brain contains the main arteries, highlighting the key clinical parts of the brain. Of note is the varying arterial intensities. In our intensity-based registration algorithm, the brighter image regions carry more weight. The brightest regions correspond to the Interior Carotid Arteries (ICA). The lower intensities correspond to the Posterior circulating arteries and Anterior circulating arteries.

2.3 Dictionary Generation Pipeline and SVM analysis

In generating our dictionary, we must check that the cases are sufficiently processed. In the original paper that described the statistical atlas, the authors describe how some cases could not be properly segmented, hence were not used in the final atlas. So in accordance, we use a quantitative quality control mechanism omitting poor quality images or large outliers that obscure our analysis. To do so, we will use the concept of the dice coefficient, [33] which shows a comparison between different segmentation techniques and uses the dice coefficient measured against a manually segmented case and an algorithmic segmentation of the same manually segmented case. We do not have access to ground truth for each case, so our reference will be a case that is the average of the dataset, so it is a template. While an imperfect benchmark, it's an adequate screening tool for poorly registered or segmented cases.

After the segmentation check, we will then set a threshold to select the valid cases. Literature using these datasets often describes the original dataset and mentions the number of data excluded but often gives qualitative reasons for exclusion. The literature includes the statistical atlas itestatistical, in which the authors ultimately used 544 cases out of 700. Our analysis starts with 570 from the same statistical atlas; using our threshold of 0.65 dice coefficient, we end up with 400 for the final analysis; eleven of these had no gender leaving 389 cases for the final analysis. In the statistical atlas paper, [22] they mention that segmentation was the number one reason for case exclusion, motivating a quantitative measure of registration to use for our work. For our case, the registration algorithm converged to a result reasonably registered to the template for all, but some data had faulty segmentation.

For the dictionary atoms, we used the grayscale images but normalized using the algorithm shown below:

Algorithm 1: Normalization Algorithms

Result: Normalized Grayscale image of [0 1]

Grayscale 2-dimensional images from the max projection of the 3D volume **while do**

There are images left to analyze maxpoint=max(image(I))

minpoint=min(image(I)) normImage=(I-minpoint)/(maxpoint-minpoint)

binarize(normImage);

end

Since the images came from different centers with varying signal strengths, we needed to account for the biases for each center. There was a specific group, the IHOP, a psychiatry center that showed images brighter than average. Using the grayscale images gives smoother atoms than using binary images.

The general workflow is shown in Figure 2.3. The flowchart has two parts. The upper left concerns processing and details for each case. The lower right describes the processing steps for the group. We check each case individually, and if the registration completes, we then calculate the Dice coefficient. There are some cases where the registration fails due to noisy imaging or a low contrast image, so in these cases, we apply an image contrast manually or apply the non-local means filter. The lower right concerns the steps taken for the SVM-based classification. These are after the images meet the criteria as outlined in the upper left.

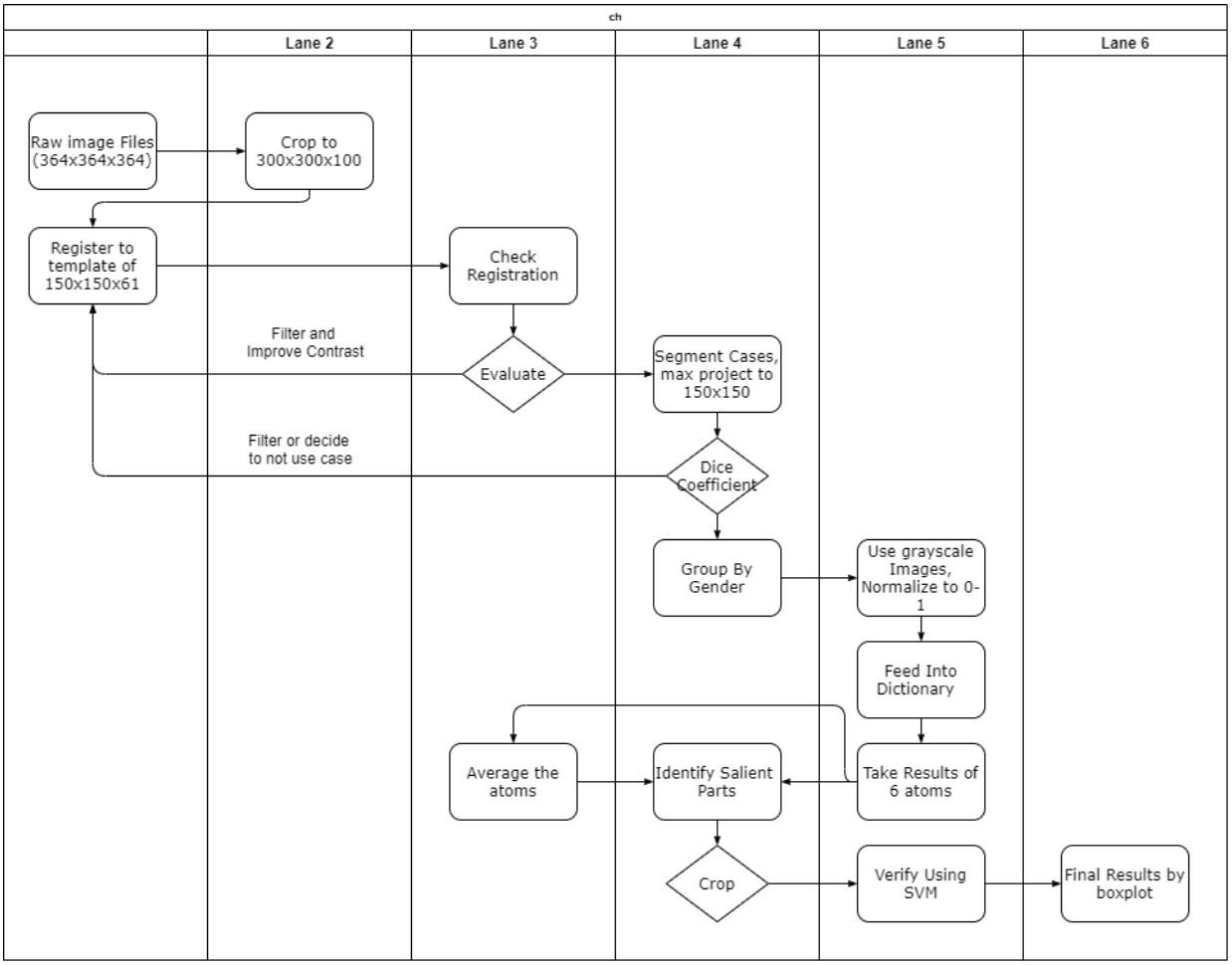

Figure 2.3: Flowchart for the image processing pipeline for dictionary learning and SVM. The upper left side corresponds to steps taken to process individual images. For instance, we would run the images through the registration again if it fails after applying a denoising operation. Afterward, once all the images were registered, the next step was segmentation. The segmentation process filters cases that may have some faulty segmentation. We cast the segmented 3-dimensional case into 2-dimensions for the dice coefficient analysis. After selecting, we construct dictionary atoms for a specific axis, either axial or coronal, then the dictionary atoms are further processed to find salient regions. Based on those salient regions, we analyze the left and right hemispheres using SVM analysis.

2.3.1 SVM Background

In this work, we use the C-SVM implementation as implemented in Matlab 2018, called fitcsvm. [21] This default mode creates a ten-fold cross-validation setup and returns the average validation error. In this implementation, nine folds are used for training, while the remaining fold is reserved as a validation dataset. We then repeat this process 1000 times and compute the mean and standard deviation for each experimental setup, as described in the results section. We pick the value of 1000 since another study using SVM for gender classification [9], used 1000 to compare various SVM kernels for gender classification. We perform the analysis on 2-dimensional binary images that are then flattened into a 1-dimensional array. The fitcsvm algorithm solves the standard SVM binary classification linear program as described below:

$$f(x) = x'\beta + b \tag{2.15}$$

In the equation above, x is the data, flattened to a 1-dimensional vector, β is a vector of coefficients that define the separating hyperplane, and b is the bias term.

The goal is to minimize the linear program:

$$0.5||\beta||^2 + C\sum \xi_i \tag{2.16}$$

Above, C is the cost function, which could be user-defined although we did not include a customized cost function, and ξ is the slack variable to accommodate data points that are further from the separating hyperplane. Fewer slack variables usually result in more robust models.

2.3.2 SVM Significance Analysis

To determine if our results using SVM have statistical significance, we will compare the output of the permutation test with the default labels. As described in [9], the permutation test allows for an estimation of the p value for multivariate analysis. The equation described

in their paper for the emprically derived p value is:

$$\hat{p} = (\#\{e_i < e^*\} + 1)/(P + 1) \tag{2.17}$$

The expression e^* is the error rate with the proper labels. The e_i is the error rate of each one of the 1000 permutations. The number of permutations that are less than the error rate of the classification error of the SVM model with the correct labels determines the statistical significance. The intuition is that when training an SVM model, these cross validation results could be a result of chance, so we shuffle the labels of 0 and 1 and see if the shuffled labels are different from the proper labels. Many studies using SVM use the notion of cross validation as statistical evidence that there is a valid classification, so it is essentially a blackbox [1]. If there is overlap between the two classes, then the results of the SVM learning is not statistically significant and the null hypothesis of classification results happening by chance cannot be rejected. Therefore, in order to achieve a significance value of 0.05 or lower we would need to have at least 1000 trials. This approach is distinct from the univariate statistical method from voxel based mophometry. [1].

The results of the SVM are to see if there is any difference between the left and the right hand sides in terms of discriminative ability. The dictionary learning results, as we discuss in the results section, do show differences between the left and the right hand side. To quantify these results statistically, one possiblity is to consider the statstical SVM validation scheme discussed in this section. The input images for the left and right hand crops are shown in Figure 2.4. To do so, we will use the anova1 function as implemented in Matlab 2018 [21].

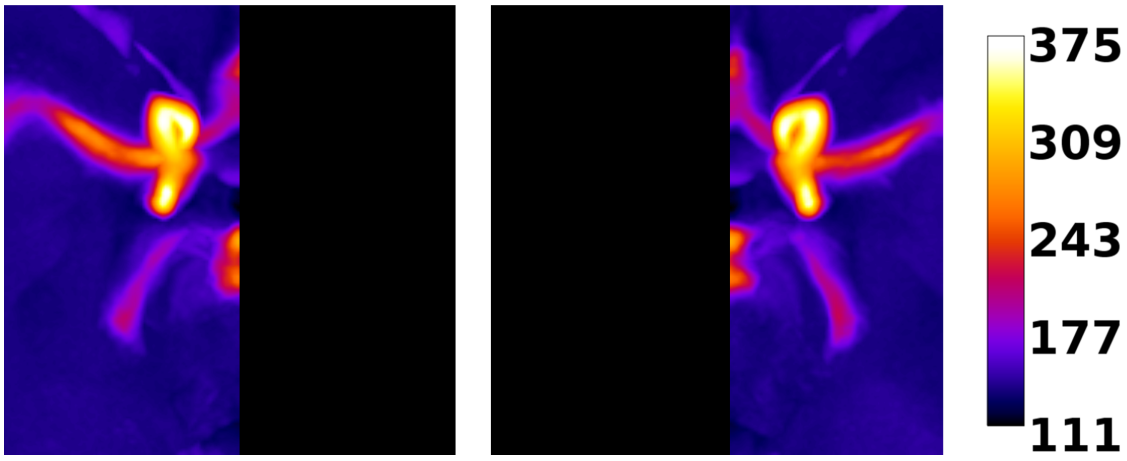

Figure 2.4: Visual description of the crop to determine statistically significant differences between the left and right hand sides. In the results section, the left and right hand sides will correspond to the left and right hand crops as shown in this figure. The crop was taken using the same statistical atlas as showin in Figure 2.2. We take the output of the registered images, segment them, and then crop them as shown in the figure for the final analysis.

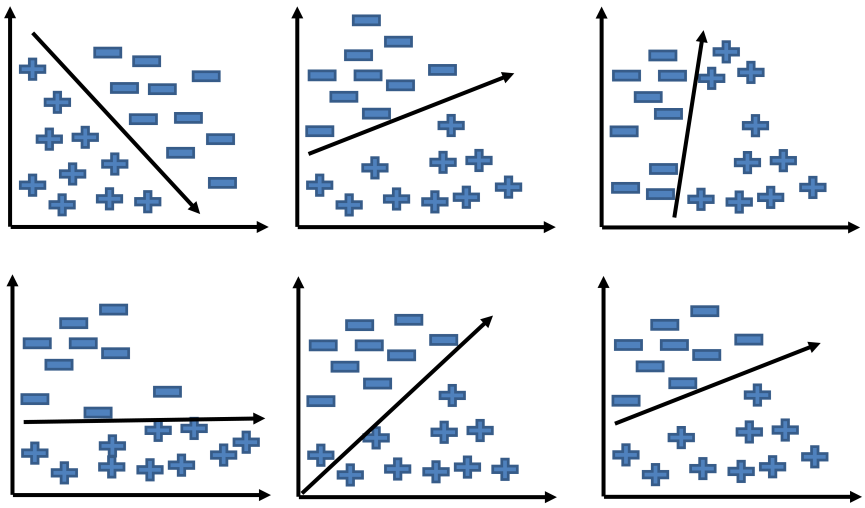

Figure 2.5: Visualization of the random permutation study. The figure is an adaption of [1]. Switching the labels of the training will change the hyperplane, and therefore will change the result of the cross-validation. This change in the cross-validation accuracy comes from the different permutations, and these

CHAPTER 3

Results

The results section has two parts. The first is registration verification showing quantitative results in a mainly graphical form showing how we validate registration and segmentation. The second shows the dictionary patterns as well as SVM classification results.

