Segmentation of Multiple Sclerosis Lesions using Dictionary Learning in Feature Space

Gregory C. Kuling

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTERS OF ARTS

GRADUATE PROGRAM IN DEPARTMENT OF

MATHEMATICS AND STATISTICS

YORK UNIVERSITY

TORONTO, ONTARIO

AUGUST 2017

Abstract

Manual segmentation is used in the diagnosis, management and evaluation of clinical trials for Multiple Sclerosis (MS), but human error makes manual segmentation variable.

Automatic segmentation has been proposed using a Machine Learning algorithm Dictionary Learning (DL). We explored using different feature spaces to automatically segment MS lesions from healthy brain tissue. Methods of image texture analysis quantify the spatial distribution of the voxels in multi-weighted MR scans. We present the results of using a single voxel, single voxel and standard deviation (σ) of adjacent voxels and a large spatial patch as feature spaces.

The single voxel method segments the MS lesions with a Dice Similarity Coefficient (DSC) of 0.985 on simulated Brainweb data, but performed poorly with noise in the image (0.654). The single voxel and σ performs at a DSC of 0.943 in the presence of 3% noise. The method should be attempted on real patient data.

To those who believed in me when I didn't.

Hongmei Zhu

Verna Kuling

Christopher Kuling

& my partner, Glen Gamble

Acknowledgements

I would like to thank and acknowledge my patient and supportive supervisor Hongmei Zhu, for giving me such a compelling research project and expertly guiding me to hone my mathematical skills. I also would like to thank my committee for pushing me to understand every detail of this project; Dr. Yunyan Zhang, and Hélène Massam. I would not have been successful in this project without their support and patience.

I would like to thank the friends and family that helped me edit: Alex Nadeau, Victor Luong, Emily Redelmeier, Mahnaz Alavinejad, Glen Gamble, Richard Le, and Eryn Frawley. Thank you to York university's mathematics department computer technician, Steven Chen. As well as the department's graduate studies team Primrose Miranda and Mike Zabroki for assistance and guidance.

Finally, I would like to thank the resilient and inspirational women in my life who are afflicted with Multiple Sclerosis; Jennifer Loughery, Michelle Snider, Kara Dawn, and Kendra MacLaren.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Segmentation is the process of dividing information into segments. For image processing, segmentation would be the act of drawing out a region of interest by selecting a collection of pixels that hold some specific characteristic. Once we have a collection of regions, labeling and classifying them can give more information about what is in the image. This assists in the comprehension and analysis of the image.

Conventionally, image segmentation has been done manually, by a human operator. For the operators, this partitioning and labeling is sufficient for their understanding, but it is not always apparent to others. Decisions on manual segmentation suffer from inter-observer variability where two people do not agree on the placement of segmentation in the same image. Over time an operator's perception of the image can change as well. The operator can decide that the segmentation they had once drawn before is incorrect and find faults in their previous work. This is described as intra-observer discrepancies. Therefore, manual segmentation can be flawed [17].

Automatic segmentation has been suggested to be a solution to the flaws of manual segmentation. Automatic segmentation tries to remove the operator variability by using mathematical methods to segment the regions of interest in a photo. These techniques analyze pixel data in the image by comparing pixel values and evaluating

the arrangements. Automatic segmentation has been very beneficial in tasks such as machine vision, object detection and recognition [28]. There is no better example for the use of automatic segmentation than finding disease in medical images [30]. So far, controversies have arisen over the optimal technique for the task and the lack of accountability for performance errors. This has resulted in an ongoing research topic to find the most accurate means possible for automatic medical imaging segmentation.

This topic of automatic medical image segmentation could not be more significant than it is in Multiple Sclerosis (MS) [18]. The autoimmune disease is character-ized by the immune cells attacking healthy neurons in the central nervous system. Healthy neurons have a fatty coating along the axonal portion of the cell membrane called myelin. This myelin sheath has self-attached antigens to identify itself. With MS, these antigens are considered an enemy to the body and the immune system is deployed to eradicate the myelin sheath.
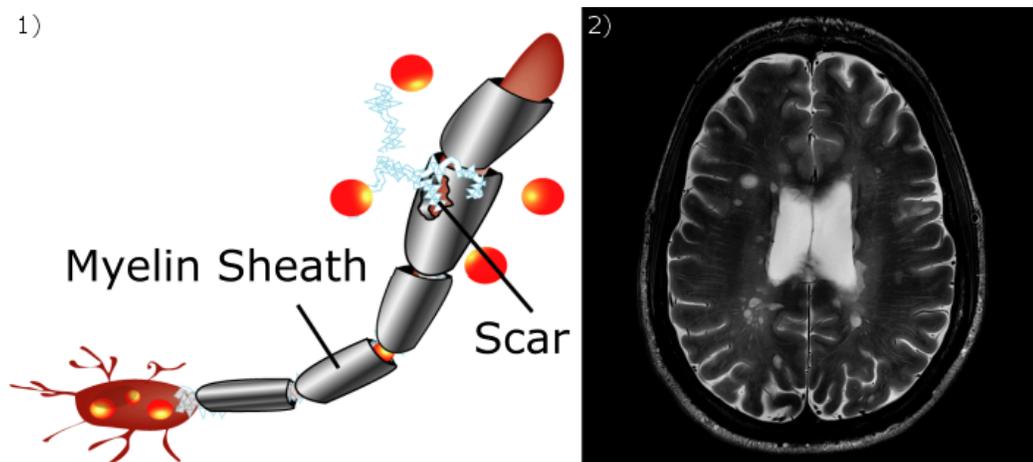
Figure 1.1: 1) diagram of a neuron being attacked by the immune system. [35] 2) a 7 Tesla T2 Magnetic Resonance image of MS lesions. [1]

Once a neuron has been attacked, inflammation increases and can result in cell death. This cell death leads to scar tissue being left behind called a lesion. It is these sclerotic lesions in the central nervous system that gave MS its name. These abnormal changes in brain tissue are characteristic to MS and the cause of neurological

symptoms in the patient. These symptoms range from motor function difficulty to cognitive delays. [11] Therefore, lesion localization is important in the understanding of patient symptoms.

As a patient's disease progresses, these lesions will develop in many parts of their brain. Some will recover, while others will relapse. The disease progression and patient status can be described using the calculated volume of lesion load in their brain [27]. By monitoring its progression, physicians can evaluate the severity of the patient's disease. Many different treatment regimens have been developed for decreasing relapses and development of lesions. Therefore, lesion load is important in the evaluation of these interventions. [16]

It has become routine to measure MS lesion that appear in a Magnetic Resonance (MR) scan of the patient's brain as an indicator to disease severity [16] [27]. MR imaging is extremely efficient at visualizing soft tissue distribution within the brain. This is because of its signal quality rendered from atomic distribution as opposed to radiation attenuation used in computed tomography. MR scanners orchestrate a sequence of magnetic field pulses to receive a signal from the atoms within the sample.

With MR imaging, the soft tissue quality is determined by the magnetic field pulse sequence. In the sequence, magnetic fields in different directions are turned on to alter the nuclear magnetic vector of the localized sample. Then the signal of the sample is measured after a given set of time. What results is measuring different relaxation times reflective of the Bloch equations [7]. These relaxation times are different for different tissues in the human body.

These sequences, notably for MS, are T1 weighted (T1w), T2 weighted (T2w), Proton Density weighted (PDw), and Fluid attenuation inversion recovery weighted (FLAIR). In T1w images, MS lesions appear as hypo-intensive regions; a collection of dark pixel intensities. However, the latter three result in hyper-intensive regions; pixels closer to white. Healthy tissue has a different coordination of intensities de-

pending on the weighting. For example, cerebral spinal fluid (CSF) is dark in T1w and FLAIR images while bright in T2w and PDw. The contrast difference depends on the tissue [18]. The combination of weighted images having characteristic lesion traits is essential to cross-validate the lesion position. Therefore, this pattern is useful in delineating healthy tissue from MS lesions.

There is a distinct pattern of image intensity variation between MR image weightings and physicians have learned to use this to their advantage in manual MS lesion segmentation. As stated before, manual segmentation is flawed, so the goal is to teach a computer to recognize the same pattern. Machine learning (ML) has been used for pattern recognition in many different tasks [22]. To automatically segment the lesion from the healthy tissue in an MR image, ML is used to recognize the patterns of healthy and lesion tissue. ML algorithms attempt to learn patterns from large sets of data, which is suitable for our goal. As an example, from the data set that is used in this study, the MR scan has 7,109,137 voxels of information for one scan weighting. Voxels are a three-dimensional concept of a pixel, holding an intensity level in a volumetric space. Therefore, ML is an excellent mathematical tool to understand the patterns of the MR scan.

In the past, mathematicians have attempted two core types of ML techniques; supervised and unsupervised [23]. Supervised ML incorporates a principle that what has been done in the past is applicable to the future and is characterized by a large database to compare the test subjects to. By learning feature patterns in previous examples of a task, the model can efficiently predict future tasks. Unsupervised ML attempts to understand and manipulate the data presently, and not consider previous examples of success. Both techniques have been proposed for automatic segmentation with varying efficacy.

Overall, most techniques that have been proposed are dependent on the voxel intensity in the MR scan. This is logical since the information is so readily available

and reflective of tissue distribution. The voxel intensities are analyzed by the ML algorithms and a pattern is found in the data. The pattern is then used to segment an incoming test. The perception is that the ML algorithm can understand the visual perception of the images in the same manner as human operators. In the following chapter, we will discuss some key studies of automatic MS lesion segmentation that show this is not always the most efficient means for analyzing the data.

The limitation of using voxel intensity is assuming the ML algorithm perceives the data in the same manner we do, when it can actually perceive more. In image processing, there are many techniques that are used to analyze the distribution and pattern in the images described as image texture. Statistically, we can analyze the gray level pattern in the same manner as a probability distribution. Then the gray level distribution can be used to calculate mean, standard deviation, etc. Another technique is using frequency-spatial information to understand the level of change in the pixels and how often it changes based on frequency [6]. The frequency is analogous to the cyclical change in a wave and describes how quickly that change occurs. Therefore, image frequency is how quickly pixels are changing in the two-dimensional space.

There are recent studies that have shown the worth of texture analysis in MS lesion diagnosis [24] [37] [38]. Image texture is described as the spatial arrangement of color or intensity in an image. Texture is described as coarse or blocky when there are collections of pixels with similar intensity, while it is described as fine if there are clean lines in the pattern. We can consider the total texture of an entire image or a small portion of it. Studies have found that different tissues in conventional MR images have different texture. Specifically, the MS lesions have different texture than white matter (WM). The texture of the lesion is coarse, which is reflective of scaring, the collection of immune cells, and inflammation. On the other hand, WM has a fine texture of crisp healthy axons.

Two recent publications have shown the diagnostic indication of texture analysis. Zhang et al. [38] analyzed the image texture of lesions using the Stockwell Transform (ST). The ST is a Fourier Transform based technique used to find local frequency information using a frequency scaled Gaussian window. While Loizou et al. [24] found differences in statistical analysis of lesion texture, remarkably with the gray level co-occurrence matrix (GLCM). The GLCM is a second order statistic that counts the frequency of two gray levels appearing within a specified vicinity of each other. These two studies have found evidence of the texture differences in lesions and WM.

The idea that MR image texture is different between WM and lesions, leads to two research questions: 1)Is image texture different for other tissue classes? And 2) Does this information make segmentation more efficient? What we set out to do in this study was investigate the textural differences between healthy tissue and find a pattern recognition algorithm to learn these textural differences. We concluded on using a pattern recognition algorithm called Dictionary Learning (DL), which is based on regression analysis principles. Regression analysis is one of the most robust and efficient means of recognizing patterns for classification [25].

DL has the capability of learning a specified feature space, thus we explored what feature spaces would improve segmentation. The result is an elegant combination of voxel intensity and statistical analysis, that did not only segment lesions with high performance but is capable of dealing with minor image distortions such as noise. The following chapter will be a literature review to give the thesis some context of the success we wished to further. Then, we will discuss the DL algorithm and the concept of using different feature spaces in Chapter 3. Sub-sequentially in Chapter 4, we will propose the methodology and evaluation criterion to test the efficacy of our proposed method. In Chapter 5, we provide numerical evidence of the method's performance; on selection of parameter values and performance of our method compared to the patch based DL in mild, moderate, and severe lesion loads with or without

the presence of the noise. We include further a discussion of our method and future initiatives in the field of MS lesion segmentation in Chapter 6, before concluding the paper.

Chapter 2

Literature Review

For the context of this thesis, there are three major concepts to consider: MS lesion Segmentation Algorithms, DL segmentation for MS, and MR image texture. The following chapter will give a collection of articles depicting these concepts.

2.1 MS Segmentation Algorithms

An article on MS lesions segmentation explains a pipeline of tasks that will fulfill the goal [18]. The pipeline involves four steps: pre-processing, feature extraction, classification, and evaluation of performance. Figure 2.1 gives a flowchart giving the framework of these pipelines. Pre-processing is a set of image processing tasks that will manipulate or standardize the data in a form that is usable for the ML algorithm. Then for feature extraction, there have been many techniques used for this step, with the most popular being voxel intensity. There have been many classifiers used in the literature many of them being based on ML.

Given that the researcher has a pipeline for the method, they will require a set of data to test against. There is real patient data and simulated patient data. There are multiple data sets online that are simulated, most remarkably is the Brainweb set from McGill University [9]. Brainweb uses an MR imaging simulator that uses
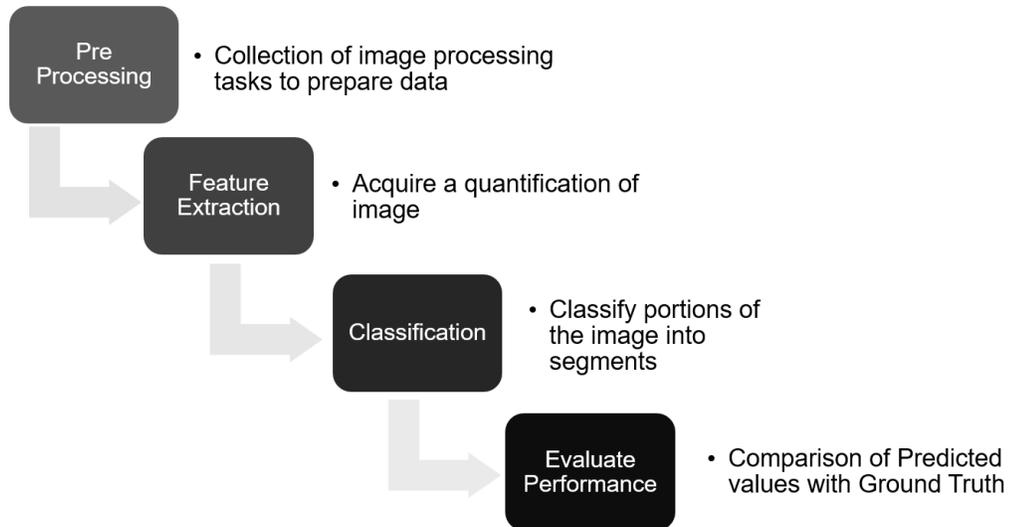
Figure 2.1: General flow chart of an MS lesion segmentation algorithm

first principles modeling based on the Bloch equations to recreate a discrete nuclear magnetic resonance signals. There are many real patient databases such as the 2008 Medical Image Computing and Computer Assisted Intervention (MICCAI) MS lesion segmentation challenge database [32]. In 2008 the MICCAI Society's convention held a challenge for MS lesion segmentation and compiled a database of patient scans with varying levels of MS lesion loads. Some studies have performed experiments on their own data set acquired from an affiliated hospital.

Multi-weighted MR scans have shown to be more efficient data sets for segmentation versus single weighted MR scans [18]. With an MR scanner, the image is reflective of nuclear magnetic resonance signal generated by the Hydrogen protons of the sample. From the nuclear spin of Hydrogen protons the nuclear magnetic resonance signal precesses like a spinning top and over time will align to the large magnetic field within the scanner. The time it takes for the atoms to return to the magnetic field determines the level of gray scale in the image. As seen from studies done with one weighting, there is not enough contrasting data for segmentation.

The use of more than one weighting provides a cross verification of tissue type and location. Therefore, new research should evaluate more than one weighting.

After the data set has been chosen and pre-processing is complete, the pipeline requires a feature extraction method to evaluate in order to classify the subjects in question. In an image, the subjects are the pixels of the two dimensional matrix. By grouping pixels into a region of interest it can be classified as a defined object. In the context of automatic MS lesion segmentation the classes are healthy tissue and lesion, with possible subdivisions of the healthy class to increase discrimination. Different methods define a pixel's feature space in different ways, typically catered to the classifier used to determine which class the pixel belongs to.

When using the pipeline on the data set with a classifier, there must be a separation of training data and testing data [22]. It is important that the training data and test data do not overlap so that there is no bias in the outcome. This is typically done by cross validation with leaving one scan out from the collection. Then the left out scan is segmented by the algorithm after training and has its answer compared to the ground truth. The ground truth are the perceived correct tissue segmentations in the image.

Once the pipeline is finished identifying the segments we must be able to evaluate the performance. Given a predicted segmentation and a ground truth segmentation we can evaluate where the answers are the same (true positives(TP) or negatives(TN)) and the answers that are not the same (false positive(FP) or negatives(FN)). When establishing ground truth, many techniques have been used. In the case of simulated data, such as Brainweb, the data has a ground truth to generate the images. Therefore, tissue and lesion segmentation is readily available prior to experimentation. In the case of real patient data, techniques have been used to segment tissues, such as voxel-morphometry [5], with the lesions in real patient scans segmented by a radio-neurologist.

Once the similarities of the predicted values are compared to the ground truth we want to be able to understand how well it did with one indicator. The industry has began following the lead of classical information retrieval systems to quantify this performance indicator. For automatic segmentation the industry has moved toward using what is called the Dice Similarity Coefficient (DSC) [39]. The DSC is the harmonic mean of precision (positive predictive value) and recall (sensitivity) [8].

$$Precision = \frac{TP}{TP + FP} \tag{2.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.2}$$

Precision is described as the fraction of correct positive answers out of all positive answers returned by the classifier. Where recall is the fraction of correct positives out of all the positive values that should be returned by the classifier. Both of these concepts are important in the evaluation of a mathematical classifier. To combine both of them in a matter that is unbiased to the larger value, the harmonic mean is taken to get DSC.

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{2.3}$$

Historically, the indicators used for testing the similarities of two segmentations have been specificity, sensitivity, and accuracy. These indicators are less intuitive to the location differences, as opposed to understanding size differences. By looking at the Dice Similarity Coefficient (DSC), we can evaluate the efficacy of the segmentation algorithm which is more location dependent. A DSC score of 0.7 or greater is considered a good agreement between two segmented images [39]. Therefore, DSC will be an acceptable means of evaluating studies. More explanation of DSC can be found in Section 4.4.

The previous attempts at solving MS lesion segmentation are divided into the two categories of supervised and unsupervised techniques [21]. Supervised techniques are considered methods that use a database of information. Previously supervised techniques have had two kinds of databases made being atlas based and clinical based. Unsupervised techniques do not require a database of information, instead they have a mentality that segmentation can be done on the data presently in the image. The subcategories of unsupervised methods are segmenting based on lesion features or based on tissue features.

Unsupervised lesion based techniques define MS lesions by the characteristics in the test image [23]. These techniques tend to be faster than supervised techniques because they do not have to cross reference with a database or train a classifier. Datta et al. (2007) [12] developed a pipeline using morphological operators and fuzzy connectivity to segment MS lesions. Morphological operators refers to concepts of set theory such as intersection and union. While fuzzy connectivity is the concept of having the probability of belonging in a set assigned to a point. Datta took a post-contrast T1w image and subtracted the pre-contrast T1w image then found the associated voxels with T2w images. Then further delineates lesions using fuzzy connectivity. The DSC for this study on average was 0.76. Critically this system would work well if all MS patients were scanned in this process. The procedure of pre- and post-contrast scans can be quite cumbersome and make the scans more susceptible to movement.

The last example of unsupervised segmentation methods are the tissue based techniques. In this technique the mentality is that tissue segmentation will guide the MS lesion segmentation [23]. This technique depends on the quality of tissue segmentation. Typically, the whole image is segmented for healthy tissues and then large errors in representation (outliers) are considered voxels of MS lesions. Garcia-Lorenzo et al. (2008) [19] developed an unsupervised tissue based segmentation algorithm that uses

two steps. First performing a local segmentation using the mean shift algorithm to generate local regions in the image. Then using a variant of expectation maximization was employed to classify the regions as healthy brain tissue or lesion. The result is an algorithm with a DSC of 0.55 on average.

Critically these unsupervised techniques suffer from the variability of image quality and artifacts [23]. Artifacts are image distortions caused various factors, such as patient movement and field of view miscalculation. Since MR imaging is dependent on the apparatus doing the imaging, acquisition parameters vary between scanners. The image quality from one scanner may be different from the next. Supervised techniques have the advantage of having a classifier trained to deal with artifacts and noise. Unsupervised segmentation algorithms may work for one scanner but not for another, although the concept of tissue based unsupervised segmentation is a logical means of delineating lesions. By using the healthy tissue to guide the lesion segmentation we get a more discriminative method. Therefore, we will use this concept in the proposed method described later in Chapter 3.

Supervised atlas based techniques use an atlas compiled from previous clinical scans [23], where a probabilistic model of tissue distribution is tabulated from the previous scans. The benefit of this is the atlas has good local information and a probabilistic framework. Mainly statistical atlases are compiled from a large set of clinical scans, from both MS and non-MS patients. Then, probability of voxel tissue classification is found from comparing to the atlas. Topological atlases are used as well, where they encode a specific topology of each structure and group of structures. These atlases preserve the topological form of the tissue being segmented. Shiee et al. (2010) [31] developed an automatic lesion segmentation method that used a combination of statistical and topological atlases. The method uses the fuzzy-C means algorithm for classification when comparing a test to the two atlases. The segmentation method was called topology-preserving anatomical segmentation. He

achieved some very impressive results with an average DSC of 0.789 an a set of real patient data using T1w, T2w and FLAIR images. This shows that supervised atlas based techniques are a competitive segmentation technique.

Unfortunately, the disadvantage of supervised atlas based techniques is that the test scan must be co-registered with the atlas to have the classification system work. Co-registration has flaws [3], in every attempt at registering photos to the same co-ordinate system, you have an uncertainty in placement. The amount of computation time to register images accurately increases the overall time of segmentation. Therefore, there is benefit in avoiding this registration step.

The alternative supervised technique avoids registration to an atlas. Supervised clinical based techniques use a database of manually segmented clinical images [23]. The manually segmented images are annotated by a neuro-radiologist. In these methods registration with the database is not needed. This results in a more robust algorithm because of its comparison to clinical results. This technique requires a classifier and a training step. These techniques incorporate a training step in their pipeline to have the classifier become familiar with the previous examples from the database. The only disadvantage of the technique is the computational time for training the classifier. Anbeek et al. (2005) [2] is a significant researcher in supervised clinical based MS lesion segmentation. Anbeek used a k-Nearest Neighbor classifier built on a feature space of voxel intensities and spatial information. The database used five different MR weighted scans. The result was a probabilistic mapping of white matter lesions to which a thresholding results in a binary map that can define the lesions. The DSC for this study was 0.805.

All of the previous techniques have challenges that should be mentioned before moving on. For supervised methods, the challenge is to find the proper database. The number of scans, the weighting and the feature set are important qualities to be considered. An incoming supervised technique will need to explore the options

of their database. For unsupervised methods, the main challenge is the variation of image quality between scanners and patients. The acquisition of MR scans varies between machines. Therefore, unsupervised methods should find an invariant method for acquisition differences.

Supervised clinical based techniques are extremely useful and seem to be the best possible technique for achieving the goal, but it is still not using all the information from the images. In the studies we evaluated with this technique, no one has evaluated the use of texture data for classification for MS lesion segmentation. All studies have focused on the voxel intensity. Studies tended to use large feature spaces of many voxel intensities increasing computational times. In looking at image texture, smaller feature sets could be calculated by measuring the voxel distribution in the MR scan.

2.2 Dictionary Learning in MS Lesion Segmentation

The proposed pipeline of this study is to use a classifier called DL [25] that has been used in recent years for MS lesion segmentation. DL is a supervised ML technique using regression analysis for classification. Typically this technique has been used on clinical based databases.

For a mathematical classifier, we have different objects that we would group together in a class. If an object is not similar to the given class, then it would be put into a new class of objects and we can discriminate between different classes. For the case of MS lesion segmentation it is customary to create classes for the different tissues visible in the MR image; healthy tissue and MS lesions. Based on clinical segmented images we can collect examples of each class into a dictionary. Then by consulting the dictionary, we can classify a new object into healthy tissue or lesion.

For images we can take the pixel intensities and concatenate them into a vector and then collect these examples into a matrix as shown in Figure 2.2.
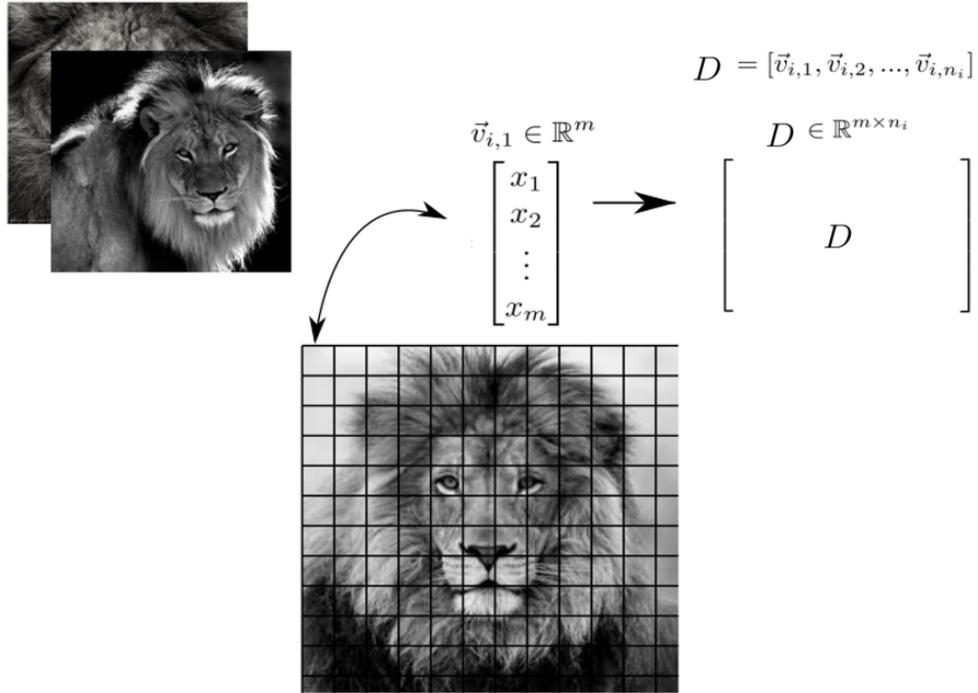
Figure 2.2: Visualization of DL: creating a dictionary for identifying a lion from examples of various images of a lion

The DL classifier uses the least absolute shrinkage and selection operator (LASSO) algorithm to calculate the regression coefficients $\vec{\alpha}$ [33]

$$\min_{\alpha} \frac{1}{2}\|\vec{x} - D\vec{\alpha}\|_2^2 + \lambda\|\vec{\alpha}\|_1 \tag{2.4}$$

where $\vec{x}$ would be a test object, D is the dictionary of a given class, and λ is a sparsity coefficient.

The LASSO finds a sparse solution to $\vec{\alpha}$ because of the nature of the ℓ_1-norm penalization. A sparse solution is desirable because it improves the interpret ability of the model and lowers the computational time of regression analysis. In the case

of MS lesions, we can create dictionaries to learn the healthy tissue in the brain and the lesions to classify image patches of a test image [13].

Weiss et al. (2013) [36] was the first to attempt this process. This study built one dictionary that was meant to represent the healthy tissue of the brain in T1w and T2w MR images. By learning the healthy tissue of the brain, any voxels that had a high representation error meant they were not reflective of healthy tissue. By thresholding out the high error voxels, Weiss found the lesions in the scan. The DSC for this study on average was 0.71 for Brainweb, and a DSC of 0.294 for a database of real patient scans. This first attempt of DL with MS lesion segmentation showed great potential.

Based on the Weiss' study, Deshpande et al. (2014) [13] was the next to explore the uses of DL for MS lesion segmentation. Deshpandes first attempt created two dictionaries as opposed to one; a dictionary to learn healthy tissue and one to learn lesions. This study also used more MR image weightings; T1w, T2w, PDw and FLAIR images. What was found was that if the lesion dictionary was smaller than the healthy tissue dictionary, the segmentation was more accurate. They found a sensitivity of 0.5414 on average from a sample of real patient data.

Deshpande et al. (2015) [14] attempted this technique again, but in this study, they used four dictionaries. The four dictionaries represented the different tissue classes of the brain; white matter (WM), gray matter (GM), cerebral spinal fluid (CSF) and lesion (LES). The results showed that having more classes guided the classification more efficiently in the same way as unsupervised tissue based techniques. They found that having a smaller dictionary for the lesion class was more beneficial again. The rationale they had was that the tissue quality of the lesions is less complex than healthy tissue and thus less examples are needed. The DSC for this study was 0.498 for the same database of patient he had used previously.

The use of DL has previously shown promise for the task of MS lesion segmentation. It has a proven robustness given the regression analysis foundation and it is computationally competitive to other supervised trained systems. A concern with DL is the choice of feature space. How much information should be used in the calculation of class definition. With voxels, we can take a patch of any size and consider a lot of information, but the computational time goes up. What we set out to do is explore image texture analysis to find a more efficient feature space for segmentation. In the following section, we will review some image texture analysis techniques and see the contrasting calculations between healthy tissue and lesions.

2.3   MR Image Texture and MS

Some recent studies have shown that voxel intensity may not be the best way to interpret the images. In image processing, texture analysis is a technique that quantifies the spatial pattern of pixels in the image. Texture analysis has been proven to have potential in the management of MS lesions [37].

Texture analysis can be described in two main ways, statistical and spatial-frequency [37]. In the statistical method, we can describe the pattern of gray levels as a probability distribution and then acquire statistical information from the distribution. Classically, we can plot a histogram of the information in the photo (Figure 2.3) and acquire statistical information (Table 2.2).

Table 2.1: Statistical look at 5x5x5 voxel patches of WM vs. LES of T2w images used in Figure 2.3

Tissue	Mean	Standard Deviation
White Matter	0.2637	0.0060
Lesion	0.4754	0.0404

The more insightful method is to use a statistic that describes the amount of times certain gray levels arise in proximity to each other. We can define a matrix for

18
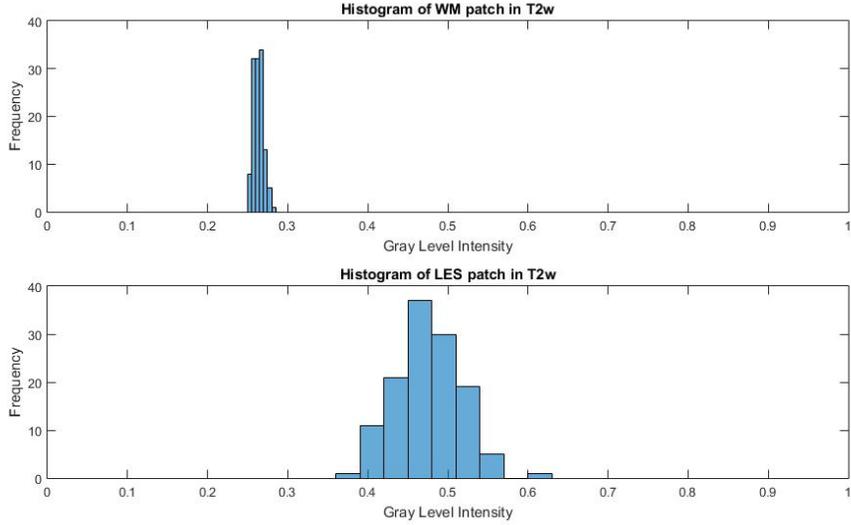
Figure 2.3: Histogram comparison of a WM patch of 5x5x5voxels and LES patch of 5x5x5voxels in T2w MR images

an image that counts co-occurring pixel values at a given offset. This is called the gray level co-occurrence matrix (GLCM). Loizou et al. (2015) [24] explored many statistical calculations of T2w MR images. What they found was the GLCM had significantly different values between MS lesion and healthy WM. The GLCM counts the frequency of times a voxel intensity appears at a certain distance and angle away from another voxel intensity $(\Delta x, \Delta y)$.

$$GLCM_{\Delta x, \Delta y}(i,j) = \sum_{x=1}^{n} \sum_{y=1}^{m} \left\{ \begin{array}{ll} 1, & \text{if } I(x,y) = i \text{ and } I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otherwise} \end{array} \right\} \quad (2.5)$$

In Figure 2.4, we calculated the GLCM for a 5x5x5 voxel patch from T2w WM and lesion, using a distance of one voxel away at any angle. We then calculated the joint probability by dividing the matrix by the sum of all of its elements. What results is a matrix of these joint probabilities, to which the 3D surfaces are graphed from. We can then see the difference between WM and LES GLCM's in this manner.