3.1 Registration Verification

We start with the registration results of various cases and show that the data has acceptable registration. Key to our strategy is a statistical atlas generated from the same dataset used in this work. The registration results should give us similar outcomes to the registration atlas. Namely, the authors have provided a probability atlas that shows the vessel probability in a specific location. The probability atlas is the average projection of all cases. The probability distributions between the two of them should be almost the same. Specifically, the probability of a vessel in the Interior Carotid Artery (ICA) location is over 0.8, while for the Posterior Cerebral Artery (PCA), the value is between 0.2 and 0.3. The variation is highest further from the ICA arteries, which are the largest. When viewing the registration template, this area is the brightest when considering an average projection. Figure 3.1 shows our experimentally generated statistical atlas scaled with the probability found in the statistical atlas. The statistical atlas from the paper is on the right. The probability distributions, particularly in the peripheral arteries, are similar between the two cases. The latter finding is important since the peripheral arteries are the hardest to segment due to their size. The corresponding probability distributions lend proof of the registration. The ICA arteries in the center are the most prominent and easiest to register. The most variation

lies within the periphery; drastic differences in those arteries may indicate some systematic bias. However, there could be subtle variations between cases that would not show using this analysis. While the broad picture shows reasonable registration quality, we then discuss individual case analysis and selecting individual cases.

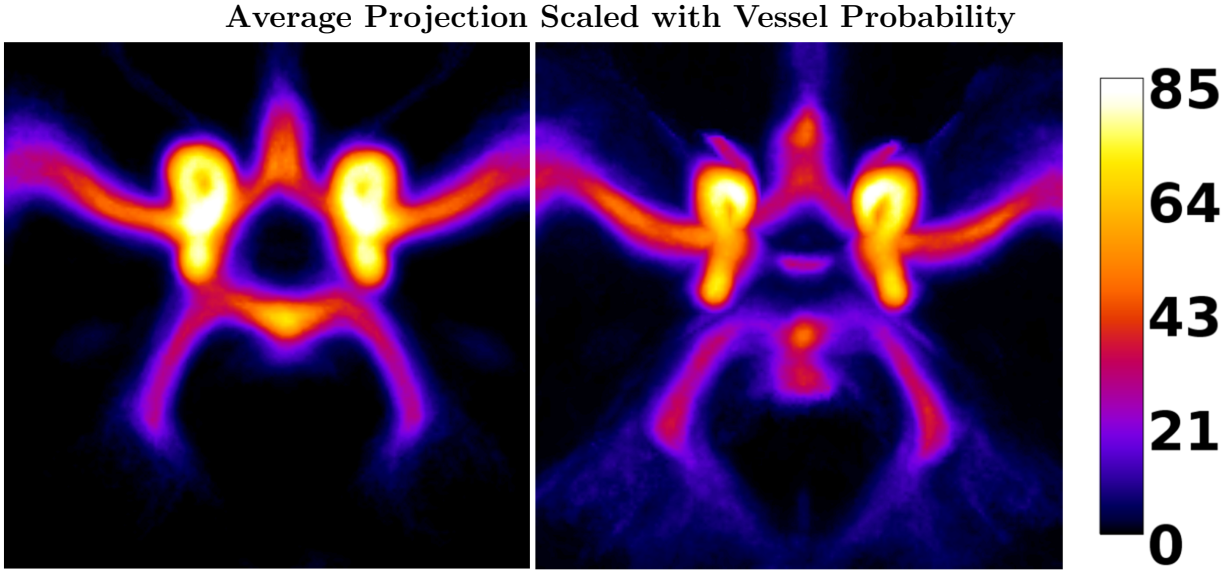

Figure 3.1: Average projection of the cases. We start with binarizing the images using the vesselness methods. Then the total of cases is scaled with the probability as expected with the statistical atlas. These steps essentially recreate a statistical atlas, allows us to compare our results with a published one. If the registration is successful, we should see a probability distribution similar to the statistical atlas.

While we have shown qualitatively that the images have some common reference point, the average of all the images converges to some common ground truth in the registration, we also quantitatively measure the distance to the reference point. Figure 3.2 shows the comparison between the registered cases. The dice coefficient threshold was above 0.65 for the final analysis; of these cases, the average Dice Coefficient is 0.70 with a standard deviation of 0.03. The principle motivation for 0.65 as the threshold is that it represents 65 percent overlap which is 2/3; the high probability regions have a probability average of 60, then these cases should be considered well-registered images.

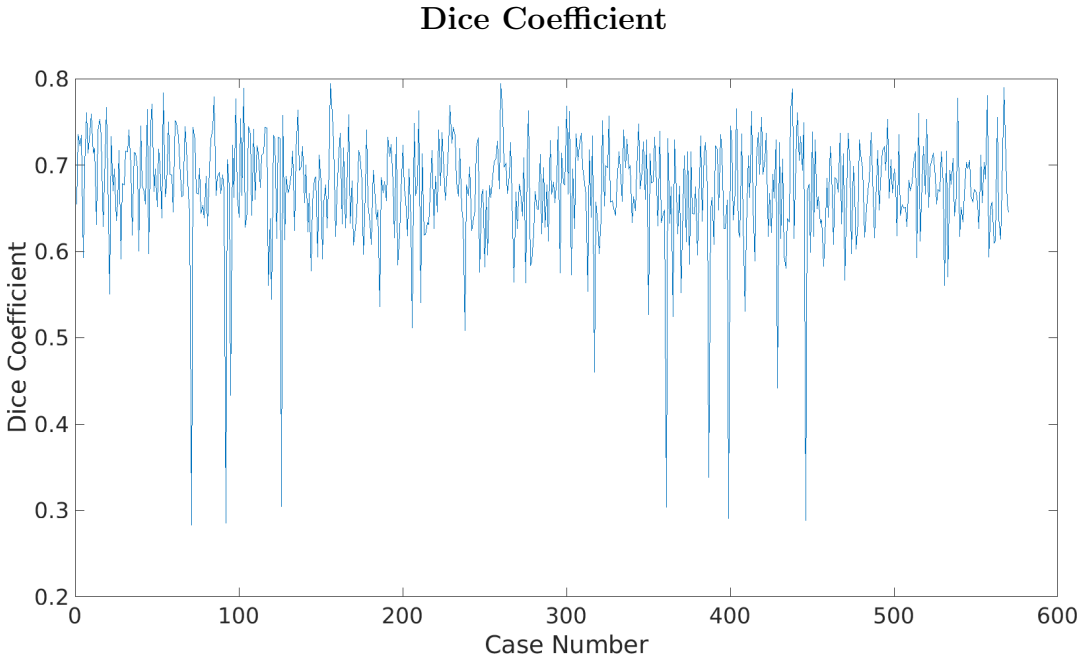

Figure 3.2: Dice coefficient to show the variation of the various cases in their dice coefficient. The graph gives an idea of outlier cases and suggests possible thresholds. While a few low-quality cases are owing to poor registration or segmentation results, most cases are good enough for further analysis.

Table 3.1: Statistics regarding the Dice Coefficient Results

Mean	**Standard Deviation**	**Outliers**	**Min**
	value	*value*	*value*
0.67	0.32	1.04	0.0032

To verify the accuracy of the registration, we will consider the use of the Radon transformation. The Radon transform is an integral transform from one plane to the set of lines in that plane, allowing one to take a plane and determine its rotation. So we have the image and can consider the rotation per angle. In a sense, it is analogous to an angle transform as a Fourier transform is a frequency transform. We use the cross-correlation function to compare different transforms. When evaluating the cross-correlation function, a cross-correlation of 1 means that two images are highly correlated. 3.3 shows the results for the axial viewpoint.

The average value for all data was a 0.87 maximum correlation coefficient with a 0.06 standard deviation. The main purpose of the radon transform was proof that the registration converged to a central axis as the method is sensitive to off-axis rotations. Based on the radon transfrom there was reasonable registration so the only figure we used for determining the cases to select was the dice coefficient. Any values that were outliers using the radon transform were also outliers using the dice coefficient metric.

Figure 3.3: Radon transform graph showing the values of the maximum correlation between the average transform value and that of an individual case. A higher correlation would mean that the rotation angle between them is similar, which means that the registration would be better. A lower correlation would signal that the rotation angle differs from that of the average case.

Table 3.2: Statistics regarding the Radon Results

Mean	**Standard Deviation**	**Outliers**	**Min**
	value	*value*	*value*
0.87	0.06	25	0.43

3.2 Extracted Dictionary Patterns

3.2.1 Our process and the axial viewpoint

Next, we will show the extracted dictionary patterns. The patterns are the main result of this work, with the bright regions showing the most distinct areas between males and females, with females having a higher probability of a complete circle of Willis. [34] We also see a discrepancy between the bilateral sides; the right side has a higher likelihood of completion.

The process to generate these images follows from the flowchart described in Figure 2.3. Briefly, after preprocessing and the selection criteria described in the first part of the results, we arrive at the dictionary patterns shown in Figure 3.4. These six results are the raw output of the program after reshaping the vectors to the 2-dimensional image. The montage also shows the atom importance order, with the top left the highest; the bottom right the lowest. The montage allows observation of salient points such as the orange background that does not contain vessels and the dark and light regions containing vessels. To interpret the meaning of these light and dark vessels, we need to know the subtraction that taken place. If we were to do image subtraction of A-B, it would differ from B-A. So to interpret them, we also need that context. In this case, the subtraction order is Female-Male, so the bright regions indicate probability regions where the female has more vessels likely than in males. The opposite is true for dark areas. One observation is that their dictionary atoms are somewhat redundant, particularly on the right-hand side, specifically the MCA.

Upon observing the results, the next step is to identify salient regions. When considering the salient rate, we will consider values above 0.5. The results are in Figure 3.5. By taking the values above 0.5, we can see the values that distinguish one population from another. The subtraction convention is the same as Figure 3.4, which is Females-Males. We can see more clearly which parts are more common while taking out the background. While it was harder to see the left and right differences in the original figure, here, they are apparent. In

addition to these differences, we also have the same process for the Male-Female differences. We use the same definition and processing for the Male-Female differences. The pattern appears across most of the dictionary atoms as shown in Figure 3.4 which shows the female-male difference, representing patterns more common in the female population. The MCA on the right-hand side is more prominent than in the male cases.

In addition to the Female-Male subtraction, we also consider the Male-Female case. Figure 3.6 shows this latter case . At first glance, the dictionary patterns are hard to decipher, with seemingly noisy images. It is easier to notice the differences with a threshold of 0.5 as seen in Figure 3.7. Here, we see that the ICA, located in the center of the circle, is the most distinct feature in males, as it is the brightest region in the first atom and shows up at another atom. The dictionary pattern indicates that the ICA is larger in males on average than in females. Also, note that the right-hand side shows the most variance, a similar trend as seen with the Female-Male cases, which would make sense. Considering the MCA differences, we would see the orientation more common in females in the Female-Male dictionary; and the orientation more common in the male in the Male-Female dictionary.

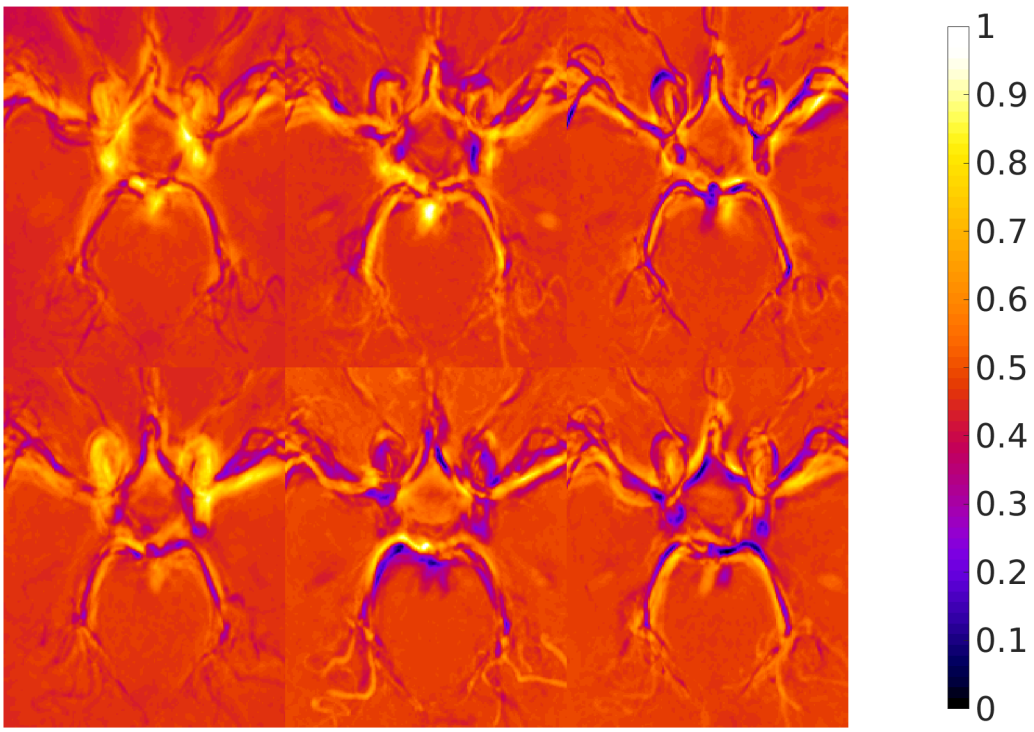

Figure 3.4: 6 dictionary atoms of the Female-Male difference images. To create this image, we train the dictionary algorithm on the image set representing female-male differences. If we were to take these images, we would see that female-male images would result in positive values where a vessel is more likely in the female case than in the male case. The atoms are scaled from 0 to 1.0, with 0.5 being background, and values further from 0.5 are regions of higher variance. The dictionary is the difference of Female-Male; bright regions correspond to structures more common in females versus males, and the dark regions the opposite.