19
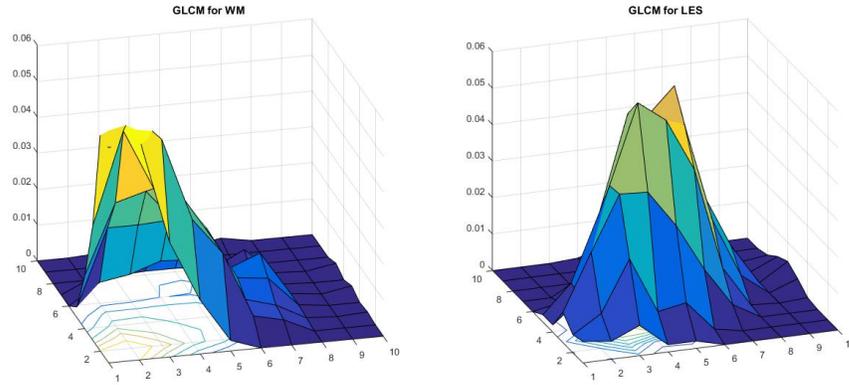
Figure 2.4: 3D visualization of a GLCM for WM and LES patches of 5x5x5 voxel in a T2w scan

The second way to explain texture is using spatial-frequency information. By using the Fourier Transform (FT) we acquire the frequencies observed in the image. The frequency of an image can be described as the rate of change between pixels. The FT gives an accurate account for the global frequency information of an image but when analyzing sections of an image FT is not accurate at finding local frequency. Therefore, alternate versions of the FT are used instead. Zhang et al. [38] used the Stockwell Transform to analyze MS lesions. ST is a variation of FT in which a frequency scaled Gaussian window is used to localize the frequency information in the image. Zhang et al. then took the polar angular integral of the frequency information of MS lesions and WM. This resulted in the local radial frequency distributions being different for the tissues.

By taking these studies of texture analysis with MS lesions, it is possible to define the examples in our dictionaries using texture quality in the effort to reduce the feature space. The DL would learn the texture of the image as opposed to the image intensities, thus resulting in a more effective segmentation algorithm.
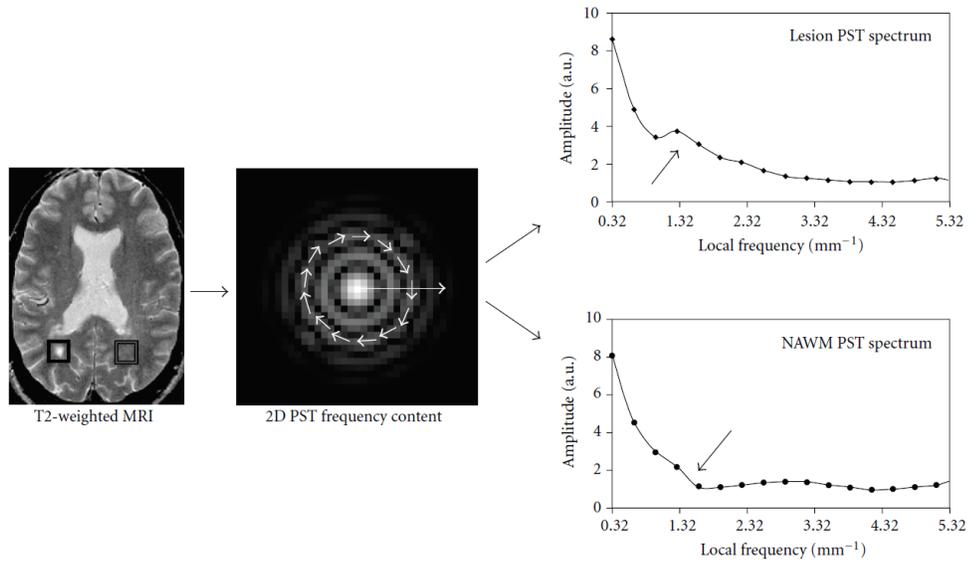
Figure 2.5: A flow chart of Zhang et al. investigation into the local frequency difference between normal appearing WM and LES [38]

2.4 Literature Review Conclusion

We have reviewed the basis of the current study by looking at previous MS lesion segmentation algorithms, significantly analyzing DL methods, and introducing MR image texture analysis. The following chapters will explain the theory and method used in this study. What was found is a feature space with reduced dimensionality for a quicker computational time and precise algorithm. The classification is done with a supervised learning classifier used in a tissue based discrimination framework. The classifier was taught using the voxel intensity in question from each MR weighting with the spatial voxel intensity arrangement calculated using the standard deviation of adjacent voxels.

Chapter 3

Dictionary Learning in Feature Space

The classifier in the proposed pipeline is a machine learning theorem resulting from the work of Tibshirani [33]. The LASSO is constructed on the principles of regression analysis. Specifically, the mathematical description of linear regression is as

$$y = \beta x + \epsilon \tag{3.1}$$

where y is the dependent variable being estimated by the independent variable x. The error of estimating the dependent variable based off the dependent variable is denoted by ϵ. If the error is normally distributed with a mean of zero, then the value of β that is the solution to the convex optimization problem

$$\min_{\beta} \frac{1}{2}\|y - \beta x\|_2^2 \tag{3.2}$$

is the maximum likelihood estimate of β. This value of β gives us an estimate of the linear relationship of x and y.

In the situation that calls for finding a limited amount of β coefficients to represent y from the available x, we can limit the number of non-zero components of β set using an ℓ_p-norm.

$$\left\| \vec{\beta} \right\|_p := \left(\sum_{i=1}^{n} |\beta_i|^p \right)^{1/p} \tag{3.3}$$

When $p = 0$, the ℓ_0-norm counts the amount of coefficients that do not equal zero. Therefore, if we wish to limit the amount of β used, then we would write the convex optimization problem as

$$\min_{\beta} \frac{1}{2} \|y - \beta x\|_2^2 \quad such\ that \quad \|\beta\|_0 < k \tag{3.4}$$

using the ℓ_0-norm to limit the amount of β coefficients. The solution to this equation would be sparse having at most k non zero solutions, thus improving the interpret ability of which x make up y. Unfortunately, the ℓ_0-norm is intractable at this time and can not be computed efficiently. In 1996, Tibshirani [33] suggested using the ℓ_1-norm instead

$$\min_{\beta} \frac{1}{2} \|y - \beta x\|_2^2 \quad such\ that \quad \|\beta\|_1 < k \tag{3.5}$$

The ℓ_1-norm results in a sparse solution because of the restrictive nature on the solution set [25]. Figure 3.1 a) is a visual description of the ℓ_1-norm ball on a 2 dimensional solution set. We can see that any solution outside of the ball will be projected onto it in the case of the LASSO. The larger red and green regions will project to the basis vectors of the space resulting in a sparse solution. Therefore, the ℓ_1-norm penalization gives a solution that is sufficiently sparse compared to using the ℓ_2-norm which results in a curving restricted solution set resulting in non sparse solutions (Figure 3.1 b) [25]. The regression coefficients are easily calculated using the Least Angles Regression Algorithm [15].
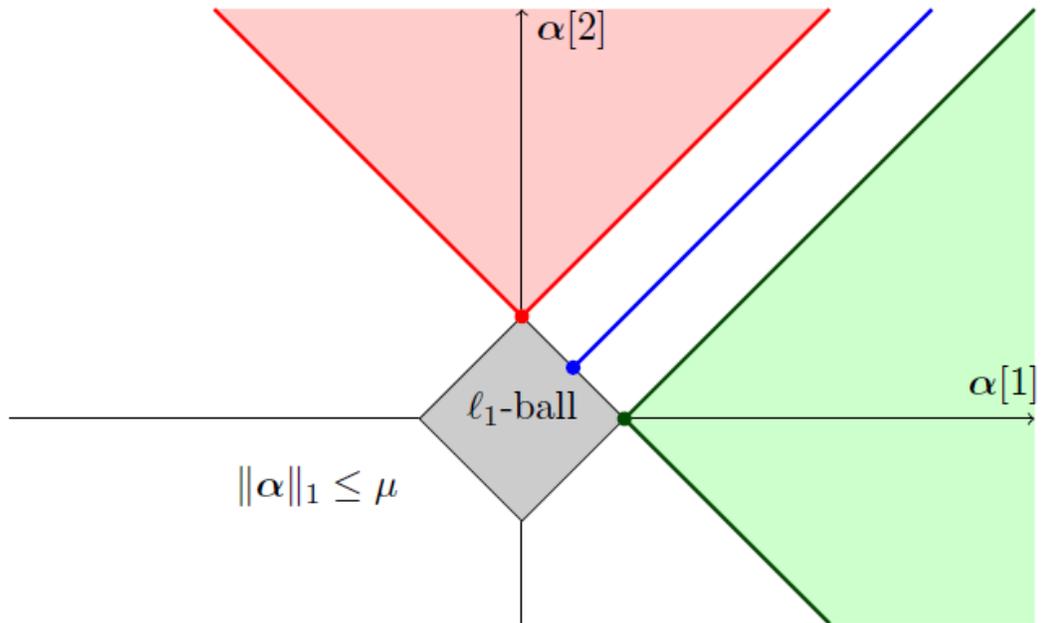
(a) Effect of the Euclidean projection onto the ℓ_1-ball.
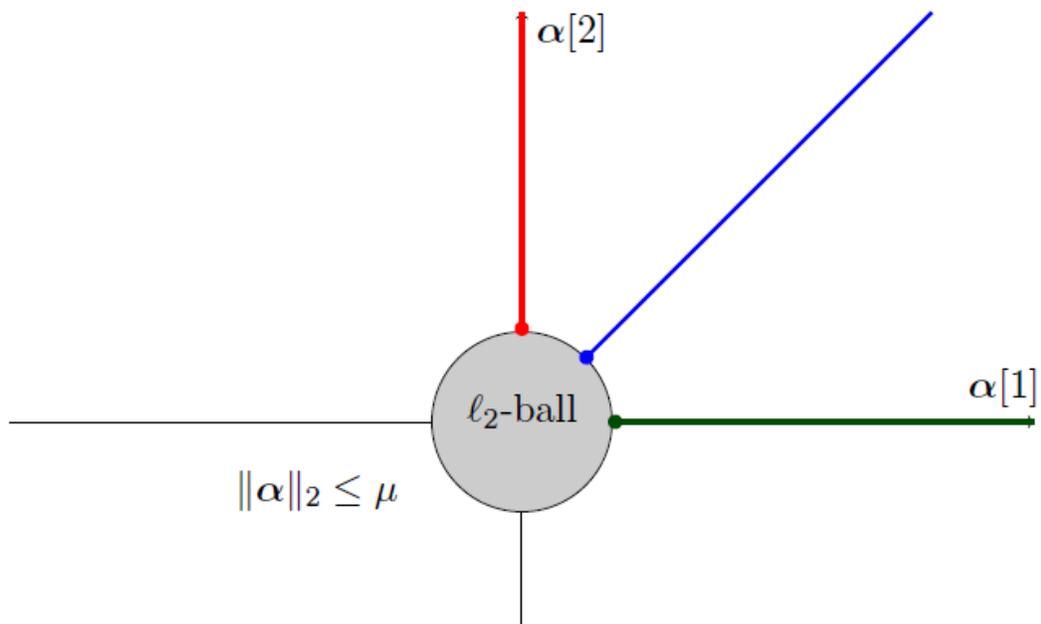
(b) Effect of the Euclidean projection onto the ℓ_2-ball.

Figure 3.1: Visualization of the ℓ_p-norm on solution set: a) solution penalization of the ℓ_1-norm, b) solution penalization of the ℓ_2-norm. Retrieved from [25]

Finally, by using Lagrange multiplier theory we can add this limitation into the original formula to get one full convex optimization problem[25]

$$\min_{\beta} \frac{1}{2}\|y - \beta x\|_2^2 + \lambda\|\beta\|_1 \tag{3.6}$$

where λ is a sparsity parameter. The sparsity parameter λ determines the proportion of the original ℓ_1-norm that will remain after optimization, as opposed to restricting the numerical amount of the ℓ_1-norm. This is the final form needed to understand the concepts of DL.

Given that we have a vector $\vec{x} \in R^n$ we want to determine if it belongs within a certain class of vectors, possibly quantified from MR images. Therefore, the classes may be healthy tissue or lesion. Assuming previous examples of the given tissue class could be combined to recreate $\vec{x}$ perfectly, then $\vec{x}$ would belong to that tissue class. We take a collection of m examples of the particular tissue class and concatenate them into a dictionary matrix $D \in R^{n \times m}$. Then by solving the LASSO we would find an estimate of $\vec{\alpha} \in R^m$ that would recreate $\vec{x}$ using a limited amount of the columns in D

$$\min_{\vec{\alpha}} \frac{1}{2}\|\vec{x} - D\vec{\alpha}\|_2^2 + \lambda\|\vec{\alpha}\|_1 \tag{3.7}$$

If $\vec{x}$ does belong to the class represented by the dictionary then the residual of the $\vec{\alpha}$ vector combination of D will be zero.

$$r(\vec{x}) = \|\vec{x} - D\vec{\alpha}\|_2^2 \tag{3.8}$$

Now, instead of having one class, we have C number of classes for our classification. Then we would have different dictionaries D_i for each class $i = 1, 2, ..., C$. To classify the test vector properly we would find the minimum residual given each dic-

tionary and the minimum residual will declare which class that $\vec{x}$ belongs to, namely

$$Class_{pred} = \min_i \| \vec{x} - D_i \vec{\alpha_i} \|_2^2 \qquad (3.9)$$

There is a complication that arises. The LASSO is quite capable of dealing with large data sets and high dimensional feature spaces n. DL is quite useful in learning complex high dimensional systems. If the choice of feature space n and the number of examples m are very large the computational time can be staggering. Therefore, a minimal amount of m for each dictionary will result in a swift algorithm.

Previously, in [34] [13] [14], DL has been used in a feature space of large n to create a discriminating feature space for MS lesion segmentation. They would take a 3x3x3 or 5x5x5 voxel patch for each MR image weighting to describe an example in the dictionaries giving $n = 108$ and $n = 500$ respectively, thus resulting in high computational times. Their rationale for a large feature space was that the dictionary is discriminating between the localized texture to classify voxels into a certain tissue class.

Although their DSC scores were comparable to previous segmenting techniques, they suffered from high levels of false positives. From an image processing point of view this could be explained by the lack of precision in large patch sizes. When segmenting voxels, a localized texture quality is evaluated but the individual voxel intensity is the main defining factor. Therefore, we propose not using a patch of localized texture but using only the voxel intensity at one point. The voxel intensity between each four weightings has a distinct combination of light and dark quality which is discriminative enough to achieve the goal, thus decreasing the feature space dramatically to $n = 4$. Figure 3.2 shows the voxel intensity pattern of different tissues over each MR image weighting.
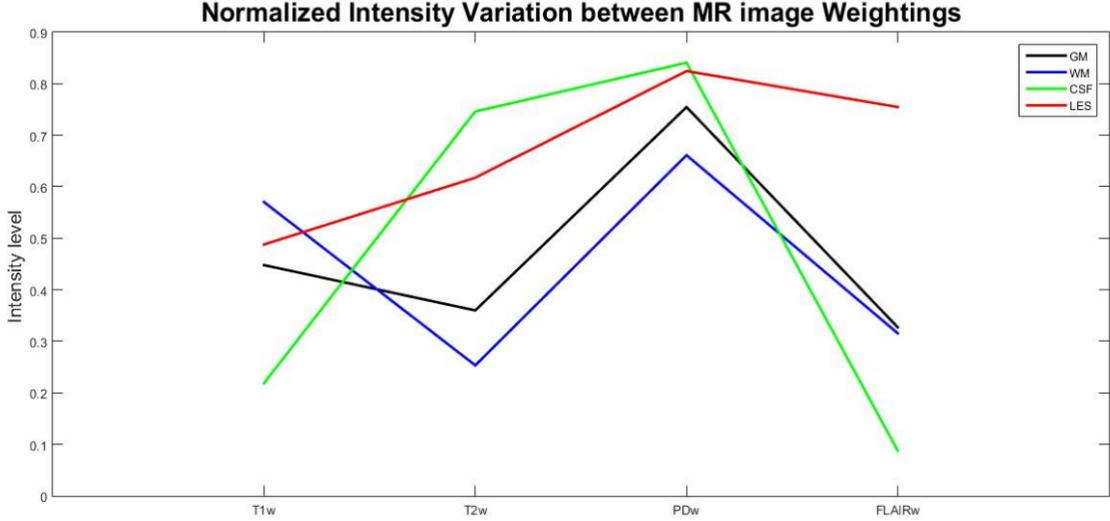
Figure 3.2: Intensity differences between MR image weightings

This pattern is very distinct, but the fact that it only uses on voxel eliminates the texture quantification that a larger patch has. Texture is described as the arrangement of many pixels and is not quantifiable if we are looking at just one pixel. Therefore, we evaluated many different texture analysis methods to find a texture quantification to pair with the voxel intensity. These techniques will be elaborated on later in Chapter 6, the discussion, since we believe there is further research to be done on this subject.

What we found to be beneficial in lesion segmentation was pairing the voxel intensity with the standard deviation of adjacent voxels.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2} \qquad (3.10)$$

By taking into account the variation around the voxel in question, the algorithm sufficiently segments in the presence of noise and image distortion. We constructed a feature space of T1w intensity level, σ of T1w 3x3x3 voxel patch, T2w voxel intensity level, and etc. which we graphed in Figure 3.3 to view the pattern of change between each feature space coordinate.

NormalizedTissue quality including intensity and standard deviation (STD)
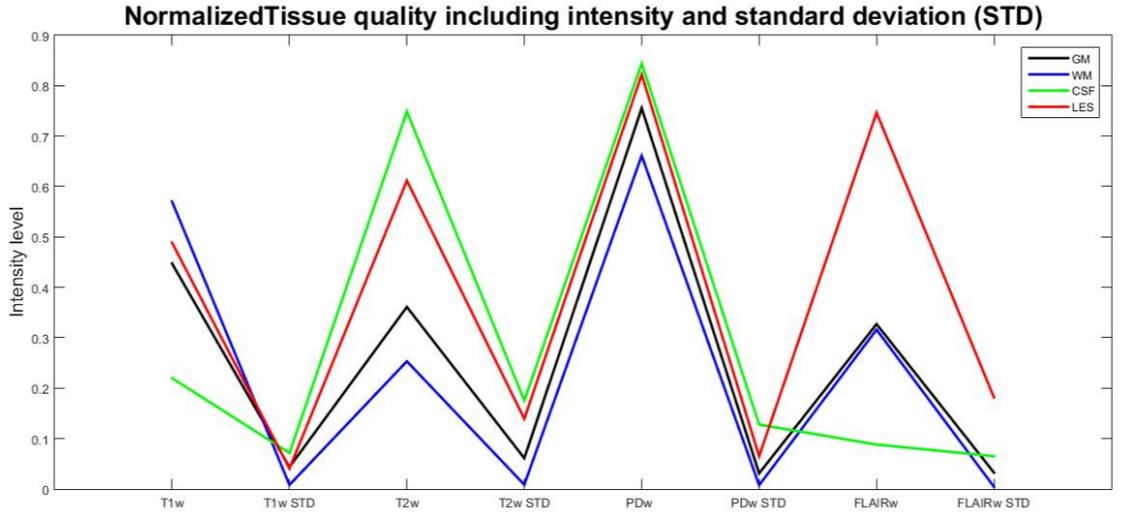
Figure 3.3: Intensity differences and standard deviations between MR image weightings

Proceeding this chapter, we will discuss the simulation methodology used to prove the efficacy of the given method. There will be a review of the data set chosen and its benefits, the feature extraction methods that are compared. There will also be a description of the DL classifier training software that was used and a comprehensive evaluation of performance indicators before giving the results.

Chapter 4

Methodology

The pipeline proposed here is a supervised clinical based segmentation algorithm that takes the concept of unsupervised tissue based methods into consideration. When performing classification the dictionaries will be learned for three healthy tissues (GM, WM, and CSF) and the lesion tissue. Based on the characteristic tissue dictionary, we can quantify the previous examples in different feature spaces to evaluate the optimal outcome. The process we took is depicted in Figure 4.1 and will be explained in the following chapter.

4.1   Pre-Processing

For this study, we decided to use Brainweb simulated scans [9]. Brainweb is appropriate because it has a known distribution of discretized tissue. The simulator takes the discretized patient and solves the Bloch equations given a set of relaxation times. In this sense we can set relaxation times manually to achieve different MR weighted images. This makes it possible to generate T1w, T2w, PDw, and FLAIR images for the study (Figure 4.2). The simulator comes available with four levels of lesion load; no lesions, mild load, moderate load and severe load. We can use three of these lesion loads in the study.
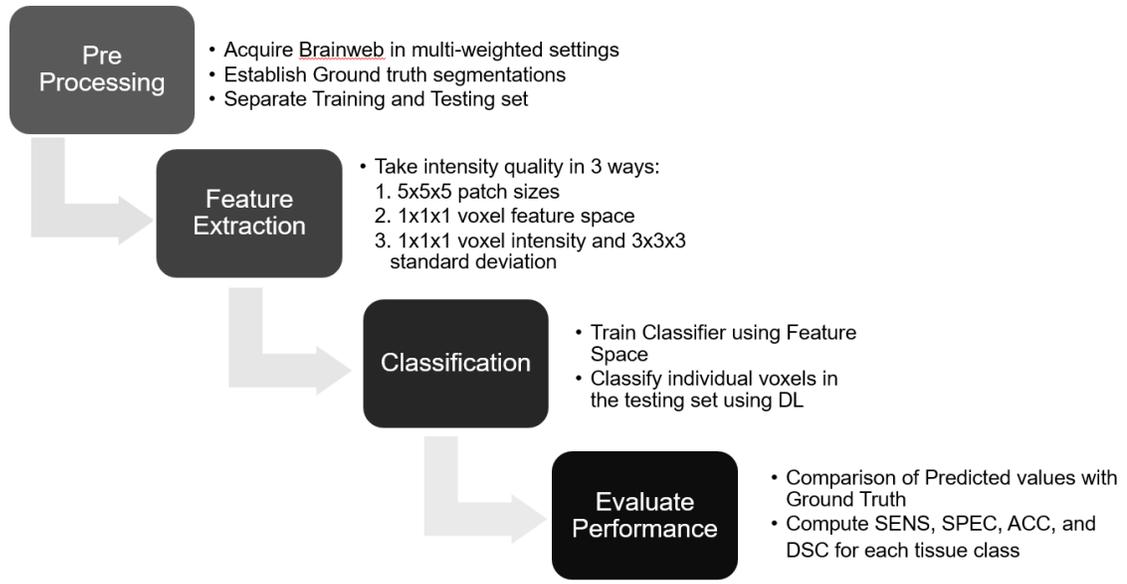
Figure 4.1: Proposed pipeline

Brainweb has tissue maps available for each tissue in the simulated brain (Figure 4.3). They include GM, WM, CSF, and lesions, as well as fat tissue, muscles/skin, etc. that are not needed for the study. We cross referenced these maps with an independent tissue segmentation technique called voxel-based morphometry (VBM) [4]. VBM is a tissue segmentation algorithm that uses statistical parametric mapping to locate the tissues of the brain. It is easy to compute this using the Statistical Parametric Mapping 12 (SPM12) software. The tissue maps in this study were defined using SPM12 and the available tissue maps on Brainweb.

The outcome of VBM is a statistical mapping of tissues, so each voxel has a probability of being in each tissue class. We decided to let the highest probability of tissue define the voxel tissue and created a binary map showing definite tissue association. Unless the given voxel was defined by the lesion map, then the voxel was considered a lesion.

The lesion tissue maps from Brainweb needed some processing before use. Since they are tissue voxels that are fed into the Bloch equations, the maps include nor-
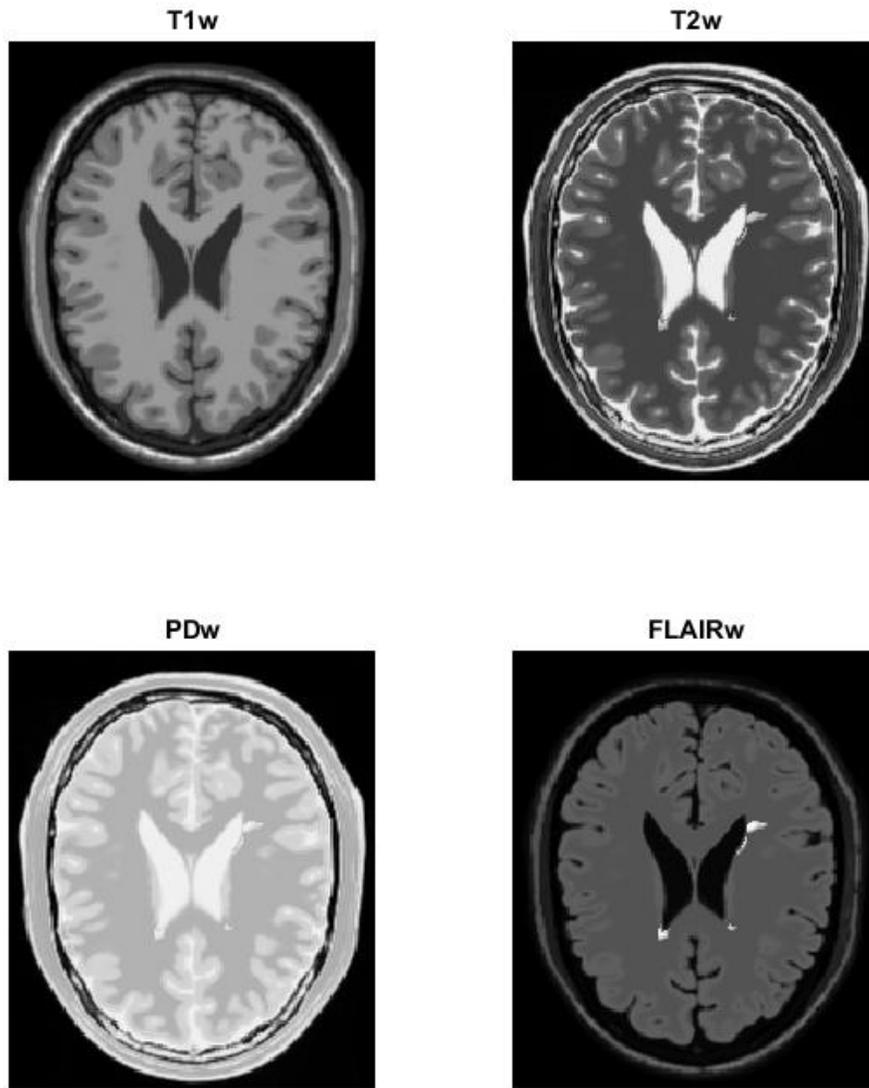
Figure 4.2: Different MR image weightings

mal appearing white matter and some invisible lesion activity for authenticity. To narrow the ground truth of the lesions, we took the voxels only contributing to scan intensities that were visibly different in all four MR weightings. Then we took the
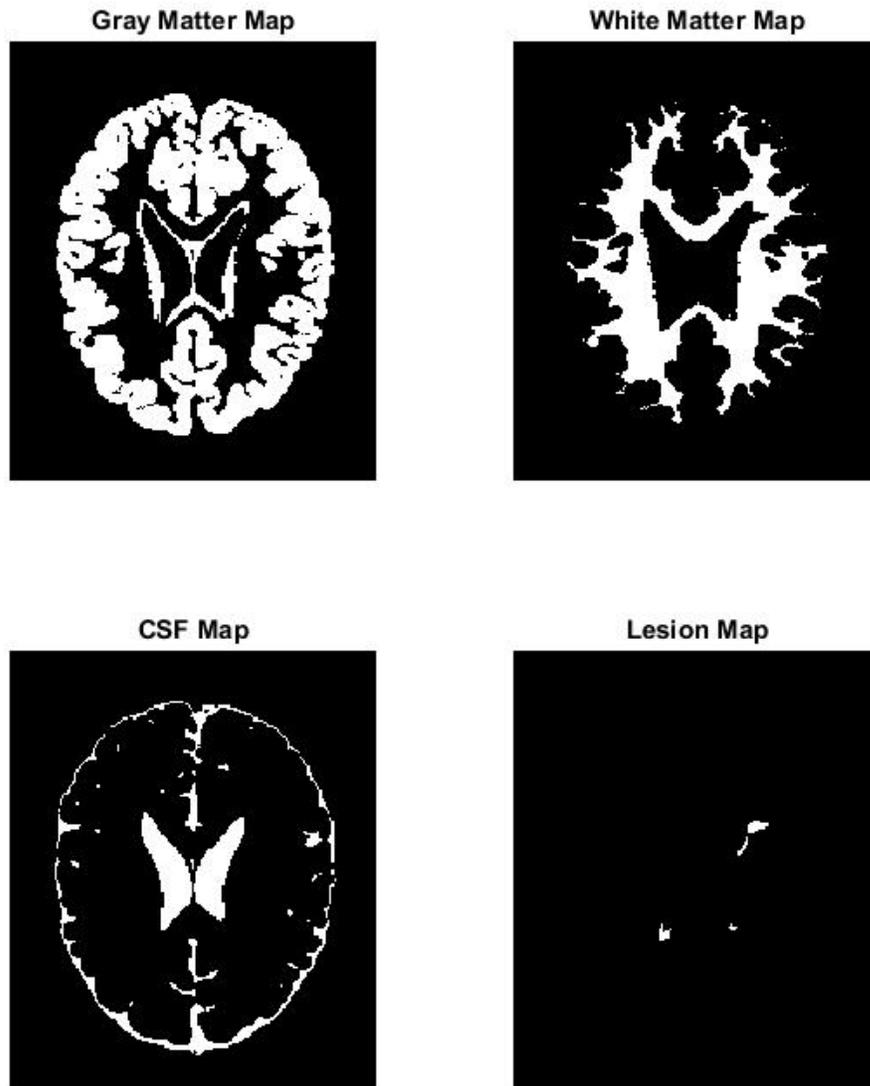
Figure 4.3: Tissue maps used in the method

voxel intensities from the lesion scans and transplanted them into the healthy scan.
Therefore, our target has been clearly defined.

Finally, the soft tissue, and bone that is outside the brain tissue was eliminated
from the scan by adding all four tissue maps together and multiplying by the original
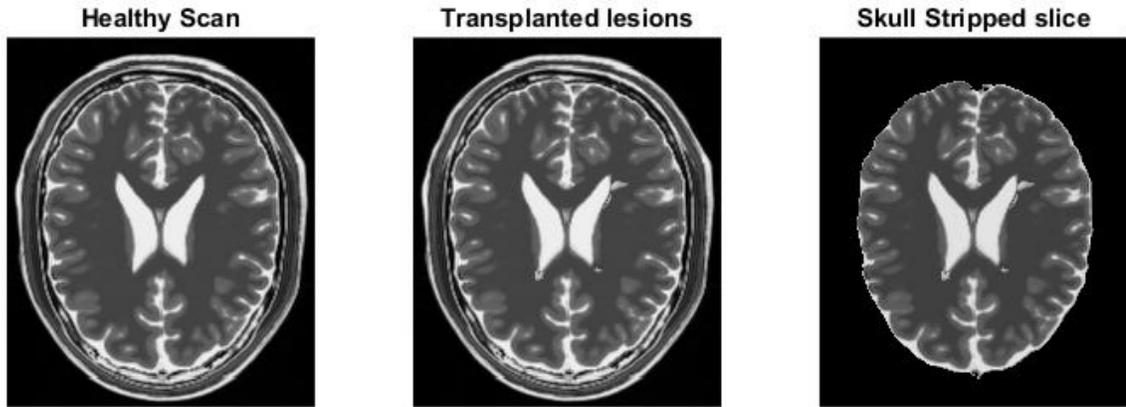
Figure 4.4: Final processing of lesion transplantation

scan. Therefore, sufficiently skull stripping the scans. This is acceptable because most segmentation pipelines will skull strip using software such as Robust Brain Extraction tool.