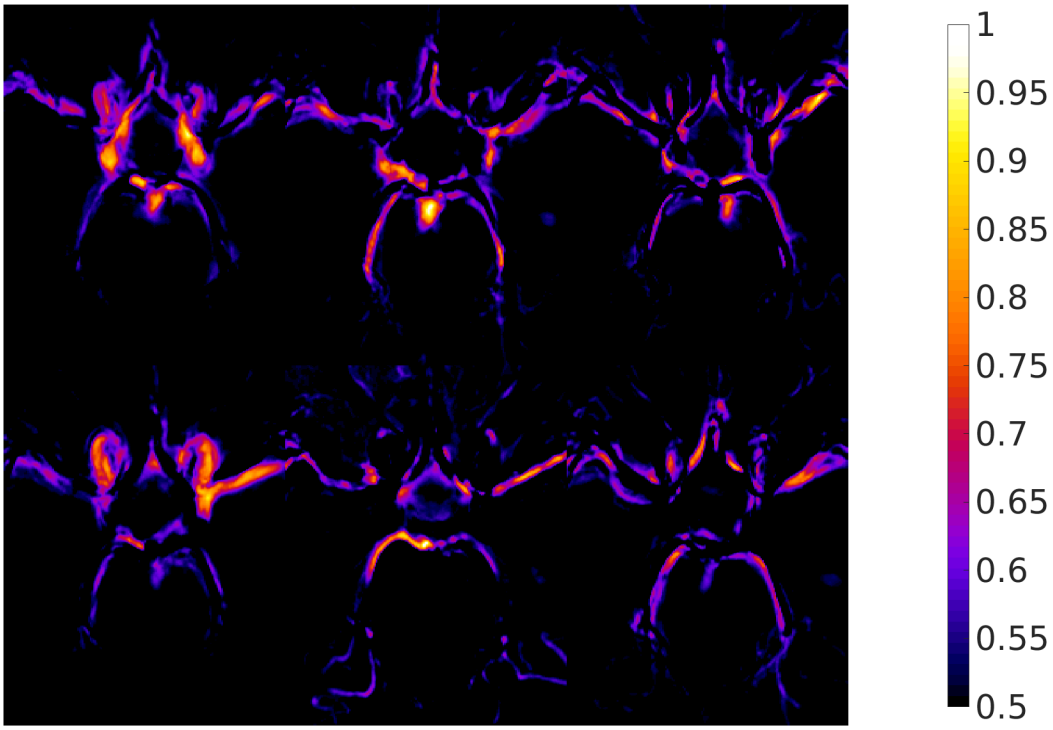

Figure 3.5: Same as in Figure 3.4 but thresholded to only show values greater than 0.5. The threshold value of 0.5 since we want to show structures found in one population versus another. Focusing on values greater than 0.5 shows the most salient comparisons. Here it is easier to note some key differences. In particular, the left-hand side of the MCA, which are the arteries extending left and right from the center of the circle, is brighter than the right-hand side, a key differentiator for the two populations.

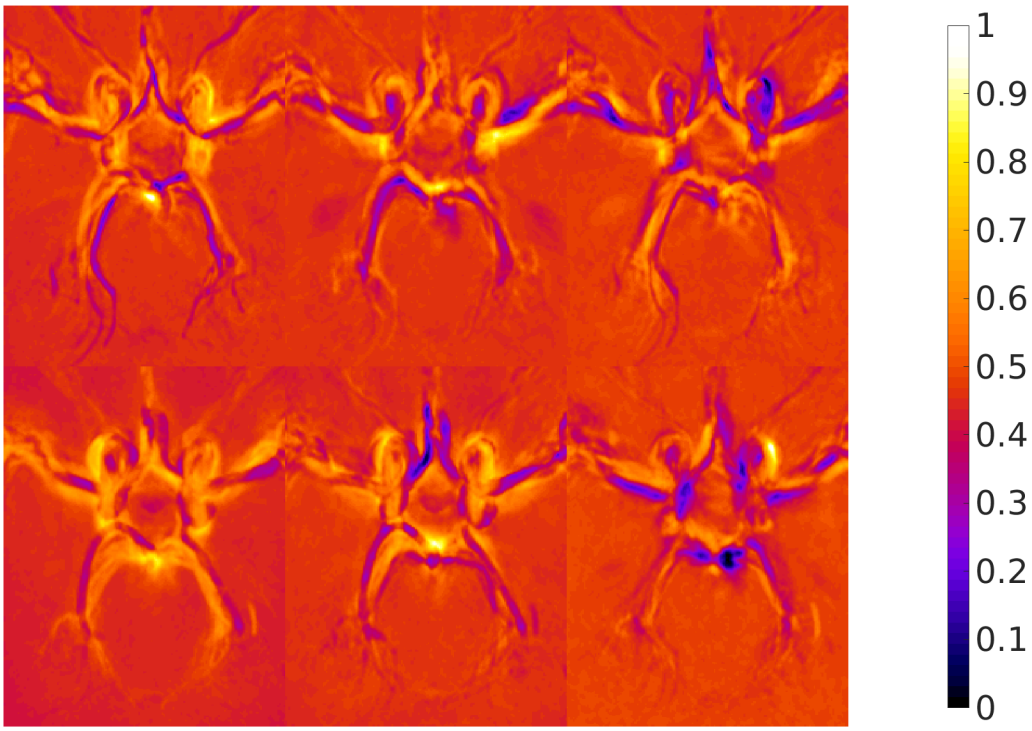

Figure 3.6: 6 dictionary atoms of the Male-Female difference images. The image in the top left shows the most significant atom, and the least significant atom in the bottom right. To create this image, we train the dictionary algorithm on the image set of Male-Female differences. If we were to take these images, we would see that Male-Female images would result in positive values where a vessel is more likely in the male case than in the female case. We do not take the absolute value of the images; the raw dictionary values would then show where the largest differences lie. The atoms have the same convention as described in previous figures.

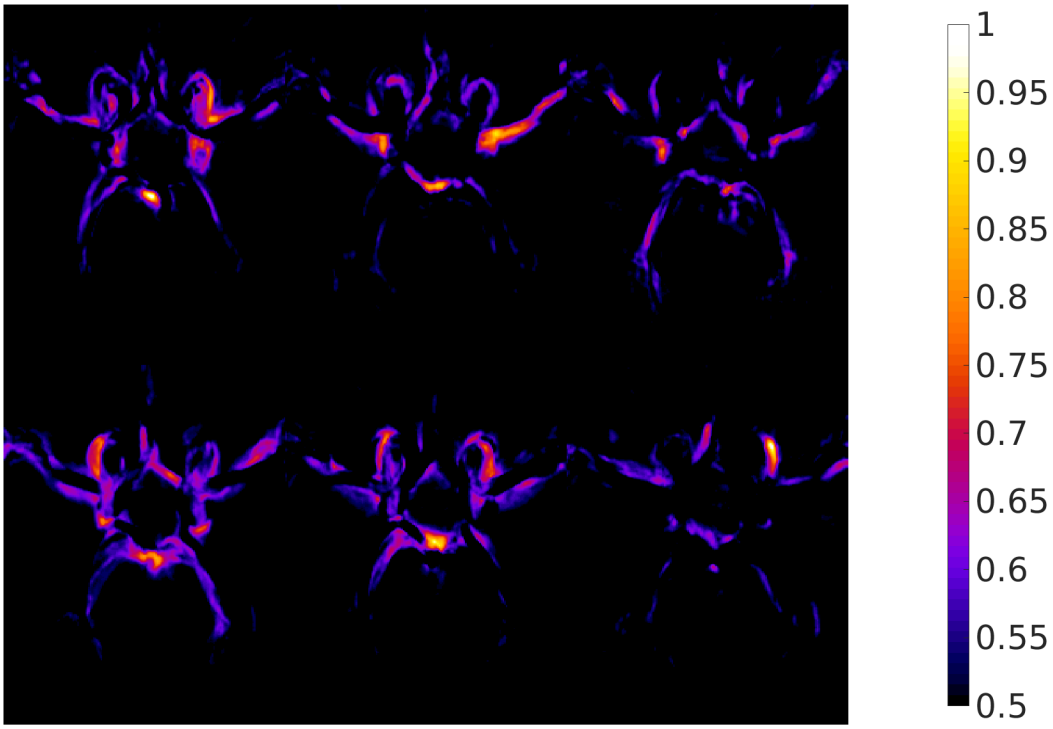

Figure 3.7: Same as in Figure 3.6 but thresholded to only show values greater than 0.5. The subtraction order is the same as in 3.6. Focusing on values greater than 0.5 shows the most salient comparisons. Here it is easier to note some key differences. The differences are not as significant as the female-male comparison, bright regions are within the ICA, seen on the top left and the bottom right.

3.2.2 Average Dictionary Atoms for Axial Viewpoint

The next step in the results is to average these dictionary patterns. The concept is to identify the most salient differences by keeping features found in most atoms. We average the raw values of all then afterward use the same threshold of 0.5 to see observed patterns in figure 3.7. We can see the most important features present in females but absent in males. We also show the reverse in 3.8. From the average Female-Male regions, we see a clear pattern emerge in the right-hand MCA; in females, it is more prominent than in males. We study this key feature later and compare it to the statistical atlas in the next section. In this section, we use this salient difference and try to find a way to quantify this difference. Other differences from the Male-Female side concern the circle of Willis size. While females are more likely to complete the Circle of Willis, males have larger sizes since we can see a larger ICA in males than in females, as shown using dictionary learning. We hypothesize that we find statistically significant differences between the left and the right-hand sides. We do this using an SVM classifier. While not the highest performing, it can show differences in classification depending on the inputted data.

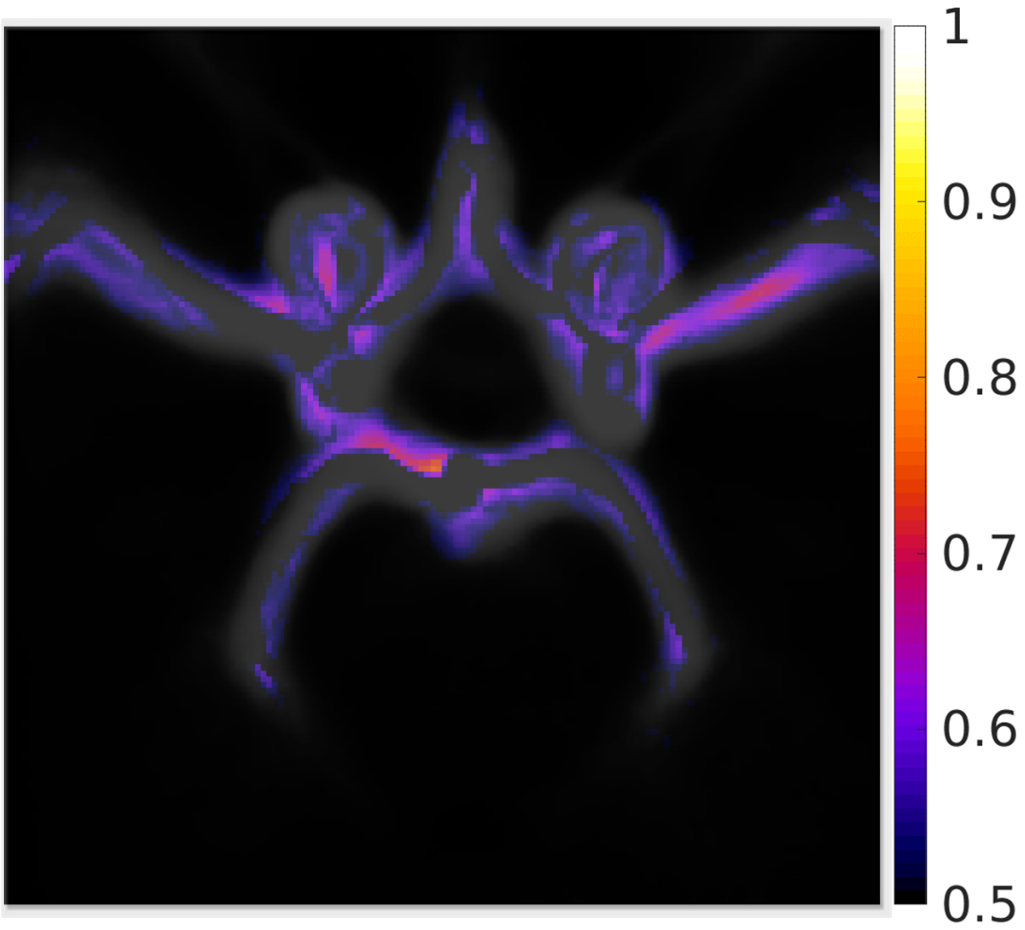

Figure 3.8: In this image, we take the six atoms then average them to see the regions with the most variance. The purpose of averaging is to highlight features found in several atoms. Here we can see that the right-hand side of the MCA is the region of the highest saliency as discussed in 3.6. The averaging does make the differences appear less pronounced, but a difference found in many different atoms is more significant than if it showed up in only one atom.