4.2 Feature Extraction

Now that we have the scans prepared, we move onto taking samples of given tissue classes. For each scan we can find all the voxels that belong to a particular tissue class and catalog them. These voxels of interest will be different for each lesion load scan.

What is more important is to find the voxels of interest pertaining to our test slices. In some previous studies they performed testing on a full scan to see how well the algorithm works. Unfortunately, with Brainweb, we can only obtain three lesion scans and leaving out one whole scan is too much information to cross validate the pipeline on. We decided to segment a collection of fifteen slices for each simulation to get statistically significant data. Therefore, the cataloged voxels of the test slices will be put aside as the test group, and then the remaining voxels are used for training as per the explanation in Section 2.1.

33

Then with the training group we take the voxel intensity of each weighting and create a vector of them for the Single Voxel method. For the Single Voxel and σ method, we take the voxel intensity and also take the standard deviation of all adjacent voxels. The standard deviation was calculated for all 27 voxels of a 3x3x3 voxel patch centred at the voxel of interest. For the large patch method, we took a 5x5x5 voxel patch centred at the voxel of interest.

4.3   Classification

Once the tissue distributions are defined in their feature space, we can then input them into the Sparse Modeling Software (SPAMS) package to get a dictionary for each tissue. SPAMS [26] [11] has a dictionary learning function that uses Online Dictionary Learning (ODL) to create an optimal dictionary for sparse solutions. ODL is on optimization algorithm based on stochastic gradient descent. What occurs is that the algorithm initiates a random collection of examples from a given distribution $p(\vec{x})$. It takes an extra sample from the distribution and solves the sparse representation using the LASSO. Then the columns of the dictionary are updated to consider the error in representation. Ultimately this results in a dictionary that gives the sparsest solution possible for every example in $p(x)$.

In the test slices, we then classify the voxels using the trained dictionaries. We will get a voxel classification depending on its feature space for each method. For each voxel, we take the same quantification and perform the LASSO to find the representation given each tissue dictionary. Then the dictionary that recreates the test with the smallest residual will classify the voxel as that tissue. Once the slice has been fully segmented, we can evaluate the results.

4.4   Evaluation of Performance

Once we have a segmented slice we compare it to the ground truth of that slice. For each tissue map we count the number of true positive, true negatives, false positives and false negatives. Table 4.1 gives a list of performance indicators that we will evaluate. As described in Chapter 2, the DSC will be our main indicator for evaluating segmentation because of it's combination of precision and recall.

Table 4.1: Indicators that are used to evaluate image segmentation [21]

Name	Description	Formula
Sensitivity	True positive rate: measure of the amount of positive values correctly identified	$SENS = \frac{TP}{TP+FN}$
Specificity	True negative rate: measure of the amount of false values correctly identified	$SPEC = \frac{TN}{TN+FP}$
Accuracy	Overall evaluation of correct answers out of all answers given	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$
Dice Similarity Coefficient	An indicator that characterizes precision and recall of a classifier	$DSC = \frac{2TP}{2TP+FP+FN}$

Figure 4.5, provides a simple segmentation example. Here we would be expecting to predict the box segmented on the right, but find the result of the left side which has over segmented the box. If we count the amount of true positive, false positive, etc. we get the following results and indicators in Table 4.2. We can see how high the sensitivity , specificity, and accuracy are in this example. Where the DSC is not as high reflecting the poor performance of this example's segmentation algorithm.

Table 4.2: Indicators that are used to evaluate image segmentation [21]

Results	Indicators
TP=361	$SENS = \frac{TP}{TP+FN} = 1$
TN=7192	$SPEC = \frac{TN}{TN+FP} = 0.95$
FP=368	$ACC = \frac{TP+TN}{TP+TN+FP+FN} = 0.95$
FN=0	$DSC = \frac{2TP}{2TP+FP+FN} = 0.66$
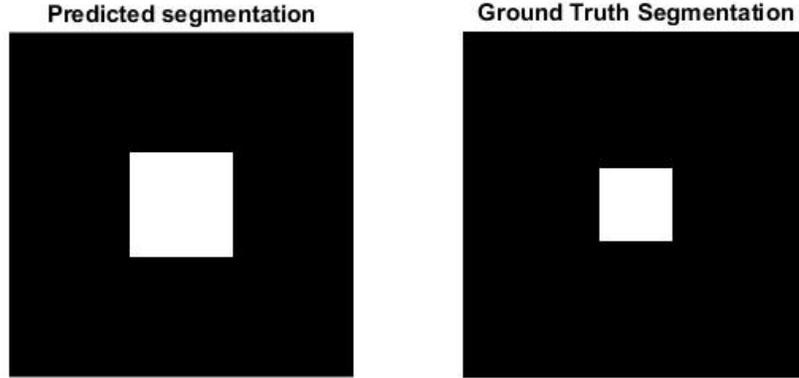
Figure 4.5: Simple segmentation of a box as an example of indicator performance

We then repeat the testing and the evaluation steps on each slice in the testing group, thus resulting in average indicators with standard deviations. The proposed outline will be the process taken to acquire the results in the next chapter. Table 4.3 gives a description of the different methods shown.

Table 4.3: Methods tested in the results

Method	Feature Space Description	Feature space n
Large Patch size	5x5x5 voxel intensity patch centred around voxel of interest for all 4 MR image weightings	500
Single Voxel	Voxel intensity of interested for all 4 MR image weightings	4
Single Voxel and σ	Voxel intensity and σ for 3x3x3 voxel patch centred at voxel of interest for all 4 MR image weightings	8

Chapter 5

Numerical Results

The results chapter is split into three parts; finding optimal Dictionary Parameters for the newly proposed methods, segmentation results on images without noise present, and segmentation results on images with noise present. The method was implemented on the super computer network at York University. The network has 16 core processors with 128 GBytes of RAM. All computational times are based off this server. The method was programmed in MATLAB using the SPAMS package [26] for the LASSO and dictionary training.

5.1 Dictionary Parameters

The first result comes from exploring the parameters of dictionary learning. We questioned which dictionary width (m) and sparsity parameter (λ) would give the best results for the classifier given the different feature spaces. This was determined by evaluating the two dimensional domain of dictionary width versus lambda. For each method we took fifteen slices spread evenly across all three lesion load scans and evaluated the DSC of the lesion map for each to get an average DSC. Then we plotted the contour map to observe the parameters that lead to the highest average DSC.

Figure 5.1 shows the contour map for the single voxel method. From what we can see there is a large domain for which the classifier gets a average DSC of 1 for the lesion map. This is in the range of $m = 10$ to $m = 100$ and $\lambda = 0.9$ and 0.95. For the domain, we evaluated the performance of healthy tissue to determine the optimal parameters for all classes. This resulted in $m = 100$, and $\lambda = 0.95$.
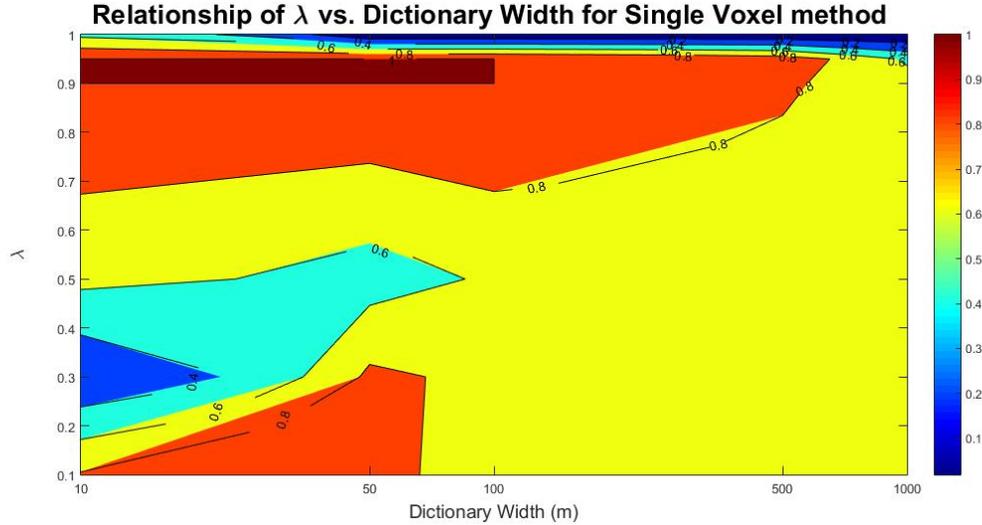
Figure 5.1: Contour map of lesion DSC with variation in dictionary width and λ for the single voxel method

Figure 5.2 shows the contour map found for the single voxel and σ method. The experiment was performed in the same manner as the single voxel method. The optimal parameters in this situation are $m = 10$, and $\lambda = 0.9$. This leads to a conclusion that with including the standard deviation in the feature space, the dictionaries do not require as many examples to identify incoming test subjects. The smaller λ parameter also shows that for classification it does not need to use as many examples as the single voxel method to identify the test subject.
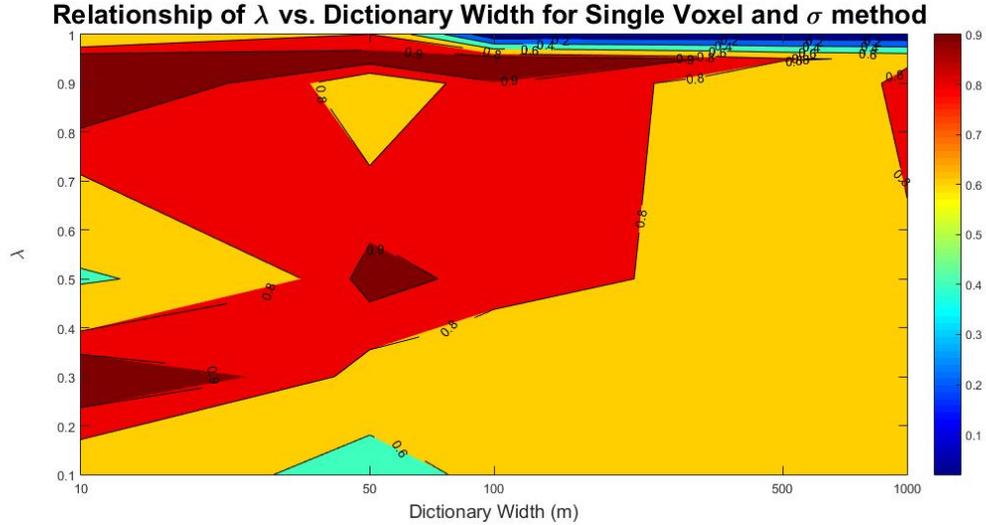
Figure 5.2: Contour map of lesion DSC with variation in dictionary width and λ for the single voxel and σ method

5.2 Segmentation without Noise

The following section evaluated the segmentation results on MR images without noise present. This simulation was performed 3 times; separately on each lesion load. A collection of fifteen slices were selected for the testing group taken from one lesion load scan. Then the next scan was evaluated and so on. The results for each lesion load will be presented and at the end, the final results were combined for an overall evaluation of the techniques.

5.2.1 Mild Lesion Load

For this simulation, fifteen slices from the mild lesion load Brainweb scans were used as test subjects. The other two lesion load scans were used for the training set. The slices were compared to the ground truth to acquire the confusion matrix components and the indicators listed in Chapter 4 were calculated for each tissue map for each method. The average of each indicator result is facilitated in Table 5.1.

For the lesion class, we can observe that the single voxel method out performs the other two. Table 5.2 gives the average DSC over the 15 test slices, the standard deviation of DSC. The time it takes to train the system and the average time it takes to test one slice are provided in the last two columns. The computational times are contrastingly different. The large patch size as presumed takes the longest amount of time to train and to test. The single voxel method takes longer to train than the single voxel and σ method, while testing times are opposite. We could explain this by observing the dictionary width for single voxel is larger than single voxel and σ, creating a longer training time. On the other hand, the feature space of single voxel and σ is larger than the single voxel, causing a longer testing time.

Figures 5.3 and 5.4 give a visual observation of each segmentation technique. These images show the increased amount of false positives given the larger patch size. As suggested in Figure 5.4, the larger patch size inaccurately recreates the slice that has no lesion tissue present.

Table 5.1: Performance indicators for each method with mild lesion load without Noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.906	0.993	0.966	0.943
	WM	0.974	0.980	0.980	0.916
	CSF	0.969	0.988	0.987	0.902
	LES	**0.895**	**0.999**	**0.999**	**0.358**
Single Voxel	GM	0.878	0.976	0.946	0.910
	WM	0.901	0.977	0.968	0.869
	CSF	0.927	0.981	0.978	0.843
	LES	**0.947**	**1.000**	**0.999**	**0.956**
Single Voxel and σ	GM	0.982	0.987	0.935	0.885
	WM	0.995	0.951	0.956	0.841
	CSF	0.873	0.986	0.978	0.837
	LES	**0.933**	**1.000**	**0.999**	**0.930**

The mild lesion load scans are reflective of patient's who have early stage disease. The lesion targets are small, and some slices have no lesions at all. In evaluation of the large patch method, the larger patches result in a higher amount of false positives.

40

Table 5.2: Performance of each method on lesion tissue map and computational time of each method for mild lesion load with no noise

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.358	0.321	3949s	11347s
Single Voxel	0.956	0.172	20.8s	59s
Single Voxel and σ	0.930	0.257	7.3s	79.8s

Therefore, the performance goes down. While the single voxel and single voxel with σ methods are more precise at detecting the smaller lesions. This leads to the conclusion that the proposed method would be more accurate at detecting early stage lesions.
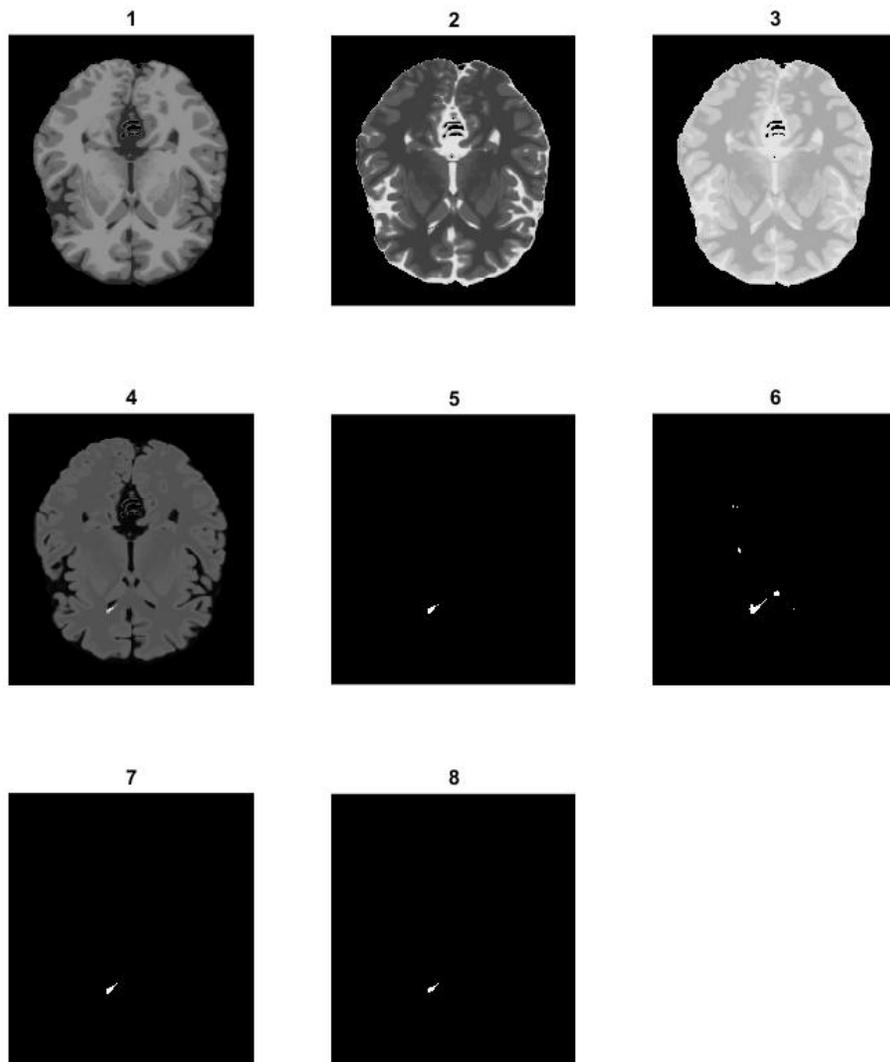
Figure 5.3: Visual look at each method for mild lesion load with no noise example 1: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation
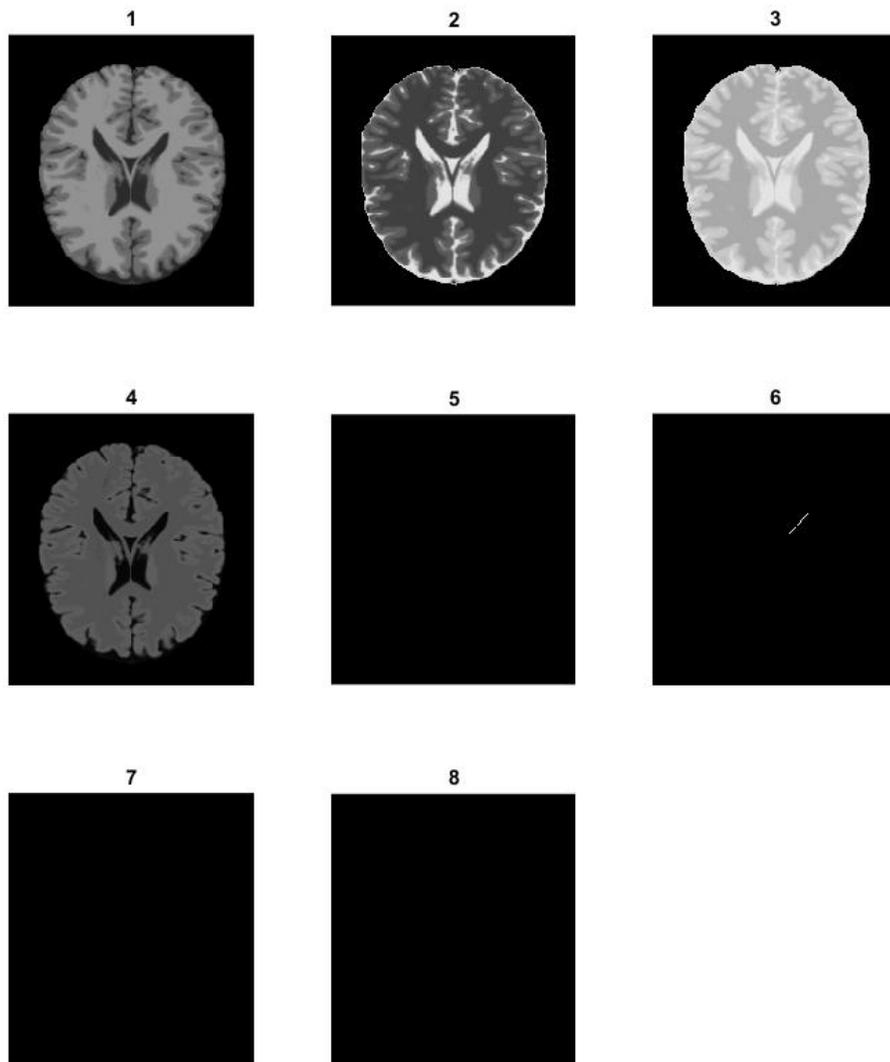
Figure 5.4: Visual look at each method for mild lesion load with no noise example 2: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation

5.2.2 Moderate Lesion Load

Here we present the results of evaluating each method on the moderate lesion load scan with no noise. The same method as previously was used, taking 15 slices from the moderate lesion load scan for testing and the rest used for the training set. Then feature extraction was executed in the three described methods before training and testing the classifier. Then the performance indicators were calculated. Table 5.3 gives the average calculated indicators for each tissue group and for each method. Then table 5.4 is a summary of each method's performance on the lesion maps and the computational times. Figure 5.5 and 5.6 are two visual examples from these results.

Table 5.3: Performance indicators for each method for moderate lesion load with no noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.903	0.993	0.965	0.941
	WM	0.968	0.979	0.978	0.909
	CSF	0.965	0.989	0.987	0.905
	LES	**0.923**	**0.998**	**0.998**	**0.531**
Single Voxel	GM	0.829	0.932	0.942	0.897
	WM	0.978	0.968	0.970	0.882
	CSF	0.951	0.973	0.972	0.813
	LES	**1.000**	**1.000**	**1.0000**	**1.000**
Single Voxel and σ	GM	0.812	0.991	0.936	0.887
	WM	0.996	0.949	0.955	0.835
	CSF	0.913	0.986	0.981	0.862
	LES	**0.990**	**1.000**	**0.999**	**0.993**

Table 5.4: Performance of each method on lesion tissue map and computational time of each method for moderate lesion load with no noise

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.531	0.161	4011s	10941s
Single Voxel	1.000	0.001	20.3s	59s
Single Voxel and σ	0.993	0.020	7.3s	78.1s

We can observe with these results that each method improved with the increase in lesion load, leading to a suggestion that all methods perform better with later stage disease. These results are satisfactory since a physician would also find larger lesions easier to segment manually. This suggests that the methods would be useful in analyzing disease progression.
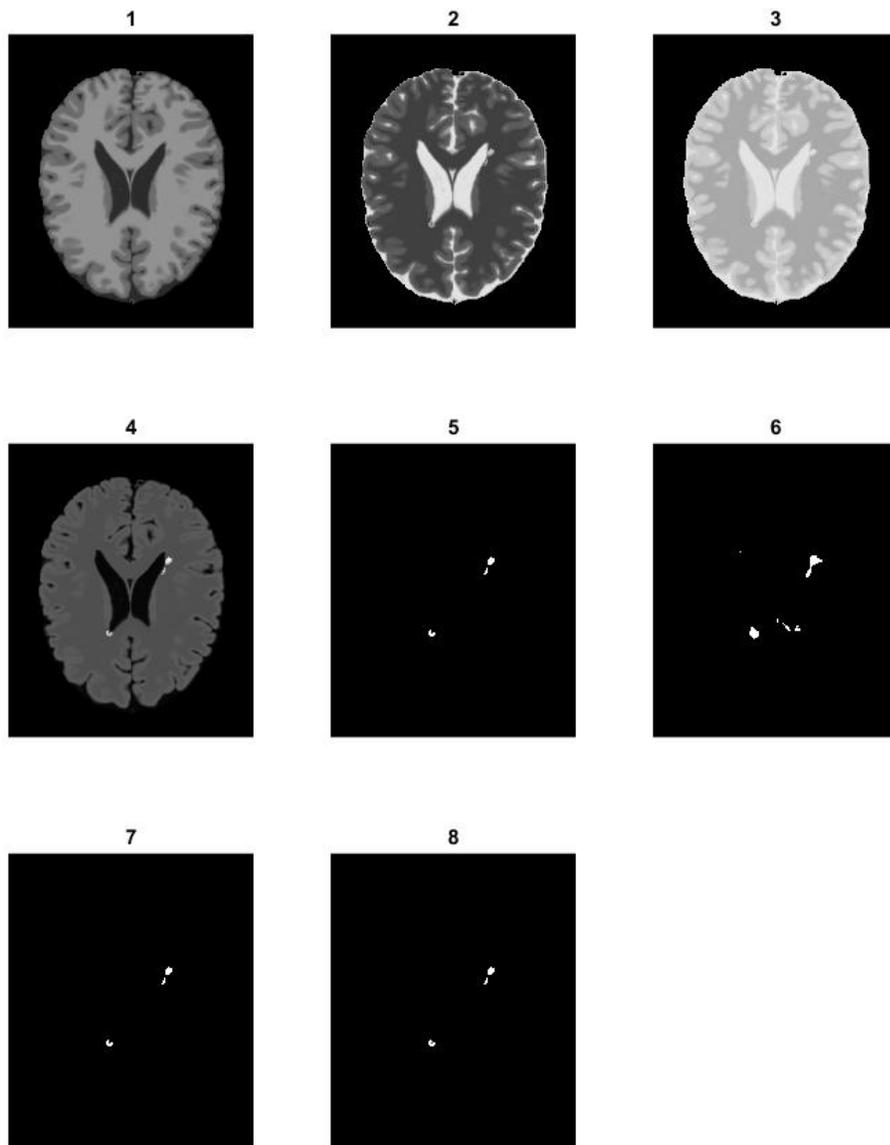
Figure 5.5: Visual look at each method for moderate lesion load with no noise example 1: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation
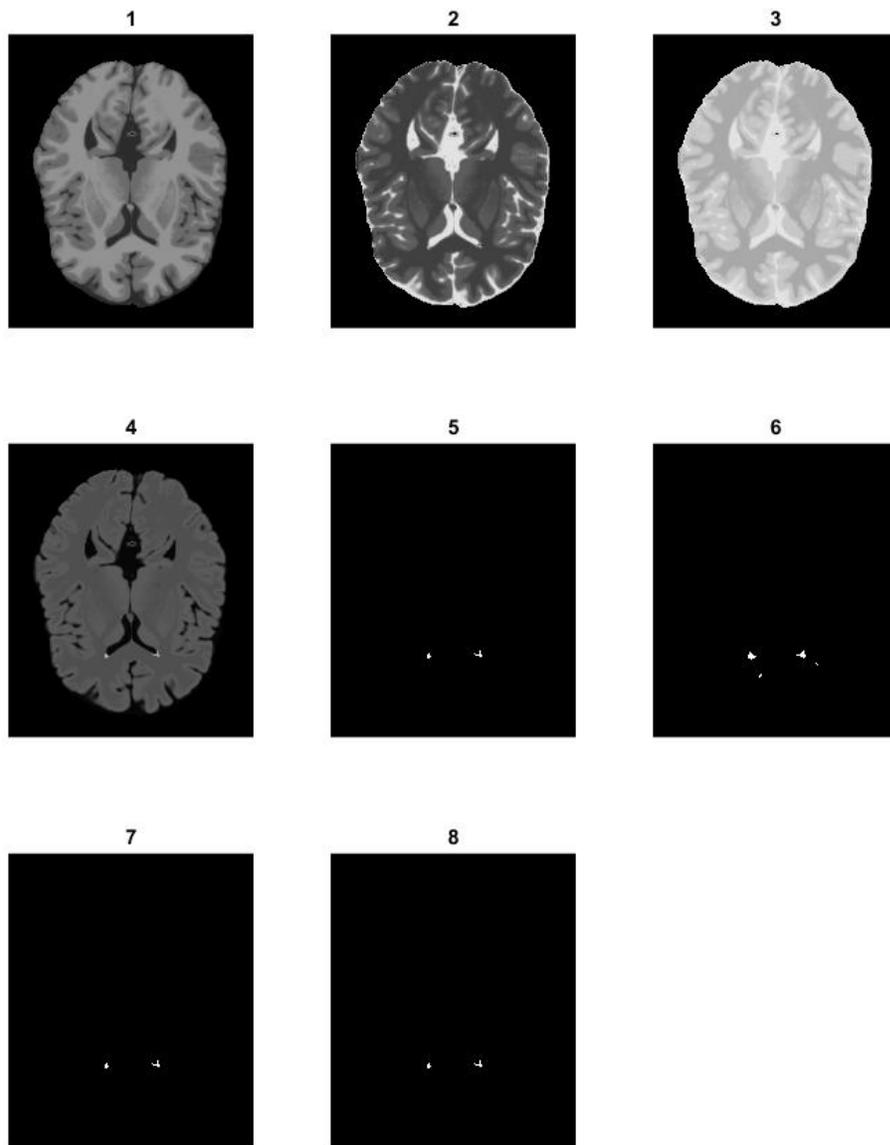
Figure 5.6: Visual look at each method for moderate lesion load with no noise example 2: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted Segmentation

5.2.3 Severe Lesion Load

For the final simulation of images without noise we used the severe lesion load for testing and the results were very impressive. The average of indicators for the fifteen test slices are given in Table 5.5 and a summary of lesion performance in Table 5.6. We can observe that the performance level has not changed much from the moderate scans, leading to the idea that the smaller lesions are the most difficult to segment.

Figures 5.7 and 5.8 are two visual examples from the severe lesion loads. By comparing the predicted maps to the ground truth maps, the larger patch size tends to over segment the boarders, resulting in increased false positives. This can be explained by the larger spatial feature space. The large patch size lacks precision to clearly delineate the lesions from the healthy tissue.

Table 5.5: Calculated performance indicators for each method for severe lesion load with no noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.900	0.994	0.965	0.940
	WM	0.943	0.980	0.975	0.896
	CSF	0.964	0.988	0.987	0.903
	LES	**0.996**	**0.993**	**0.993**	**0.513**
Single Voxel	GM	0.801	0.993	0.934	0.882
	WM	0.983	0.960	0.962	0.855
	CSF	0.955	0.973	0.972	0.812
	LES	**1.000**	**1.000**	**1.000**	**1.000**
Single Voxel and σ	GM	0.809	0.991	0.935	0.885
	WM	0.996	0.948	0.954	0.831
	CSF	0.915	0.986	0.981	0.863
	LES	**0.998**	**1.000**	**1.000**	**0.977**

Table 5.6: Performance of each method on lesion tissue map and computational time of each method for severe lesion load with no noise

| Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.513	0.108	4663s	14056s
Single Voxel	1.000	0.0002	23.6s	61.5s
Single Voxel and σ	0.977	0.035	8.2s	83.7s
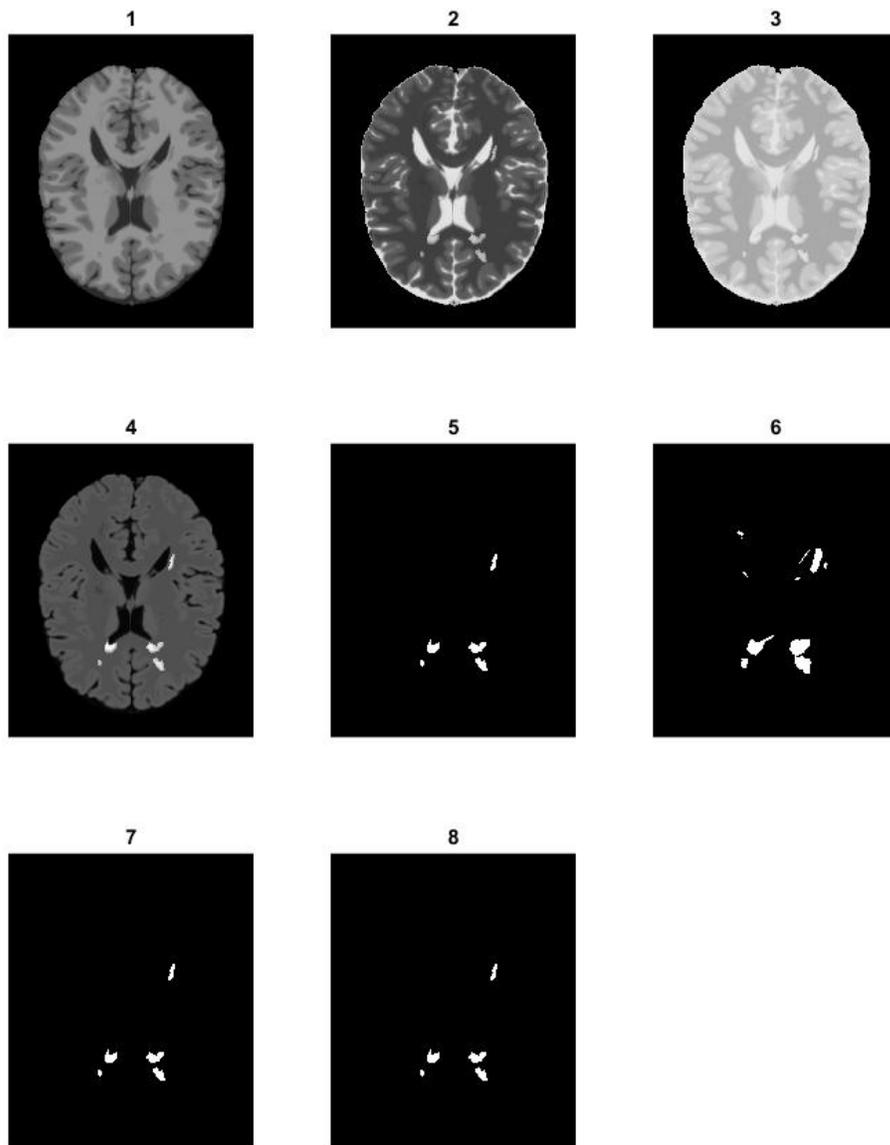
Figure 5.7: Visual look at each method for severe lesion load with no noise example 1: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation
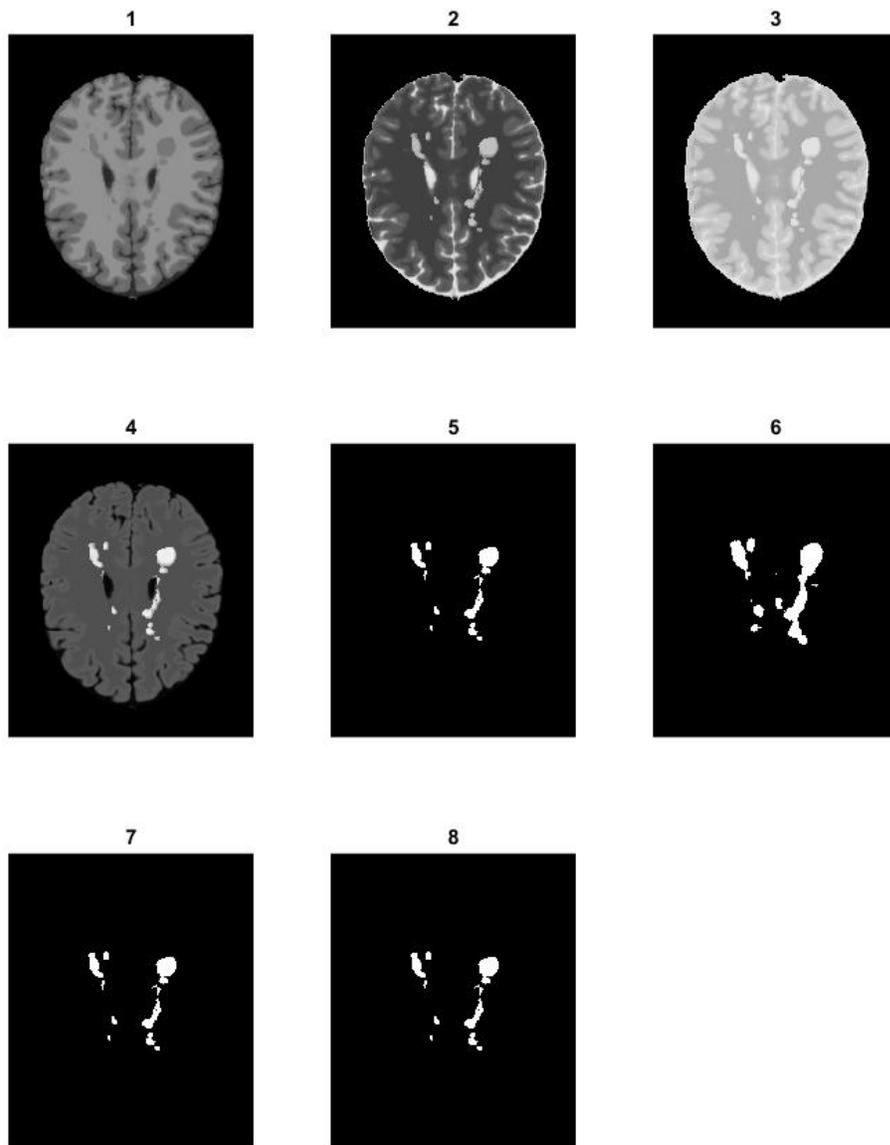
Figure 5.8: Visual look at each method for severe lesion load with no noise example 2: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation

5.2.4 Segmentation without Noise Overall Performance

Table 5.7 gives the overall result of the methods used on all 45 slices from the collection of all lesion load scans. We can see that the proposed method is more efficient at segmenting the MS lesions compared to using a large patch size, as previously suggested [13] [14] [36].

Table 5.7: Over all performance of each method on lesion tissue map and computational time of each method

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.467	0.226	4208s	12115s
Single Voxel	0.985	0.099	21.6s	59.8s
Single Voxel and σ	0.967	0.150	7.6s	80.5s

In looking at each lesion load scan individually we have shown a consistent efficacy of the proposed method. The ability of the algorithm to manage segmentation of different lesion loads shows a promise for clinical usage. Brainweb is a simulated patient and is an idealized situation for testing the algorithm. Working with a data set of real patient data should be evaluated.

The improvement could be attributed to the fact that the dictionary is now learning the intensity differences between MR image weightings for each class, as opposed to learning the texture difference between MR image weightings. Therefore, it is learning the different relaxation times of the lesion tissue and healthy tissue. The complexity of encompassing the texture difference between tissues seems to confuse the DL classifier as opposed to clarify the problem. Leading to the idea that learning texture quality from large patches of voxel intensity and MR multi-weighted values together, complicates the solution as opposed to aiding it. Although some form of texture quantification must be used to help in distinguishing tissues in the presence of noise, we see how evidently this is needed in the next section.

5.3 Segmentation with Noise

The following section will evaluate the robustness of the methods with varying levels of noise added to the images. We will look at each lesion load scan individually and show within each subsection the results of having 3% noise and 9% noise added to the image. Noise is generated within the real and imaginary parts of the relaxation times using a pseudo random Gaussian noise field [9]. The noise percentage is relative to the average brightness values from the phantom. The tissue dictionaries are trained with the noise level present to show it's ability to use the noise quality.

5.3.1 Mild Lesion Load with Noise

From looking at fifteen slices from the mild lesion load scan with noise present we can see the methods break down and are unable to preform as accurately as before. Table 5.8 gives the indicators for 3% noise added to the images and Table 5.9 gives the indicators for 9% added to the image. Tables 5.10 and 5.11 give a summary of each method's performance on MS lesion segmentation. Figure 5.9 shows a visual example from the 3% noise images and Figure 5.10 is a visual example of images with 9% noise.

In the terms of the mild lesion load the large patch based method is the most consistent and appropriate method to use. We observe that the single voxel method completely fails to delineate any important segmentations from the slice. We presumed that this would occur given the lack of texture quantification in evaluating a single voxel. The noise alters the voxel intensity directly and changes the pattern of relaxation time magnitude between weightings, making it more confusing to delineate one tissue from another. The different classes will have more overlap of voxel intensity quality compared to no noise. Therefore, some texture quantification is needed.

We observe that the standard deviation of neighboring voxels is adequate enough to compete with the large patch based method. The 3% noise simulation for single voxel and σ method got a DSC=0.856. Proving the adequacy of this method in the presence of slightly noisy images. Unfortunately, with a noise increase to 9% the method falls to a similar DSC of the large patch method of 0.348. We can explain this by the complexity of finding smaller lesions. A physician will request de-noising of the MR images before evaluating the scan, and even then smaller lesions would be more difficult to delineate from healthy tissues. De-noising may be a proper tactic before using DL for MS lesion segmentation.

Another suggestion to aid in the segmentation of smaller lesions would be to use post-processing tasks. In Figure 5.10 panel 8 we can observe that the false positives are a few scattered single voxel segments. Post-processing encompasses taking the results of the DL classifier and applying image processing methods to "clean up" the segmentation. These single voxel segmentations could be considered as inadequate and we could delete any segmentations that are single voxels, thus leaving behind clusters of more than one voxel. This would improve the DSC of the single voxel and σ method in the presence of 9% noise of mild lesion load scans.

Table 5.8: Performance indicators for each method for mild lesion load with 3% noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.901	0.993	0.964	0.940
	WM	0.975	0.978	0.978	0.908
	CSF	0.967	0.988	0.987	0.904
	LES	**0.6482**	**0.999**	**0.999**	**0.349**
Single Voxel	GM	0.948	0.821	0.860	0.804
	WM	0.043	0.999	0.883	0.082
	CSF	0.869	0.989	0.982	0.858
	LES	**0.613**	**0.995**	**0.995**	**0.068**
Single Voxel and σ	GM	0.842	0.977	0.935	0.889
	WM	0.951	0.954	0.954	0.827
	CSF	0.843	0.990	0.981	0.847
	LES	**0.854**	**1.000**	**1.000**	**0.856**

Table 5.9: Performance indicators for each method for mild lesion load with 9% noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.853	0.994	0.950	0.913
	WM	0.972	0.966	0.967	0.869
	CSF	0.981	0.983	0.983	0.882
	LES	**0.660**	**0.999**	**0.999**	**0.374**
Single Voxel	GM	0.760	0.911	0.864	0.773
	WM	0.317	0.976	0.896	0.423
	CSF	0.917	0.979	0.975	0.823
	LES	**0.667**	**0.940**	**0.940**	**0.008**
Single Voxel and σ	GM	0.694	0.952	0.872	0.769
	WM	0.793	0.906	0.892	0.633
	CSF	0.874	0.987	0.980	0.848
	LES	**0.653**	**0.999**	**0.999**	**0.348**

Table 5.10: Overall performance of each method on lesion tissue map and computational time of each method for mild lesion load with 3% noise

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.349	0.309	4236s	13517s
Single Voxel	0.068	0.062	21.8s	58.9s
Single Voxel and σ	0.856	0.315	7.8s	84.4s

Table 5.11: Overall performance of each method on lesion tissue map and computational time of each method for mild lesion load with 9% noise

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.374	0.307	4110.1s	16911.0s
Single Voxel	0.008	0.007	29.1s	73.1s
Single Voxel and σ	0.348	0.269	7.5s	85.9s
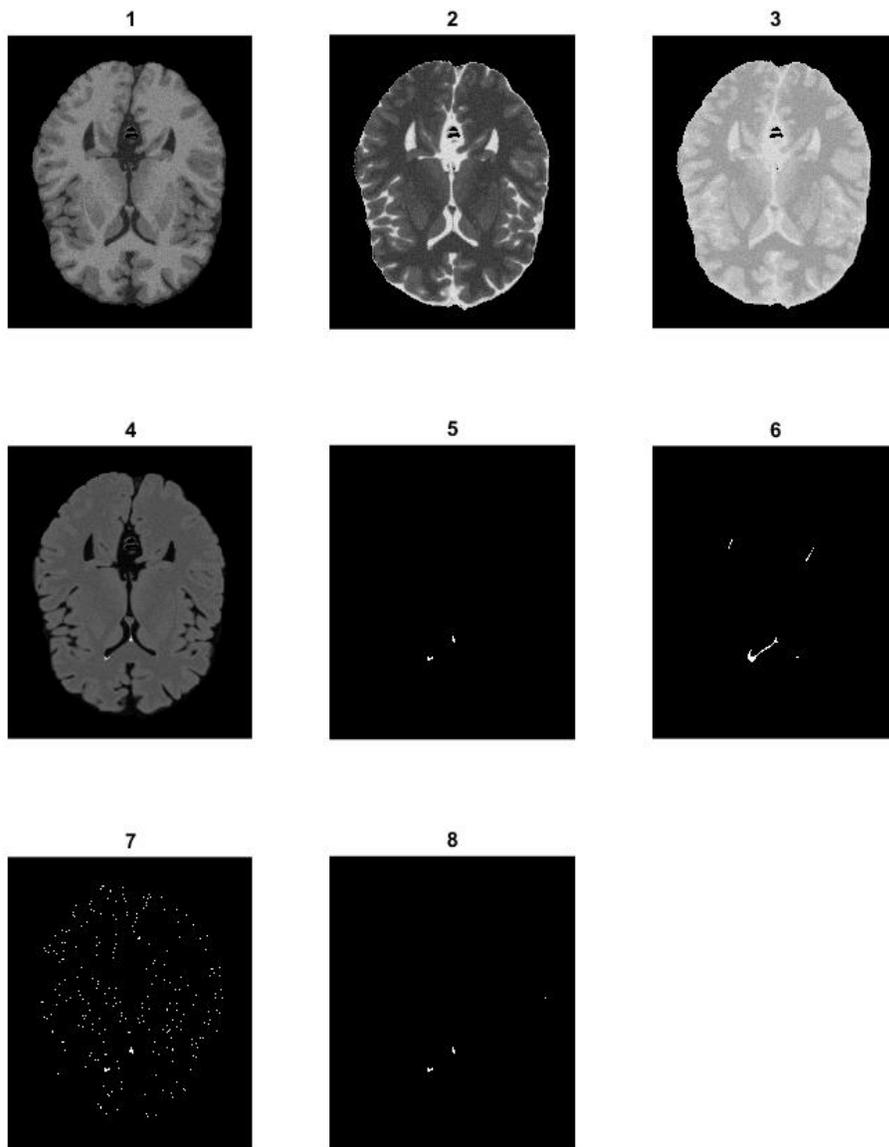
Figure 5.9: Visual look at each method for mild lesion load with 3% noise: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation
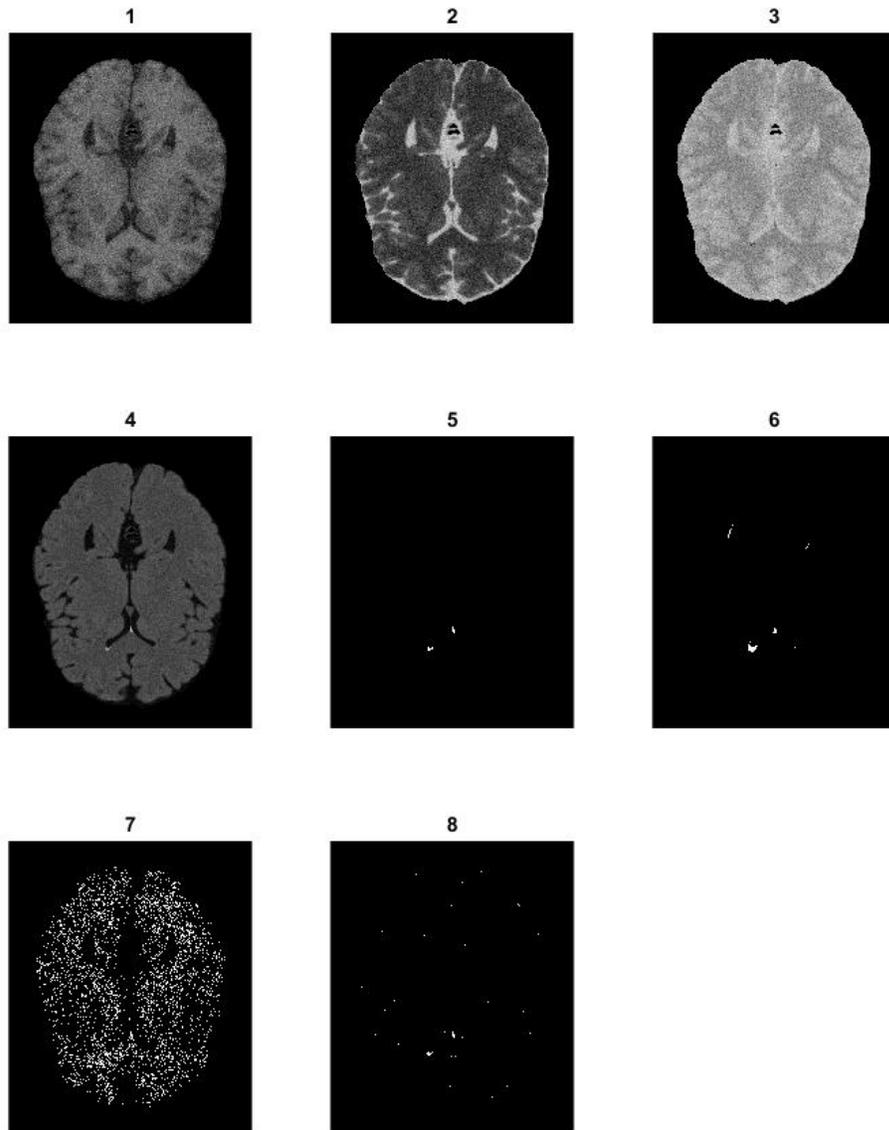
Figure 5.10: Visual look at each method for mild lesion load with 9% noise: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation

5.3.2   Moderate Lesion Load with Noise

The Moderate lesion load scans with noise had a differing outcome from the mild lesion load scan. Tables 5.12 to 5.14 gives the same calculated indicators as the previous sections and then a review of lesion segmentation. Figure 15.11 is a visual example of 3% noise added to the images and Figure 5.12 is a visual example of 9% noise added to the images.