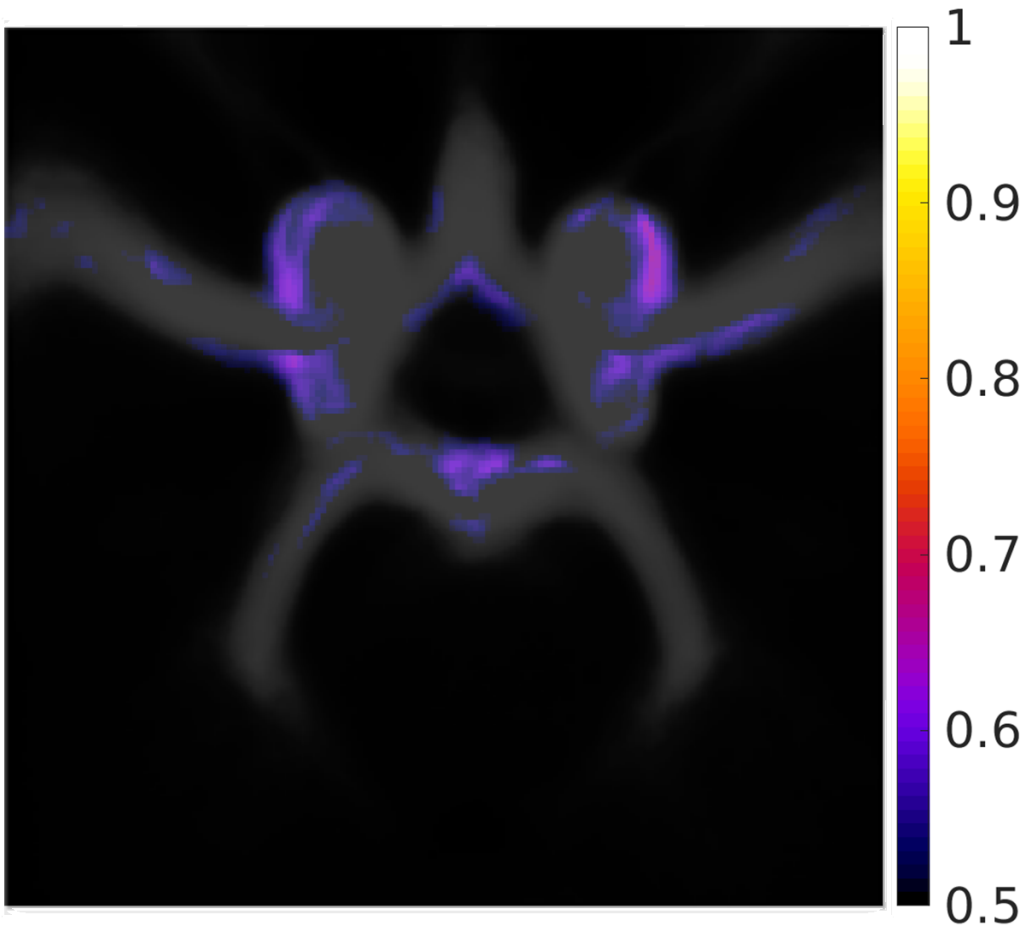

Figure 3.9: Using the same process as for making 3.8, we average the male-female dictionary atoms, the result would then show the structures more common in males than in females. Here we can see that the ICA is larger than in females in general. The other structures did not show significant differences.

3.2.3 Coronal Viewpoint

In 3 dimensional imaging, we have to consider different spatial viewpoints for analyzing volumes. This other viewpoint is known as the coronal viewpoint, extending from the front to the back of the head. The axial viewpoint extends from the bottom of the brain to the top of the brain. There is also a third viewpoint known as the sagittal, extending from the right-hand side to the left-hand side. As we will see, the results of our dictionary patterns vary in this viewpoint, and in the next section, we attempt to quantify these patterns and reconcile these viewpoints. As shown in 3.13, the coronal viewpoints look more symmetric on the left and right-hand sides than in the axial view. We discuss statistical differences in classification accuracy between the two axes, which can help interpret the dictionary patterns.

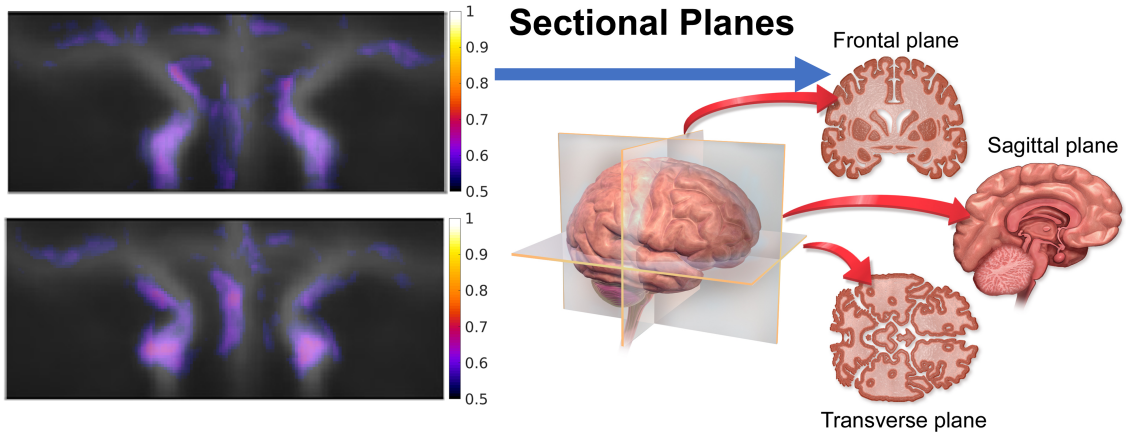

Figure 3.10: Left, dictionary atom comparisons of patterns more common in females at the top, those more common in males at the bottom; right, 3D map of the brain showing the perspective of this slice, from [28]. We average the six atoms found in the coronal views using the same technique for figures 3.8 and 3.9. The top shows Female-Male dictionary atoms, while the bottom shows Male-Female subtraction. The brightest regions correspond to differences in the ICA angle, with males having a larger ICA, while in females, the ICA extends. This ICA difference is similar to the first atom that we see from 3.4.

3.3 SVM analysis and Significance

To test the validity of the dictionary atoms, we will use an SVM classifier and first verify that there is a statistically significant classification. We have not tuned the SVM classification; the default parameters yielded statistically significant results on their own. The first test compares regular data and random permutations of the same dataset. For this experiment and subsequent experiments, we will use binary images in the SVM classifier; since this machine learning algorithm performs best with normalized or binary data where all values in a vector are between 0 and 1. We will show a table with the results then detail the results using boxplots. We use a basic SVM using a standard kernel with default parameters as a MATLAB library [21], called fitcsvm. We discuss the coronal and axial results separately then consider their differences. The key is to gain intuitions behind the saliency maps and how they can identify differences between regions.

3.3.1 Axial View

We start our SVM analysis with the axial viewpoint. The results in Table 3.3 are also represented in boxplot format in figures 3.11 and 3.12. The first comparison, shown in the boxplot of 3.11 and the first two rows of Table 3.3, is between random permutations, where we shuffle the labels and do validation tests 1000 times. The results show a significant difference between them, with an error rate of 48 percent with the permutations, essentially random chance, and a 38 percent error rate using the original labels. The split between male and female subjects is 42 percent and 58 percent. So we can see that the random permutations give results consistent with random chance. The error rate is 10 percent lower with the proper labels, which is comparable to earlier SVM work in classification between males and females [9]. Figure 3.12, and the bottom two rows of Table 3.1, show the left and the right-hand sides of the axial viewpoint. As discussed earlier in this section, we can visually observe differences between the left and right-hand sides, particularly in the MCA. We want to quantify these differences by considering classification strengths when considering

only one side or the other. Thus we define a binary mask, where we blank out part of the hemisphere. Right and left classifications mean that only the right and left-hand sides are intact; the other side is blank. Using this logic, as seen in 3.12, we see some differences in classification in the right and left-hand sides. The error rate is lower using the right-hand side rather than using only the left-hand side. The error rate difference means that the right-hand side has more discriminatory power than the left-hand side. This finding makes intuitive sense when considering the dictionary patterns shown previously. The right-hand side, as seen in Figure 3.8, is brighter than the left-hand side. According to the dictionary learning scheme, the right-hand side shows more salient regions. The SVM results also support that notion, lending credibility to the dictionary approach for finding salient regions. We quantify the statistical significance on Table 3.4, which shows the ANOVA results when we consider 10,100, and 1000 cases. The results are not significant until we consider 100 cases, which coincides with SVM results in another paper that performed 1000 SVM permutations using cross validation [9]. These results do not disprove the null hypothesis that there is no statistical significance between them, this is a proof to show how many samples we would need to show that there are differences between the means of these two cases. Also 10 permutations would not be enough trials to determine statistical signficance, since more data would be needed in a permutation sense to determine if outcomes can arise by chance. The results with 10,100,and 1000 cases shows how we would need to simulate 1000 permutations for statistically significant results. Table 3.7, shows the calculated p-values for the cases of the entire image, the left and right crops. For all three cases, we follow the definition of the p-value as in [9]. There is a statistically different value of the proper labels and of the random permutations, but for the left, there is a greater probability that the error probabilities can come from chance. The results give more quantitative validation to the dictionary method of finding statistically significant saliency regions.

Table 3.3: Results of SVM Permutation with Axial View.

Condition	**Mean Error Rate**	**Standard Deviation**	**95 ± C.I.**
	%	%	%
Correct Labels	37.01	1.31	0.08
Random Labels	47.41	3.07	0.19
Left Side	40.73	1.49	0.09
Right Side	39.31	1.46	0.09

Table 3.4: Statistical Analysis for the Axial Dimension data. The test performed is the ANOVA test, with different sample sizes of 10,100, and 1000 samples sampled randomly from each distribution.

Sample Size	**P-Value**	**F**	**MS**
10	0.32	1.04	0.0032
100	6.36e-10	42.26	0.01
1000	1.60e-92	462.9	0.1

Table 3.5: Estimated P-values for the axial dimension comparison between the left and the right hand sides.

Crop taken	**P-Value**
Entire Image	0.0009
Left	0.006
Right	0.002

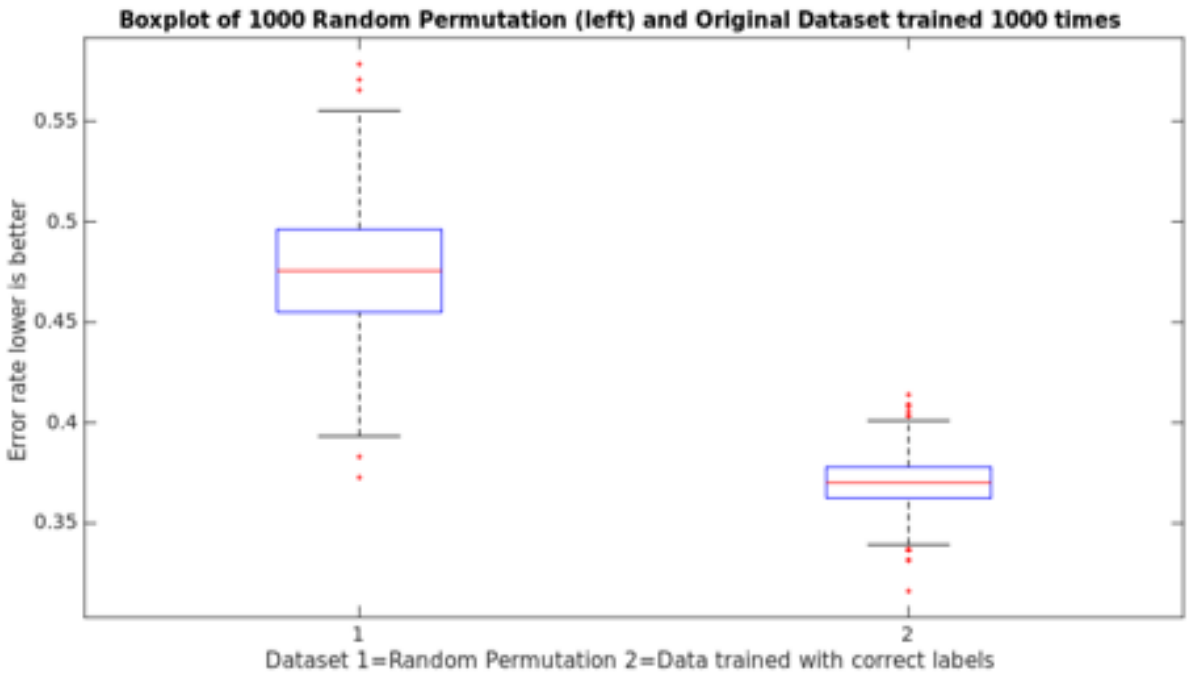

Figure 3.11: SVM comparison using cross validation with different training/validation splits showing 1000 trials/permutations. The main takeaway is that there is a statistically significant difference between the proper labels and the randomized labels.

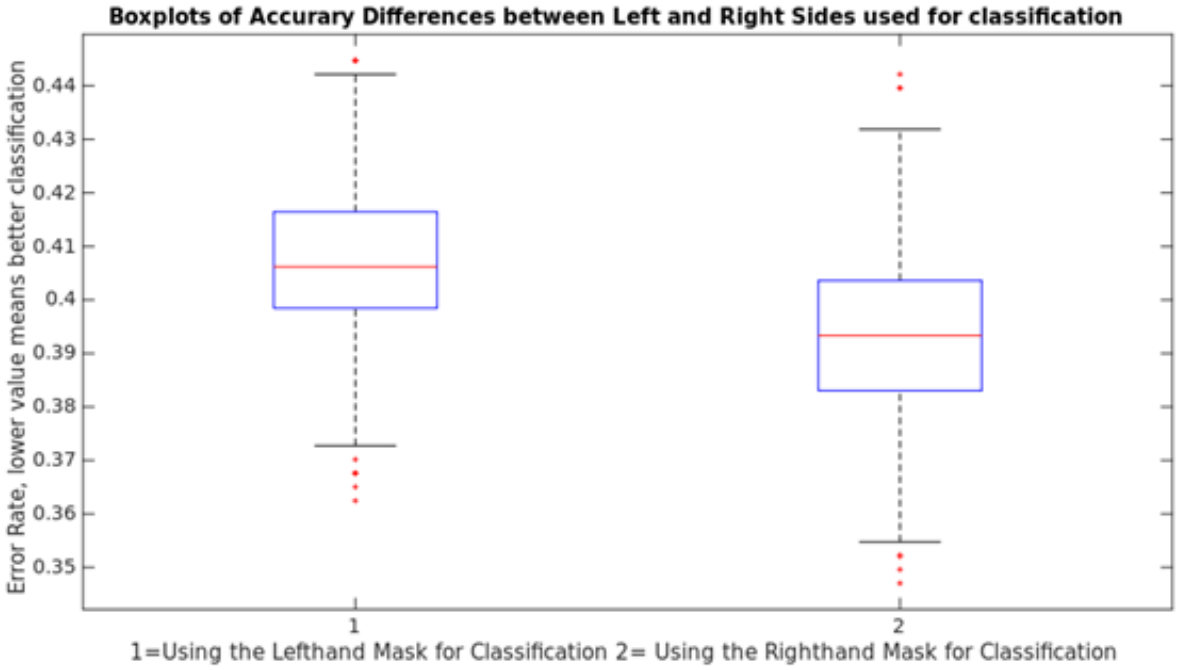

Figure 3.12: SVM comparison using cross-validation with different training/validation splits showing 1000 trials/permutations. In this case, we compare the classification strength between the left and the right-hand side. The left-hand plot shows the classification results if we zero out the right-hand side and only keep the left-hand side. The right-hand plot was vice versa the left-hand plot. We see that the classification error is lower when we only use the right-hand side versus only using the left-hand side.

3.3.2 Coronal View

The coronal viewpoint is shown in Table 3.6, and the corresponding boxplots are shown in figures 3.13 and 3.14. Compared to the axial view, the coronal view shows less accuracy in classification. However, there is still a statistically significant result in classification. The random labels permutation test accuracy is similar to the axial view; the permutation test is a valid control demonstrating that the classification results are different concerning the different viewpoints. In particular, the left-side only standard deviation is higher than the right side, meaning that the statistical difference is low for both hemispheres. The left and right-hand sides for the SVM classification do not show a significant difference; the left and right-hand sides do not show striking visual contrasts. We can visually see these results on the boxplot and we also see this through the statistical analysis as shown on Table 3.7. The F score is less for the cases of 100 and 1000 cases, as compared to the axial view. The 10 cases is larger but since these results are not statistically significant, we focus on the 100 and 1000 cases. Lastly, we consider the estimated p-value of the SVM classification. The estimated p-value calculation follows the same logic and is found on 3.8. We see there is no statistically significant difference between random permutations on the coronal viewpoint when considering the left and right hand sides individually. This is in contrast to the axial view, where even though there was a statistical difference between the left and right hand sides, in terms of p values, there was also sufficient information to reject the null hypothesis that the data is independent of the labeling.

Table 3.6: Results of SVM Permutation with Coronal View. C.I= Confidence Interval

Condition	**Mean Error Rate** %	**Standard Deviation** ± %	**95 % C.I.** ± %
Correct Labels	41.66	1.33	0.08
Random Labels	47.76	3.13	0.194
Left Side	44.33	1.54	0.09
Right Side	43.55	1.36	0.09

Table 3.7: Statistical Analysis for the Coronal Dimension data. The test performed is the ANOVA test, with different sample sizes of 10,100, and 1000 samples sampled randomly from each distribution.

Sample Size	**P-Value**	**F**	**MS**
10	0.1369	2.39	0.00053
100	1.41e-05	19.82	0.005
1000	4.85e-43	198.45	0.045

Table 3.8: Estimated P-values for the axial dimension comparison between the left and the right hand sides.

Crop taken	**P-Value**
Entire Image	0.0009
Left	0.12
Right	0.06

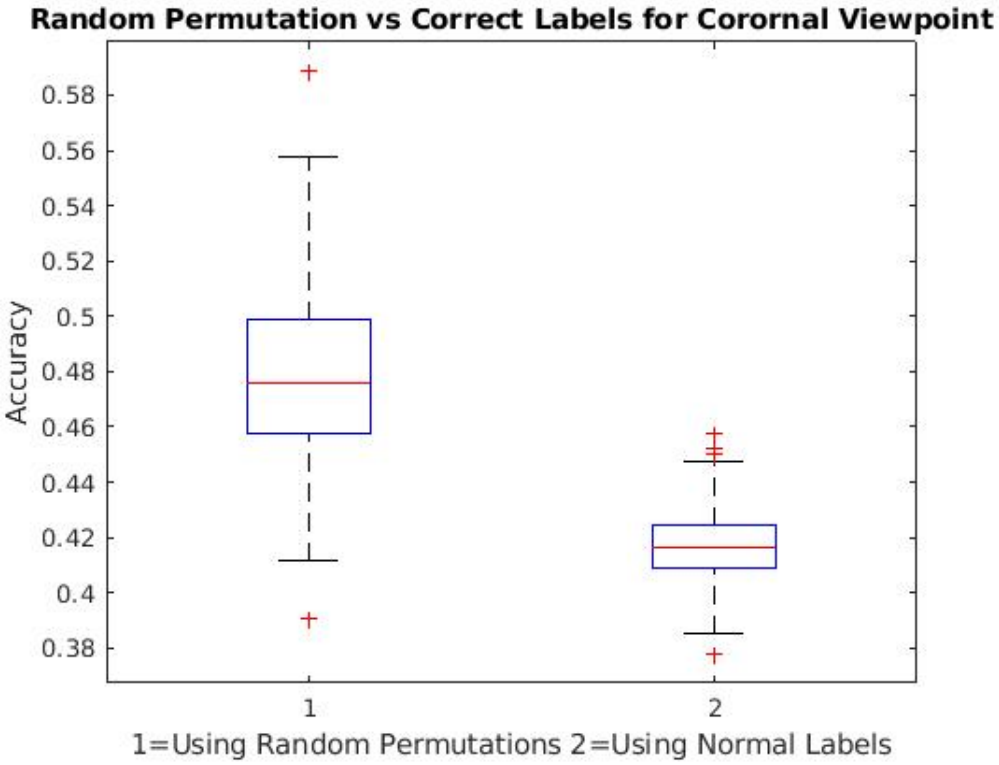

Figure 3.13: Coronal Viewpoint classification. This figure follows the same logic as figure 3.5, as we use random permutations of the labels, but we run it for 1000 trials of 10 fold cross–validation. We see differences in the classification accuracy, with an average classification error of 0.42 for the coronal viewpoint as compared to the axial viewpoint of 0.39.

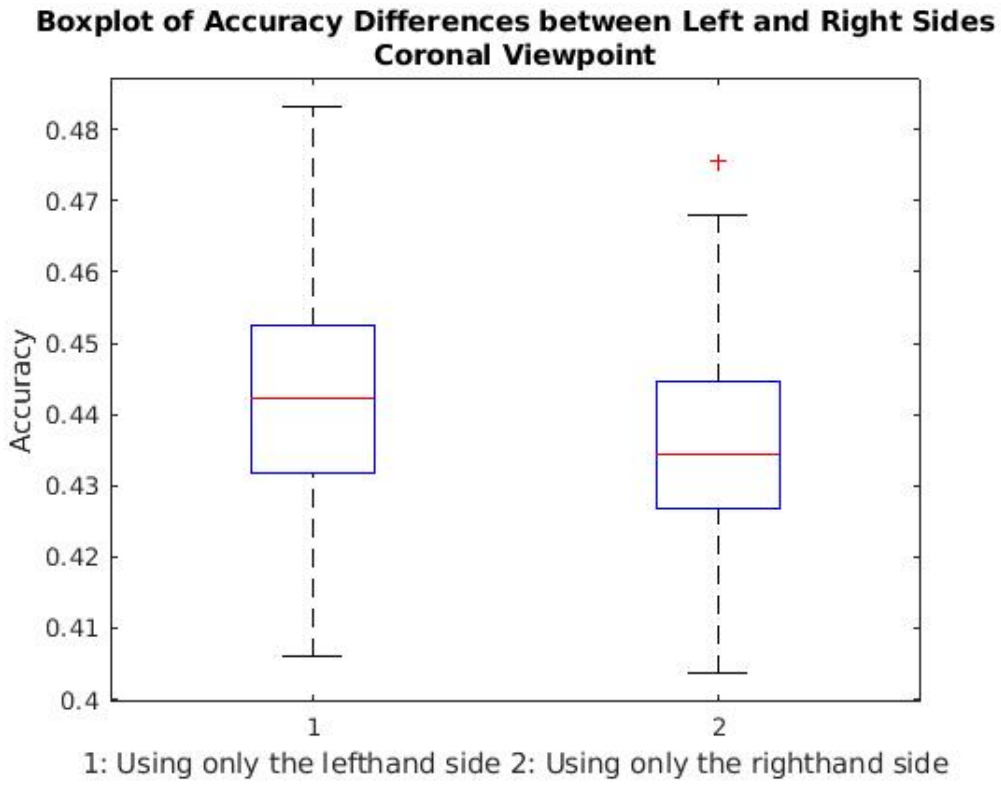

Figure 3.14: SVM comparison using cross-validation with different training/validation splits showing 1000 trials/permutations. We compare the left and right-hand sides of the coronal viewpoint; the left and right-hand sides are not distinct, in contrast to the axial dimension. The difference between the right and left-hand sides shows in the boxplot and the dictionary atoms. There is good correspondence between the saliency map and the results given by classification through SVM.

3.4 Dictionary Variation

The next step in our analysis is to see the variations in the dictionary as we rerun the same analysis. We chose the parameters that we did since the RMSE, or the root mean square error was consistent throughout several iterations. After about 20 iterations, the RMSE was constant, but we ran the dictionary to about 50 iterations. In practice, running the dictionary learning algorithm stops when the RMSE shows no change or atoms are no longer replaced. Although we have run the same algorithm and obtained the same RMSE, the resulting dictionaries do vary somewhat. However, the general dictionary structure is preserved. We also create a statistical map to determine the statistical significance of our dictionary. The inputs will be the Female-Male and Male-Female averaged dictionary atoms as discussed earlier. We use the average of the atoms since the average of the atoms can give clues as to the most salient features in the image, and somewhat account for variations in the algorithm running. Every iteration of the dictionary is slightly different but the visual features are similar. In other techniques using multivariate analysis, the other papers observe some variations with the runs, necessitating preprocessing with a gaussian kernel to smooth the results. [5] We use the notion of permutation testing as described when we implemented the SVM analysis. Recall that we find statistically significant classification results when considering the axial viewpoint but not when considering the coronal view. In our case, for permutation testing, we train the dictionary with 1000 random permutations of the training labels, and then see which of the dictionary atom values are greater in a pixel by pixel manner. While this is a simple concept, it is not mathematically rigorous as there is variations in training the dictionary and these can only really be estimates of statistical significance. [1] There is no method of statistical validation using dictionary learning that does not involve permutation testing of some kind. [2].

The results of the dictionary permutation testing are shown in Figure 3.15. The statistical significance is greater for the Female-Male comparison, showing more robust comparison between the parts present in females than those present in males. This is also in accordance

with the SVM results between the left and right hand sides. The right hand side had more discriminatory features than the left hand side. So there are two pieces of permutation testing that show the asymmetries. However, an important thing to note is that we have not quantified the effect of the dictionary learning. The higher values do indicate the increase presence of a vessel in one location versus another but it does not show how often this effect is. The only thing we have done is to show that globally, there are differences within populations and that the information content globally is different within different regions of the image. This is because the classification cross validation accuracy is a measure of information rather than of activation differences. [5]. So we will implement population based inference that would indicate quantitatively how prevalent an effect is in the population. [5].

We also compare the cases by transforming the average dictionary atoms into z scores. We take the average measure of the dictionary atoms and subtract the mean. The absolute value will give the variance of the method. The distance from the mean would show significance, and these will then be compared to the previous analysis performed, especially for SVM.

3.4.1 Correcting for Possible Confounds

In using multivariate pattern analysis, it is essential to differentiate between the two goals of finding functional differences between groups and constructing a high accuracy classifier. These are two different tasks, and depending on the problem to solve, one might take precedence over the other. When first using these techniques, the neuroimaging community found out that there were SVM classifiers that would achieve high classification accuracy [15], but would use features that would not indicate functional differences between groups. So there is a need for statistical tests or simulated data to control for possible confounds. Another purpose for doing so is to increase the feature sensitivity, [27]. While we have statistically significant regions for dictionary learning, it is still possible that there are registration confounds. These confounds could also be specific to demographic groups. There are a few papers that describe motion confounds [4], Among the key results is that males

display larger in scanner motion than females do. These artifacts are closer to the skull, and so we expect more discrepancies in outer vessels versus vessels closer to the brain center. Anatomically, the MCA artery is closer to the skull, away from the center of the circle. The circle of Willis does not show the same scanner motion dependence the outer vessels show. We test for possible confounds by using a method to introduce rotations. The experimental setup simulates +/- 1-degree rotations, with each case having a random rotation between these two values. If the experiment recreates some of the dictionary patterns found in the original data, it would signal that there would be some registration errors. The technique essentially regresses out some of the registration errors, leading to those features that are informative.

To control for these effects, we will use SVM analysis for the dictionary patterns found. There are significant patterns shown in the right MCA and the left PCA. To implement, we will use methods similar to those used in a paper to select voxels significant in aging using a sparse representation [29]. The group implements sparse representation, using a linear program, to select voxels important in aging. We will use similar logic except that we will use the dictionary voxels found through permutation testing. The group [29] found that using 20000 voxels, they were able to construct an SVM classifier of higher accuracy than using less significant voxels. We evaluate similar techniques to build a classifier and try to interpret the machine learning models.

To measure the effect of different voxel quantities on classification, we take the right MCA region, then see what the classification accuracy is when considering different sections. We find classification results similar to those using the whole image with just using a small subsection. We start by comparing the left and right sides and show the classification results. This reasoning is similar to that of another paper that tries to control for confounds in brain imaging [27]. We start with a small crop of the original data then expand it to one brain hemisphere, comparing ten different crops in total. In the next section, we repeat the same analysis but try to correct for small registration confounds by only considering

cases that are well registered. However, there is a high chance for positive bias, meaning higher than expected classification results, after confound correction due to the elimination of challenging edge cases. On the other hand, eliminating the confound could decrease the classification accuracy. Lastly, another technique uses cross-validation to find regression parameters, leading to efficient feature extraction and minimal bias.

We can go further than comparing regions of interest to isolating small ROIs to see if we could achieve a statistically significant classification. While we were able to show statistically significant classification using a relatively high amount of voxels, such as that shown in Figure 3.16 we wanted to pinpoint an area of high accuracy. Small regions that the dictionary learning indicates were significant did not always show statistical strength, and these small regions were only a few hundred voxels as compared to a few thousand voxels we considered when making Figure 3.16.

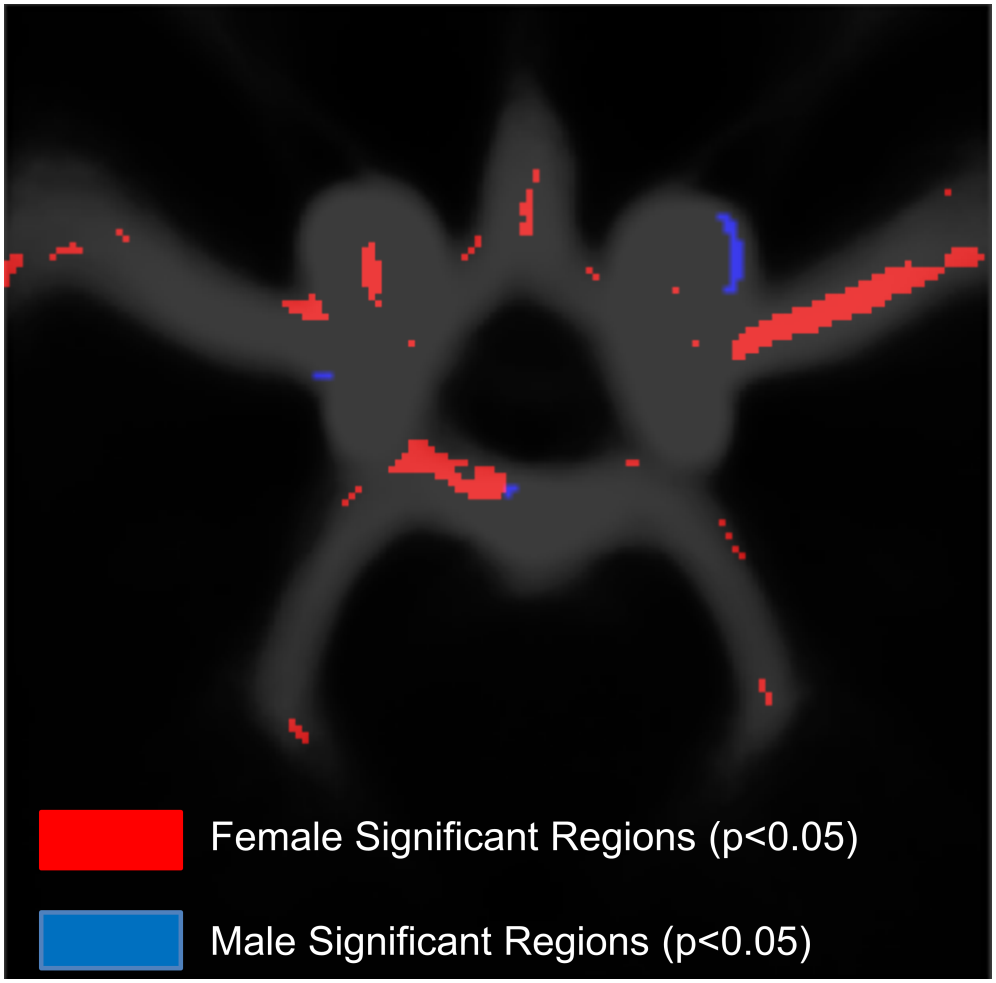

Figure 3.15: Dictionary Learning significance Map. The Red represents regions where a vessel is statistically significant more likely to be found in a female case than in a male case. The Blue color represents the opposite case. The white is a background template, which is an averaging of all the cases found. We can see that there are more statistically significant regions found in the female cases than in the male cases. Compared to the other figures where we show dictionary images, some of the bright regions may not be statistically significant, but it is hard to tell sometimes since the dictionary atoms themselves have some error associated with them.

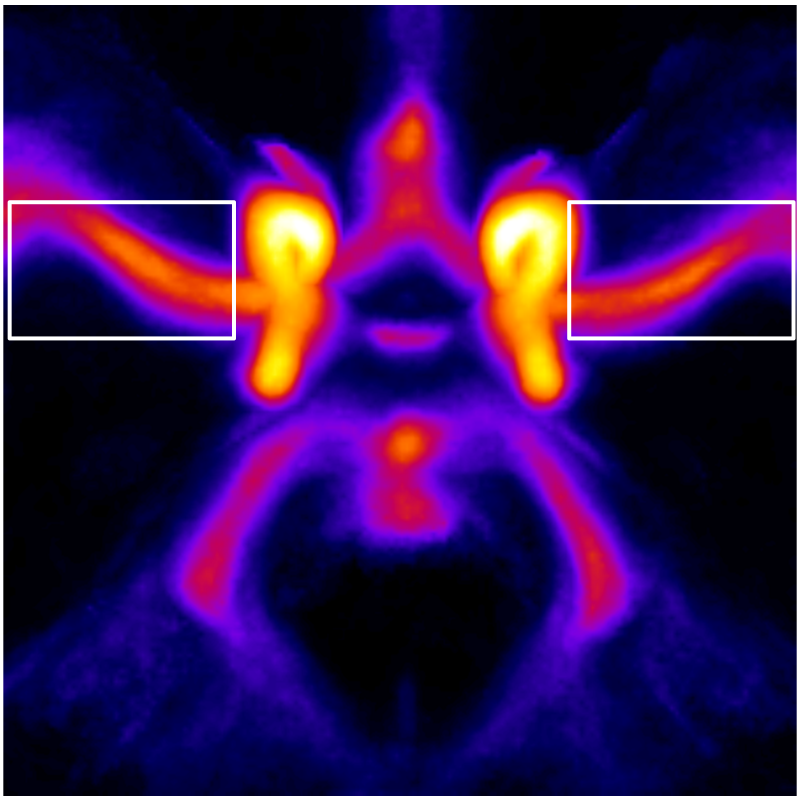

Figure 3.16: Testing crops in order to determine statistically significant classification. We take these two regions as shown in the figure and we wanted to see if they yield classification results similar to those found when comparing the left and right hand hemispheres. We see that when using the cropped regions as shown in the white regions of interests, the right hand side had statsically significant classification, meaning that we were able to train linear SVM programs in order to classify between male and female subjects. The left hand side did not however, and this also corresponds to the dictionary learning results that we also showed in this work.

CHAPTER 4

Discussion

In this section, we compare our findings and techniques in two ways. We first discuss comparisons between our anatomical patterns with prior literature on Circle of Willis anatomy. The second comparison will focus on two main machine learning works. One work uses an SVM kernel to classify males and females via a graph-based approach [9]. The second work is a graph-based deep learning network that tries to visualize salient regions used in deep learning-based classification. We also review our image processing pipeline and machine learning algorithms and propose future developments or enhancements.

4.1 Anatomical Comparisons

In general, as mentioned and shown in the results section, there are differences between the left and the right hemispheres. In the original atlas in the axial projection, we can see that the right and left hand MCA arteries have a different probability as the left is fainter than the right. The difference shows in dictionary imaging as the MCA between the male and female patients is the most significant difference between the two populations. We also have tested the significance of the left and right-hand sides in terms of classification and noted that if we only included the right-hand hemisphere, we would have a better classifier between the two genders than if we only used the left-hand side. So three different data points show a difference between the right and left-hand sides of the circle of Willis. A statistical atlas of both populations serves as a general starting point for visualizing differences. The dictionary learning provides more details of the same population. These patterns can serve as a platform for comparing other salient maps, such as deep learning

saliency maps. While it has not been discussed in this work before, there are techniques such as Grad CAM [26] which allow the end-users to see which parts of the image are more important from a classification point of view. A possible experiment is to study the salient features of deep learning models and seeing possible correlations between salient features and model performance. In particular, if we were to see similar salient features like the ones the dictionary learning pipeline finds, then it could help interpret the deep learning model. Intuitively, we would expect deep learning models to be responsive to salient features that are easy to observe with other algorithms. However, images are very high dimensional; the deep learning representation may instead pick lower-level biases that may not correspond well to actual semantic differences. There are studies [3] explaining differences in deep learning performance depending on the specific dataset used to train the model. The paper [3], compared brain tumor segmentation algorithms trained on different institutions. The paper evaluated a model trained on one institution on a separate institution, and the dice coefficient decreased from an average of roughly 0.7 to 0.6. Given these results, it would be interesting to develop a method of determining machine learning validity. The hypothesis is that robust deep learning models should find sound saliency maps that correspond to saliency maps constructed using basic techniques such as dictionary learning detailed in this work. The Grad CAM project has emphasized visualizing saliency maps to help end-users trust and understand deep learning models [26].

4.1.1 Circle of Willis Completion

To compare our results, we will discuss literature on gender anatomical differences in the circle of Willis. Among the interesting findings is the rate of the Circle of Willis completion. Some papers [34] mention that the Circle of Willis completion rate is higher in females than in males, which could explain some of our results. The same paper [34] showing evidence of the discrepancy of the Circle of Willis completion also speculated that the completion rate could have clinical implications for stroke outcomes. Our results support the papers [34] conclusion, key evidence of which comes from considering the Dice Coefficient of the genders,

with females on average having higher dice coefficients. The Dice Coefficient is a higher value when there is a complete circle. It is hard to distinguish image processing artifacts from subtle differences in population anatomy without expert anatomical knowledge. When reading some papers referenced in this work, case selection justification was often qualitative in judgment. The paper [22], on statistical atlas generation, mentioned that they would not consider cases that were not well segmented but only gave qualitative reasoning. For large datasets, manual processing and selection are impractical. Although the case selection did not have input from an expert in segmentation and medical imaging processing, we believed it to have reached plausible post-processing, giving some results comparable to previous anatomical studies. Although the most challenging part of the work was image processing validation, we have tried quantitative justification given prior work. Also, another benefit to our work would be the automatic detection of differences between populations in a systematic way since many of the methods involve time-intensive manual analysis [34].

4.1.2 ICA Size Comparison

Another finding of this work that has some experimental support is the comparison between the ICA between males and females. A paper [18] mentioned that the ICA diameter of males is larger than that of females. The ICA is the largest artery in the circle of Willis, found in the bottom of the neck. Other variables could account for this difference, such as neck circumference, height, age, and weight. The paper found that these differences still were found after controlling for these confounding variables. The ICA diameter was 4.66 ± 0.78 mm, in females, and 5.11 ± 0.87 mm in males. From the Male-Female average dictionary atom, as shown in Figure 3.9, the ICA arteries are the most prominent, showing that they are the most significant arteries found more common in that position rather than females. The white background is the same for both the male and the female cases, so the white background provides a good comparison between males and females. With more rigorous validation of the dictionary learning technique, specifically in characterizing the quantitative vessel differences, the dictionary learning technique could help in automatically identifying

clinically relevant regions. The main application of these studies [18] is to provide gender-based considerations for stroke surgeries. The surgeries may be harder to perform with smaller arteries; both the paper mentioned and this work suggests gender-based differences that could affect surgery and possible stroke prognosis.

4.1.3 Axial vs Coronal Viewpoints

While our results are more reliable in the axial viewpoint, the outcome could be due to the axial axis being the acquisition axis. However, other axes can be suboptimal. When building the coronal axis atoms, we used the same cases as the axial viewpoint for consistency. While dictionary learning can create patterns, these patterns may not indicate actual structural differences; but rather variations in the registration processing. We saw some structure, particularly when considering the axial viewpoint and the left and right-hand sides. The MCA was the most different between males and females, and this difference also shows in the statistical atlas itself. While the statistical atlas is a population-level measure, any asymmetries in the atlas can signal asymmetries between groups indicating an area where population differences are likely. Figure 4.1 shows the statistical atlas for the axial viewpoint; this figure shows a difference in the MCA on the left and right-hand sides. We will display the anatomical convention viewpoint so that it is easier to see vessel locations according to a canonical textbook definition. In particular, looking at the MCA in the statistical atlas and comparing it to the dictionary pattern found earlier in the results section can provide additional support for the patterns found. The dictionary learning pipeline described is essentially a population difference clustering map, so asymmetries seen in an atlas should be found in the dictionary atoms, corresponding to our findings.

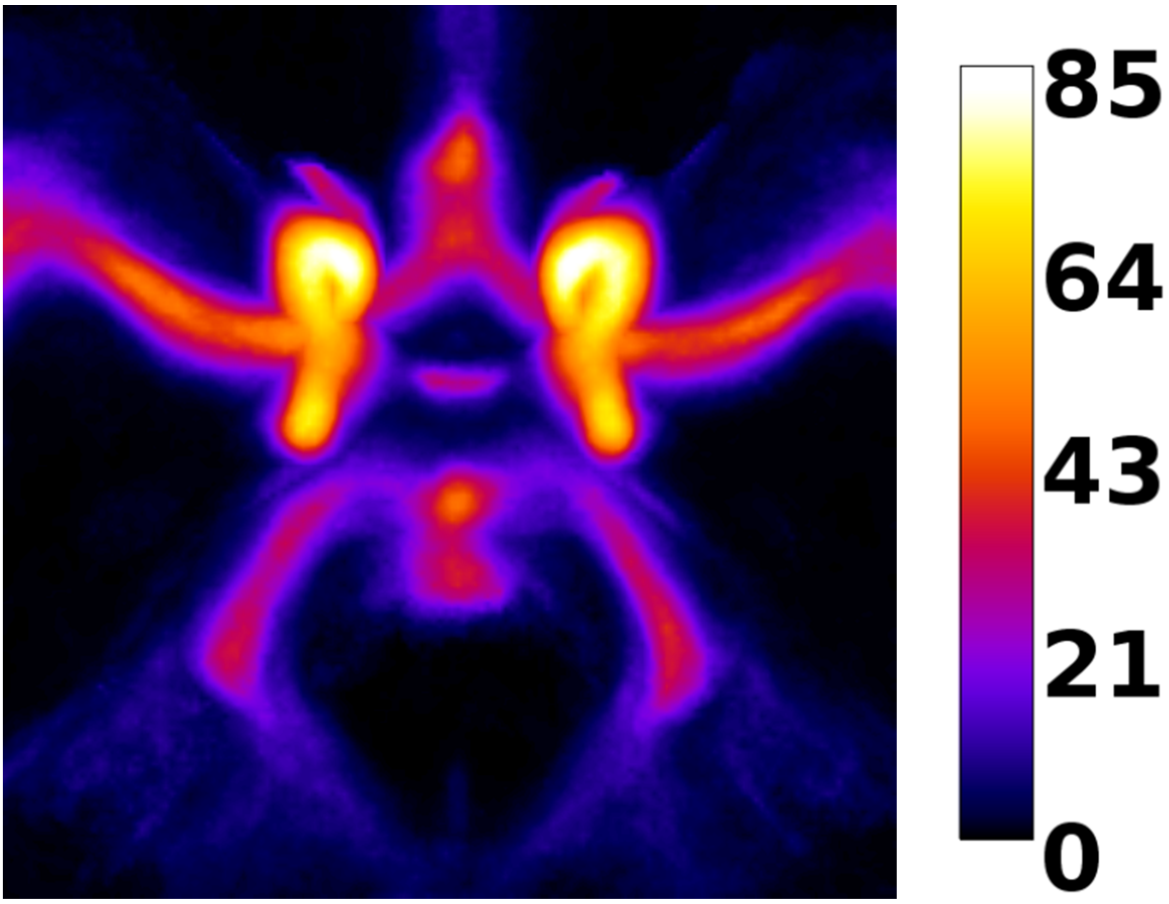

Figure 4.1: Probability Atlas of the axial projection showing the anatomy that has the highest probability to be found. Since the vessels are of various sizes and shapes, we have shown statistical atlas to showcase the most variable vessels. The graph scales to probability, with 100 corresponds to a vessel probability of 1.0, and 0 corresponds to probability 0. From the image, we can see that the ICA, the central arteries, are found with high probability, while the peripheral arteries have less probability at around 0.2 to 0.3. The highest probability is 0.85, while the blue regions are below 0.2. Red is about 0.4

4.2 Dictionary Learning Saliency Maps vs SVM saliency maps

Some relatively recent work has criticized the use of percentage based classifications in terms of buliding a saliency map of the SVM regions. [5] However, our method builds on this by using an additional saliency method using our dictionary learning to learn how the classifier works. THe main lesson from this work is that there are some registration confounds due to systematic biases of male and female brains. The differences in registration and movement artificats are the main source of this confound and has been most studied in structural MRI. While we do not implement the method used in [5], we provide an additiona support for the SVM saliency map, so that we have a more complete understanding of how the MVPA algorithms work. This can be used as a stepping stone for more complex classifiers including deep learning models.

4.3 Discussion of Alternative Registration, Segmentation and Dictionary Learning Algorithms

4.3.1 Registration Algorithm Alternatives

In this work, we evaluate the registration results by considering the dice coefficients and the Radon transform. There are other registration validation techniques, among them being the concept of mutual information. The essence is that in well-registered images, the images should have corresponding histograms in the same spatial regions. Thus, this can help quantify registration quality without relying on segmentation, as the dice coefficient as defined only works on binary data. Additionally, the registration algorithm needs improvement. In our work, we use image crops of 150x150x61, since registering the entire image would take 20 hours per CPU. Alternatives to CPU-based algorithms include GPU implementations of the CPU algorithms and deep-learning variations. These deep learning-enabled registrations run at constant time and natively run on GPU, offering a faster alternative allowing us to perform registration of potentially hundreds of cases within a few days instead of a

few weeks as would be the case with the ANTS registration implementation we have. A potential algorithm is the Quicksilver algorithm [32]. The deep learning algorithm works on the LDDMM principle of registration described earlier. It tries to predict the registered images by predicting the transformation parameters using the deep learning network. So instead of performing the gradient descent step, it infers the optimal parameters using the deep learning network. However, in terms of registration performance, these deep learning networks may fall short of the state-of-the-art optimized registration algorithms [13]. The group trained the deep learning network on various image patches of 15x15x15 in size using a sliding window to capture most of the distinct subregions in the volume. The patchwork is needed since the deep learning networks need lots of training data and image patches offer more data than trying to infer using whole images. Another reason this patchwork strategy works is due to medical image redundancies; at a finer scale, it is possible to integrate patterns in a bottom-up fashion. A downside to the bottom-up integration is boundary effects on the global image representation. So there is another part of the network that works on the inference of the first part, then has a fine-tuned prediction using the prediction of the first network. For this study, we would have to retrain the network using MRA images; since the original network used MRI images for training.

4.3.2 Segmentation Algorithm and Possible Alternatives

When implementing machine learning, the first step is to normalize the data. While we have presented what we have done with our preprocessing pipeline, we have had some problems interpreting the segmentation results. Most segmentations performed in medical imaging are deterministic; the structure is either present or absent. However, some low contrast vessels are still tricky to threshold using an automatic algorithm, and even the best may require some fine-tuning. Despite the difficulty in segmentation, it is necessary to correct the biases in imaging across different setups and patients. To generate the dictionary atoms, we used a normalization algorithm to correct for grayscale imaging variations, but did not use segmentations since the atoms were noisy. An intermediate representation is one where

the segmentation result is a probability, a continuous value, rather than a binary value. The dictionary atoms produced using normalized grayscale images are smooth, but other normalization techniques can help. The main theoretical component is the expectation-maximization concept. The concept is described in the preprint [30], where the work details a probabilistic model to segmentation. While the paper aims to create general probabilistic models to segment any image, this concept applies most strongly to the life sciences. In general, biological information has noise and uncertainty; the brain images that we study in this work are composed of blood vessels that are small and of complex geometries. There are also imaging artifacts and resolution limits that can make it hard to define the segmentation boundaries. The main mathematical principle used in this probabilistic treatment is to treat the pixel label as a random variable, then an expectation-maximization algorithm is implemented to solve the probability distribution. Intuitively, the integrated probability distribution values would result in the segmented image. We would take the average of these segmented results so that ambiguous segments are assigned some grayscale value, while areas that are almost certainly vessels would be assigned greater numbers.

4.3.3 A generalization of K-SVD algorithm to train dictionaries

In our work, we used the original algorithm of K-SVD, which is a generalization of the k-means algorithm. While the algorithm is fast and converges, there are some cases where it could fail, especially when there is an ambigious partition between two groups. Thus many groups have tried to derive better k-means algorithms, one of which [25], shows promise and is relatively fast but is based on expectation-maximization optimization. The authors call their algorithm the maximum a posteriori Dirichlet process since the optimization attempts to find the most probable probability density functions that fit the data. A Dirichlet process is a probability distribution model that models the probability of the data fitting a certain probability distribution. The clusters are probability distributions; the probability of data points assigned to any one of these clusters is also a probability distribution. The theoretical framework to build clusters using this methodology is known as the Chinese restaurant

process (CRP), in which the analogy is to a restaurant where the data points are customers; and the clusters are tables. The customers coming in parameterize the number of tables used, in contrast to having a fixed number of tables. This probability distribution of data points to clusters is as follows:

$$p(customer\ i+1\ joins\ table\ k) = \begin{cases} \frac{N_k}{N_o+i}\ if\ k\ is\ existing\ table\ (cluster) \\ \frac{N_o}{N_o+i}\ if\ k\ is\ new\ table\ (cluster) \end{cases} \tag{4.1}$$

this corresponds to draws z of the CRP distribution, as defined below:

$$(z_1, ..., z_N) \tilde{} CRP(N_o, N) \tag{4.2}$$

The probability of any given distribution using the CRP distribution is the result of multiplying each of the cases.

$$p(z_1, ..., z_N) = \frac{N_o^k}{N_o^{(N)}} \prod_{k=1}^{k} (N_k - 1)! \tag{4.3}$$

The N_o comes from considering each table added (or cluster) and is thought of as a density parameter. The factorial expression is the mathematical expression for considering all the individual customers.

The above process can be derived from the Dirichlet process [25], and the above formulation of the CRP will be the foundation for deriving the modified k means problem. We will now present the generalized form of the modified k means. One additional consideration for the final formulation is considering the predictive distribution, with x concerns the data probability given the hyperparameters θ. The probability model is the CRP described above and the likelihood distribution:

$$x \quad \tilde{} f(\theta_{z_i}) \tag{4.4}$$

combining the expression above with the CRP and multiplying gives us the likelihood expression below:

$$p(z_1, ..., z_N) = \prod_{i=1}^{N} \prod_{k=1}^{K} f(x_i|\theta_k^{-i})^{\delta(z_i,k)} \tag{4.5}$$

Next, we take the negative log likelihood of the above equation:

$$E = -\sum_{i=1}^{N} \sum_{i:z_i=k} \ln f(x_i|\theta_k^{-i}) - K \ln N_o - \sum_{k=1}^{K} \ln \Gamma(N_k) - C(N_0, N) \tag{4.6}$$

where $C(N_o, N) = \ln \frac{\Gamma(N_o)}{\Gamma(N_o+1)}$. Equation 4.6 is the equation we want to minimize and is the equivalent to the minimization process in the normal K-means equation. The analogy to our case would be taking the difference images as stated earlier, then taking each of these difference images, and clustering them. The clusters would then be determined using the minimization criteria, so there could be a variation in the cluster number. This cluster number may not necessarily be the sparest representation, as was theoretically the case using K-SVD, but is another method to determining possible clusters. To interpret the output of Equation 4.6, we would then average the clusters to arrive at a global saliency map.

4.4 Deep Learning Comparisons and Machine Learning Comparisons

The work closest to this work is a paper on using graph-based kernels to classify males and females based on the whole brain segmented MRA vasculature [20]. In that work, the group only considered 43 cases out of a dataset of 109. They also relied on expert segmentations to achieve a non-trivial SVM classifier. The next step to that paper was to integrate their graph-based algorithm along with dictionary learning. There is interest in dictionary learning and morphometry, but the field is relatively new. Among the problems is that it is hard to standardize the patterns gained. Another is that the process for getting these results from case selection to registration and segmentation depends on manual annotation and

case selection, making it hard to scale. A recent paper [22], on a statistical atlas of MRA images, was crucial to our registration pipeline as the first paper [20], relied on MRI brain imaging to serve as a template for registration. The work helped to streamline the processing and allow the analysis of hundreds of cases instead of dozens. We also sought to quantify case selection. Medical imaging papers rely on an expert to decide what cases are selected and what preprocessing is needed. Hence a part of this work was to have a workflow so that there are ways to have standardized analysis and case selection so that the results are easier to reproduce. With datasets containing thousands of patients, a protocol for case selection and criterion that is automated and streamlined is crucial.

As the last part, we will also consider a paper that could be relevant as the next step in this work. This work uses deep learning to find saliency maps in graphical networks using a graph convolutional network or GCN. [6]. The context of this experiment was in trying to find gender differences in fMRI, or what is known as functional MRI. This study used 5000 subjects from the UK Biobank database, which is among the largest brain databases [6]. It's essentially a modality of MRI that attempts to measure how blood flow changes through an experiment to show the relative brain activity during different tasks. The main idea of the work is to extend the principles of [26] to visualize gradients, using a graphical representation where the graph represents brain activity correlation. After building a graph representation, the network then finds features to classify between two different populations. The final layer before the classification is a summation layer known as global average pooling. An analogy to our project, the second to last layer is like our dictionary atoms, and the average pooling is similar to averaging all of the dictionary atoms. The global average pooling is below:

$$F_i = \frac{1}{d_z} \sum_v f_i(v) \tag{4.7}$$

Equation 4.1 represents the average pooling operation. The $f_i(v)$ represents the node activation and d_z the number of nodes. In this work, the $f_i(v)$ represents a node, but in our dictionary learning, it represents a vessel. The primary difference in this work is that

the euclidean representation, which is typical in deep learning and used for our dictionary learning implementation, is replaced with connectivity graphs as shown in Equation 4.1. The weight vector summation is analogous to summing the dictionary atoms to create a global saliency map for the differences between populations.

The main advantage to using the dictionary learning method is that it would give you salient features without tuning many hyperparameters, as the original work on pattern-based morphometry [14] found robust dictionary patterns regardless of the hyperparameters. Figure 4.2 shows a network graph. It also has a schematic of how these graphs and representations are combined to construct a classifier.

A possible follow-up to this work is to compare the adjacency graph representation of the vasculature using our dictionary learning pipeline and the deep learning approach. If both converged to similar saliency maps, it could imply a link between dictionary learning and deep learning. However, there is no guarantee that they would have to agree, so then the next step would then be to evaluate the same thing but on different datasets. If there is a correlation between salient features and generalizability, it would be a significant finding since we would have a way to evaluate the deep learning models and find ways to interpret them.

4.4.0.1 Applications and Relevance

While deep learning has had much success in the medical imaging field, it is opaque and relies on specific training data. Part of the motivation behind this work is to use sparsity-based methods to understand deep learning models. The platform described here uses multiple data sources and applies to different populations without extensive hyperparameter optimization. Also, the main hurdle in applying deep learning to medical data is controlling for experimental conditions. The main problem to solve is to find a model that is robust to different setups and conditions or to find a preprocessing pipeline so that the data is appropriately normalized. While the deep learning test would work well in medical imaging in one institution, there are no guarantees that it would work in another. [3] While it is hard

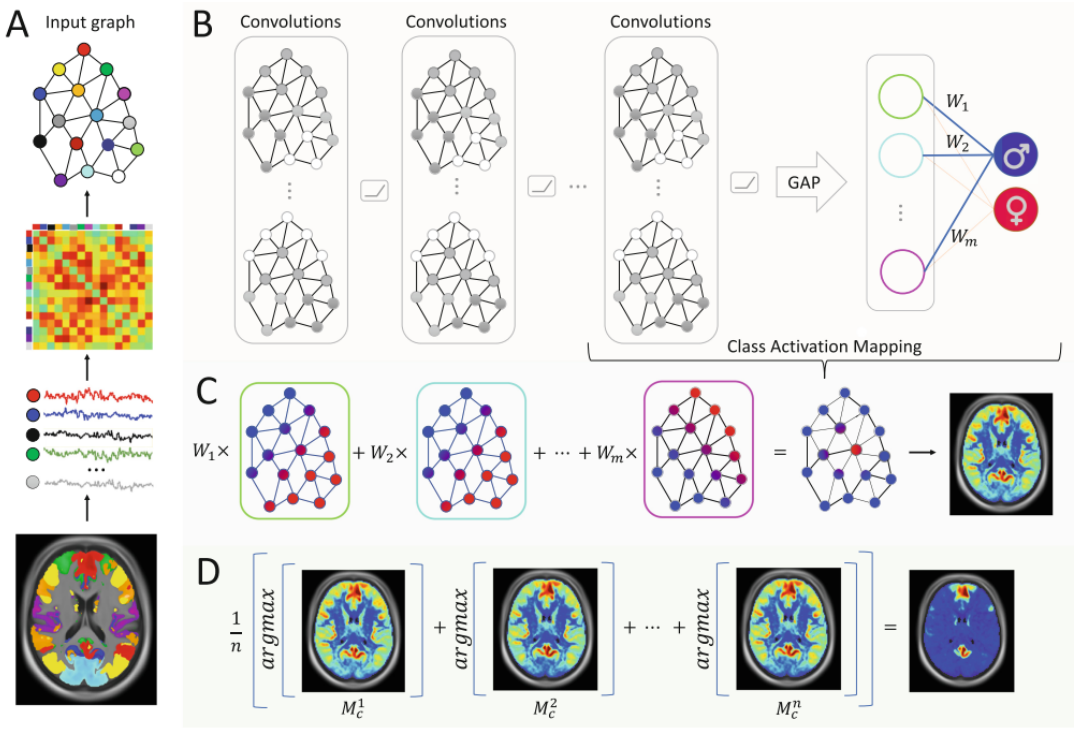

Figure 4.2: Network of the GCN classifier. The salient features found in the image figure are analogous to our dictionary atoms that show saliency differences. In particular, the last part of figure part D shows how taking a projection of that data yields a saliency map, of which the result and interpretation are similar to our method.

to know how deep learning works on a detailed level, knowing which features it relies on for its decisions; could help to diagnose some of these shortcomings.

When implementing these techniques into the medical setting, we also need to consider group differences in normal patient populations. As a simple example, on average, males have larger brains than females, and these differences might be of clinical significance. While clinicians may be well aware of these population differences when analyzing images, algorithms are not guaranteed to have these considerations in mind. Therefore, there should be methods to correct these confounds. There are various methods to do confound control, one of these [27], focuses on the problem between brain MRI size between male and female patients. The method incorporates SVM cross-validation to find features to control. In this work, we focus on possible registration confounds between males and females. While we did not prove that these registration confounds are robust across all data sets, one possible hypothesis difference is motion differences between populations. There is a paper [4] detailing differences in MRI volume segmentation between different groups. Translated to MRA imaging, there can be motion confounds that can translate to registration confounds. In the same paper, they mention that on average, motion confounds are more significant in males than in females. Motion confounds, if they exist in a statistically significant manner, are readily detected using machine learning. It is wise to control for these differences as an algorithm could find statistically significant results with high accuracy between groups, but if these differences are due to motion differences, the results have less clinical relevance.

4.4.1 Confound Regression using Cross Validated SVM

As mentioned before, there is a paper that claims to have a low bias approach to confound analysis. While the paper focused on gender size in brain MRI confound, our approach will use a similar approach to the paper, but with the regression out of motion confound. It can be difficult to find features that could be confounds, but our methodology with dictionary learning can show how to find these confounds. Once these confounds are found using our technique, we can apply methods similar to [27]. We give a brief overview of the methods that

[27] uses, and how these techniques could augment our arguments. Their novel technique is termed cross validated confound regression (CVCR). To go over the CVCR method, we first go over the method of whole dataset comfound regression, the key equation is shown below. This method tries to do a global confound regression and tries to regress the whole dataset using a parameterized version of the variable to regress out. In the context of their experiment, their confound is brain size, in our context, our matrix will be a transformation regressing out small registration error.

$$\beta_c = (C^T C)^{-1} C^T X \tag{4.8}$$

Here, X is the dataset, C is an Nx2 matrix, with one column having a bias term, and the other has the confound. B is a Kx2 array. To represent taking out the variance due to confound we introduce the equation included below:

$$X_{corr} = X - C\beta_c \tag{4.9}$$

The X_{corr} is thus the data after subtracting for the confound. While this method is straightforward, it does have a high risk of negative bias since it does reduce correlations, with the estimation of the confound based on the entire dataset, it is an imprecise method since there could be considerable variation within the dataset and cases. An alternative is to estimate the confound within a given fold of training data and using those parameters to regress the confounds out. The equation logic is shown below in terms of training the parameters.

$$\beta_{c,train} = (C_{train}^T C_{train})^{-1} C_{train}^T X_{train} \tag{4.10}$$

To regress out the parameters, it is similar to the whole dataset confound regression method except that we use the trained parameters $\beta_{c,train}$ to regress out the parameters in the training set:

$$X_{train,corr} = X_{train} - C_{train}\beta_{c,train} \tag{4.11}$$

and the test set:

$$X_{test,corr} = X_{test} - C_{test}\beta_{c,train} \tag{4.12}$$

The paper authors have implemented the methods using the python scikit module. Out of the methods mentioned, such as having a smaller dataset without confounds and regular confound regression, the cross validated confound regression yields minimal bias. In future work, we plan to implement similar confound control with motion and registration artifacts. With similar refinement of confounds, that we can see using our dictionary learning, the end goal is to find subtle group differences that could have clinical relevance, and ones unrelated to experimental conditions and data taking. While high accuracy classifiers are now commonplace due to deep learning, we argue that visualization of group differences and confound analysis are of equal importance, and perhaps detailed confound analysis can help in increasing deep learning robustness.

CHAPTER 5

Conclusion

In this work, we propose a pipeline for MRA population vasculature analysis. Our technique can complement machine learning algorithms, such as SVMs that are powerful but opaque discriminators. However, the main limitation in this approach is that the values in which we can make inferences are not quantitative, as they would be for a classifier. The plan for this dictionary learning model is to explain opaque but high-performance classifiers.

. Future directions would include trying to train a deep learning network naively on the data as presented, afterward applying the graphical representation described to compare the saliency graphs generated through this technique with our dictionary learning model. While we have found some patterns through dictionary learning, it would be reassuring if deep learning found structures similar to classical methods. However, it might be the case that we find that deep learning based features more general and more useful than those found using classical methods.

BIBLIOGRAPHY

[1] Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. *NeuroImage*, 78:270–283, 2013.

[2] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[3] Ehab A. AlBadawy, Ashirbani Saha, and Maciej A. Mazurowski. Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. *Medical Physics*, 45(3):1150–1158.

[4] Aaron Alexander-Bloch, Liv Clasen, Michael Stockman, Lisa Ronan, Francois Lalonde, Jay Giedd, and Armin Raznahan. Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo mri. *Human brain mapping*, 37(7):2385–2397, 2016.

[5] Carsten Allefeld, Kai Görgen, and John-Dylan Haynes. Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *Neuroimage*, 141:378–392, 2016.

[6] Salim Arslan, Sofia Ira Ktena, Ben Glocker, and Daniel Rueckert. Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity. In Danail Stoyanov, Zeike Taylor, Enzo Ferrante, Adrian V. Dalca, Anne Martel, Lena Maier-Hein, Sarah Parisot, Aristeidis Sotiras, Bartlomiej Papiez, Mert R. Sabuncu, and Li Shen, editors, *Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities*, pages 3–13, Cham, 2018. Springer International Publishing.

[7] J. Ashburner and K.J. Friston. Voxel based morphometry. In Larry R. Squire, editor, *Encyclopedia of Neuroscience*, pages 471–477. Academic Press, Oxford, 2009.

[8] B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008. Special Issue on The Third International Workshop on Biomedical Image Registration – WBIR 2006.

[9] Stephen R. Aylward, Julien Jomier, Christelle Vivert, Vincent LeDigarcher, and Elizabeth Bullitt. Spatial graphs for intra-cranial vascular network characterization, generation, and discrimination. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005, pages 59–66. Springer Berlin Heidelberg.

[10] Fred L. Bookstein. "voxel-based morphometry" should not be used with imperfectly registered images. *NeuroImage*, 14(6):1454–1462, 2001.

[11] S. Chelbi and A. Mekhmoukh. Features based image registration using cross correlation and radon transform. *Alexandria Engineering Journal*, 57(4):2313–2318, 2018.

[12] IXI Dataset, 2021.

[13] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. Deep learning in medical image registration: a review. *Physics in Medicine & Biology*, 65(20):20TR01, oct 2020.

[14] Bilwaj Gaonkar, Kilian Pohl, and Christos Davatzikos. Pattern based morphometry. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 14(Pt 2):459–466, 2011.

[15] Martin N Hebart and Chris I Baker. Deconstructing multivariate decoding for the study of brain function. *Neuroimage*, 180:4–18, 2018.

[16] Tim Jerman, Franjo Pernuš, Boštjan Likar, and Žiga Špiclin. Enhancement of vascular structures in 3d and 2d angiographic images. *IEEE Transactions on Medical Imaging*, 35(9):2107–2118, 2016.

[17] Paul Johns. Chapter 10 - stroke. In Paul Johns, editor, *Clinical Neuroscience*, pages 115–128. Churchill Livingstone, 2014.

[18] Jaroslaw Krejza, Michal Arkuszewski, Scott E Kasner, John Weigele, Andrzej Ustymowicz, Robert W Hurst, Brett L Cucchiara, and Steven R Messe. Carotid artery diameter in men and women and the relation to body and neck size. *Stroke*, 37(4):1103–1105, 2006.

[19] R.K. Kwitt. Python pattern based morphometry. https://github.com/KitwareMedical/TubeTK-pypbm, 2013.

[20] Roland Kwitt, Danielle Pace, Marc Niethammer, and Stephen Aylward. Studying cerebral vasculature using structure proximity and graph kernels. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16(Pt 2):534–541, 2013.

[21] MATLAB. *version 9.5.0.1067069 (R2018b) Update 4*. The MathWorks Inc., Natick, Massachusetts, 2018.

[22] Pauline Mouches and Nils D. Forkert. A statistical atlas of cerebral arteries generated using multi-center mra datasets from healthy subjects. *Scientific Data*, 6(1):29, 2019.

[23] Rodica E. Petrea, Alexa S. Beiser, Sudha Seshadri, Margaret Kelly-Hayes, Carlos S. Kase, and Philip A. Wolf. Gender differences in stroke incidence and poststroke disability in the framingham heart study. *Stroke*, 40(4):1032–1037, 2009.

[24] Robert A. Pooley. Fundamental physics of mr imaging. *RadioGraphics*, 25(4):1087–1099, 2005. PMID: 16009826.

[25] Yordan P Raykov, Alexis Boukouvalas, Fahd Baig, and Max A Little. What to do when k-means clustering fails: a simple yet principled alternative algorithm. *PloS one*, 11(9):e0162259, 2016.

[26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[27] Lukas Snoek, Steven Miletić, and H Steven Scholte. How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage*, 184:741–760, 2019.

[28] Blausen.com staff (29 August 2014). "medical gallery of blausen medical 2014". *Wiki-Journal of Medicine, 1 (2).*

[29] Longfei Su, Lubin Wang, Fanglin Chen, Hui Shen, Baojuan Li, and Dewen Hu. Sparse representation of brain aging: extracting covariance patterns from structural mri. *PloS one*, 7(5):e36147, 2012.

[30] Jonathan Vacher, Pascal Mamassian, and Ruben Coen-Cagli. Probabilistic model of visual segmentation. *arXiv preprint arXiv:1806.00111*, 2019.

[31] Jennifer L. Whitwell. Voxel-based morphometry: An automated technique for assessing structural changes in the brain. *Journal of Neuroscience*, 29(31):9661–9664, 2009.

[32] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration – a deep learning approach. *NeuroImage*, 158:378–396, 2017.

[33] Xin Yang, Kwang-Ting Tim Cheng, and Aichi Chien. Accurate vessel segmentation with progressive contrast enhancement and canny refinement. In Daniel Cremers, Ian Reid, Hideo Saito, and Ming-Hsuan Yang, editors, *Computer Vision – ACCV 2014*, pages 1–16, Cham, 2015. Springer International Publishing.

[34] Orel A. Zaninovich, Wyatt L. Ramey, Christina M. Walter, and Travis M. Dumont. Completion of the circle of willis varies by gender, age, and indication for computed tomography angiography. *World Neurosurgery*, 106:953–963, 2017.