Remarkably the 3% noise level images redeemed their efficacy. Showing that with larger lesions present it is easier to find them in the presence of noise. This could be explained by the fact that smaller lesions are cloaked by the noise and make them harder to find, therefore the moderate lesion load will have a better performance.

The 9% noise in the images resulted in inadequate results for the single voxel method. This further supports the need for noise reduction before segmentation is performed. This is not an unreasonable requirement since most radiologists will request noise reduction before manual segmentation.

Table 5.12: Performance indicators for each method for moderate lesion load with 3% noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.902	0.993	0.965	0.941
	WM	0.966	0.979	0.977	0.906
	CSF	0.969	0.988	0.987	0.904
	LES	**0.914**	**0.998**	**0.998**	**0.521**
Single Voxel	GM	0.936	0.822	0.856	0.798
	WM	0.006	1.000	0.880	0.013
	CSF	0.932	0.979	0.976	0.833
	LES	**0.982**	**1.000**	**1.000**	**0.985**
Single Voxel and σ	GM	0.817	0.985	0.933	0.884
	WM	0.979	0.947	0.950	0.819
	CSF	0.882	0.990	0.983	0.867
	LES	**0.986**	**1.000**	**1.000**	**0.984**

Table 5.13: Performance indicators for each method for moderate lesion load with 9% noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.826	0.989	0.938	0.892
	WM	0.978	0.945	0.948	0.813
	CSF	0.893	0.995	0.988	0.908
	LES	**0.929**	**0.997**	**0.997**	**0.477**
Single Voxel	GM	0.800	0.908	0.875	0.796
	WM	0.583	0.944	0.900	0.579
	CSF	0.785	0.994	0.980	0.825
	LES	**0.977**	**0.993**	**0.993**	**0.273**
Single Voxel and σ	GM	0.752	0.9930	0.875	0.787
	WM	0.772	0.914	0.897	0.636
	CSF	0.670	0.998	0.977	0.789
	LES	**0.942**	**1.000**	**1.000**	**0.961**

Table 5.14: Overall performance of each method on lesion tissue map and computational time of each method for moderate lesion load with 3% noise

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.521	0.175	4890s	13745s
Single Voxel	0.985	0.022	24.7s	64.8s
Single Voxel and σ	0.984	0.021	7.2s	91.5s

Table 5.15: Overall performance of each method on lesion tissue map and computational time of each method for moderate lesion load with 9% noise

| Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.477	0.166	4037.1s	14593s
Single Voxel	0.273	0.198	29.4s	70.1s
Single Voxel and σ	0.961	0.067	7.7s	81.5s
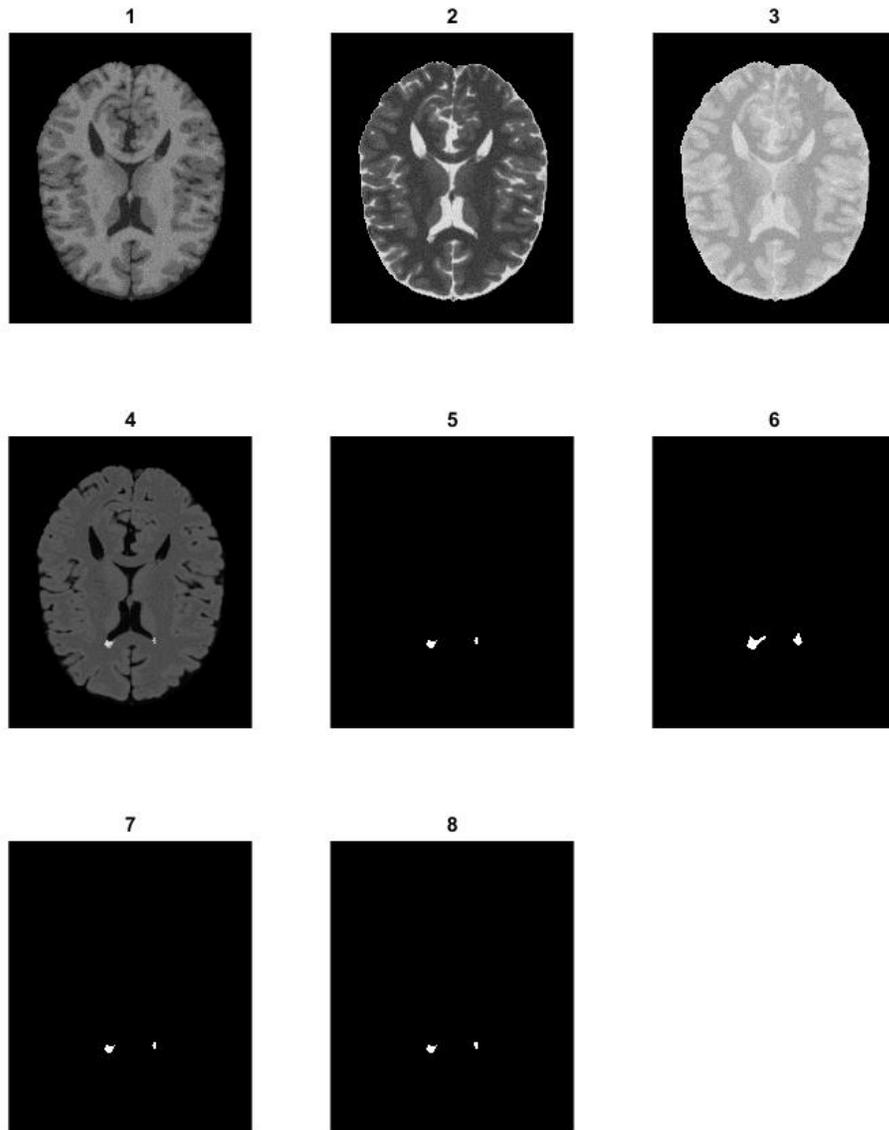
Figure 5.11: Visual look at each method for mild lesion load with 3% noise: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted S]segmentation
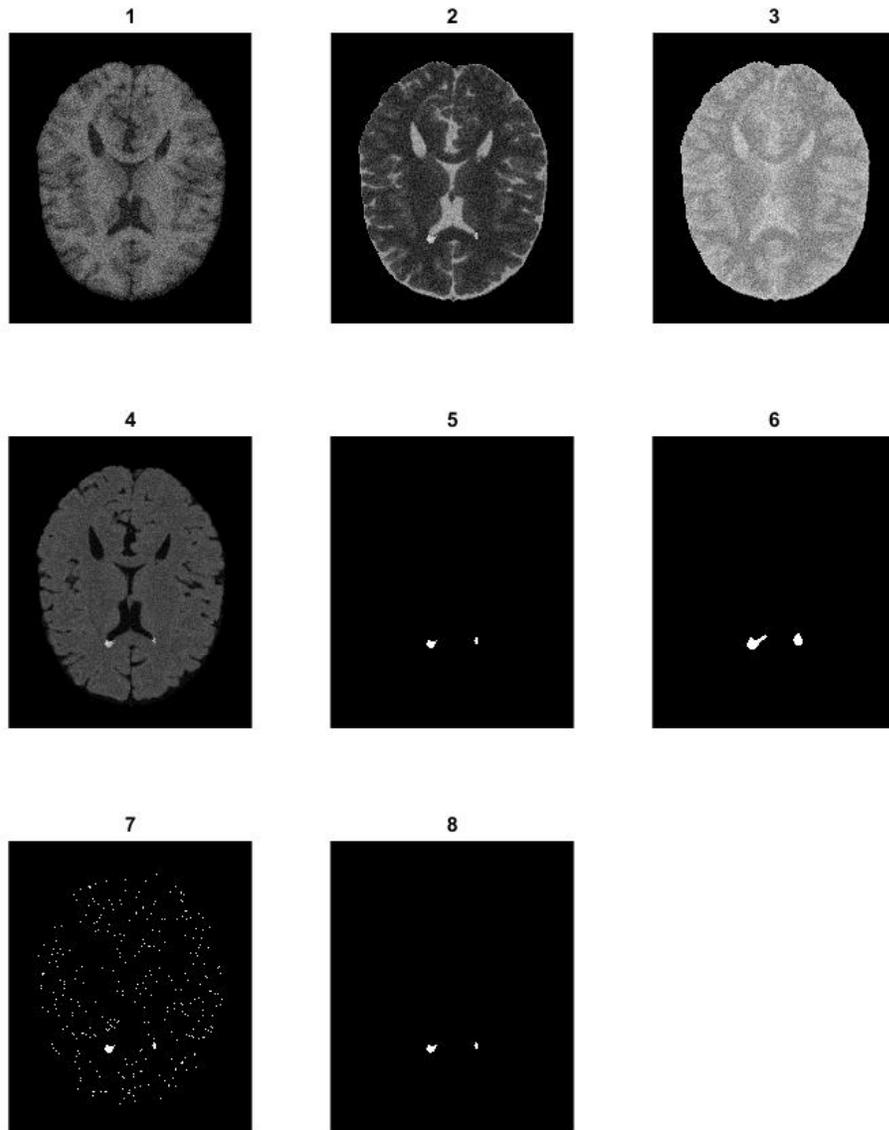
Figure 5.12: Visual look at each method for mild lesion load with 9% noise: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted S]segmentation

5.3.3   Severe Lesion Load with Noise

The final simulation is performed using the severe lesion load scan with noise present. In the same manner as the previous sections Tables 5.16 to 5.19 give the indicator results for each method under each noise level. The latter two tables give a summary of performance on lesion load. Figures 5.13 and 5.14 are two visual examples of 3% and 9% noise images respectively.

As we have evaluated each lesion load progressively in the presence of noise there has been a pattern, the more abundant and increased size of the lesions the easier it is to automatically segment them. This simulation on the severe lesion load concretes that fact. Here we observe the single voxel method competitively keeping up with the single voxel and σ method contrasting from previous simulations with noise in the image. This supports the interpretation that the segmentation algorithm would be adequate for calculating total lesion volume and aiding in the evaluation of disease progression.

Table 5.16: Performance indicators for each method for severe lesion load with 3% noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.891	0.9995	0.963	0.936
	WM	0.947	0.978	0.974	0.890
	CSF	0.970	0.987	0.986	0.898
	LES	**0.996**	**0.993**	**0.993**	**0.521**
Single Voxel	GM	0.937	0.846	0.873	0.818
	WM	0.136	0.996	0.895	0.233
	CSF	0.906	0.984	0.979	0.846
	LES	**0.998**	**1.000**	**1.000**	**0.910**
Single Voxel and σ	GM	0.789	0.986	0.925	0.867
	WM	0.986	0.939	0.944	0.799
	CSF	0.869	0.988	0.980	0.849
	LES	**0.999**	**1.000**	**1.000**	**0.994**

Table 5.17: Performance indicators for each method for severe lesion load with 9% noise

Method	Tissue	Sens	Spec	Acc	DSC
Large patch size	GM	0.848	0.995	0.949	0.912
	WM	0.946	0.968	0.965	0.859
	CSF	0.987	0.981	0.981	0.869
	LES	**0.989**	**0.995**	**0.995**	**0.622**
Single Voxel	GM	0.768	0.920	0.873	0.787
	WM	0.552	0.953	0.906	0.575
	CSF	0.967	0.968	0.968	0.792
	LES	**0.827**	**1.000**	**0.999**	**0.899**
Single Voxel and σ	GM	0.661	0.957	0.865	0.751
	WM	0.839	0.892	0.885	0.626
	CSF	0.843	0.989	0.980	0.844
	LES	**0.869**	**1.000**	**0.999**	**0.929**

Table 5.18: Overall performance of each method on lesion tissue map and computational time of each method for severe lesion load with 3% noise

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.521	0.106	5339s	16166s
Single Voxel	0.910	0.082	23.3s	62.9s
Single Voxel and σ	0.994	0.011	7.1s	85.0s

Table 5.19: Overall performance of each method on lesion tissue map and computational time of each method for severe lesion load with 9% noise

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.622	0.080	4593.0s	18714.0s
Single Voxel	0.899	0.051	21.9s	65.7s
Single Voxel and σ	0.929	0.038	7.2s	89.0s
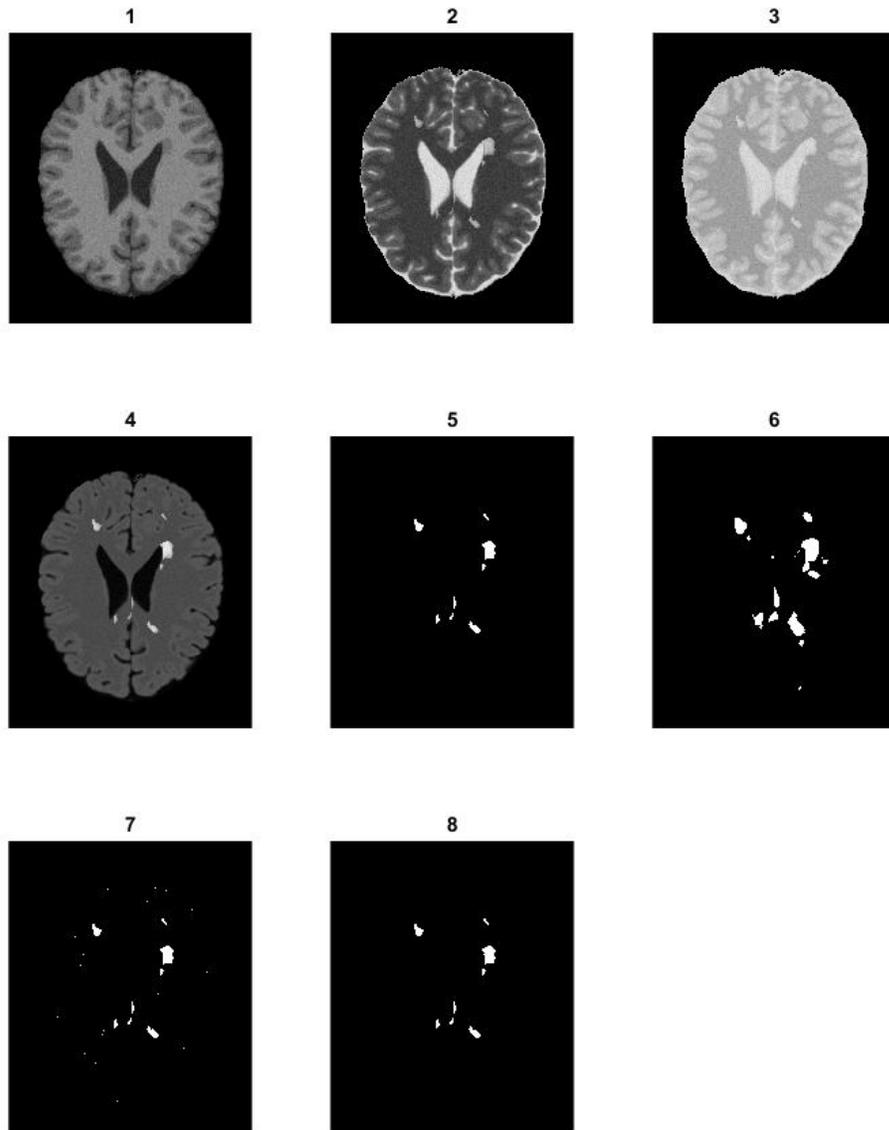
Figure 5.13: Visual look at each method for severe lesion load with 3% noise: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation
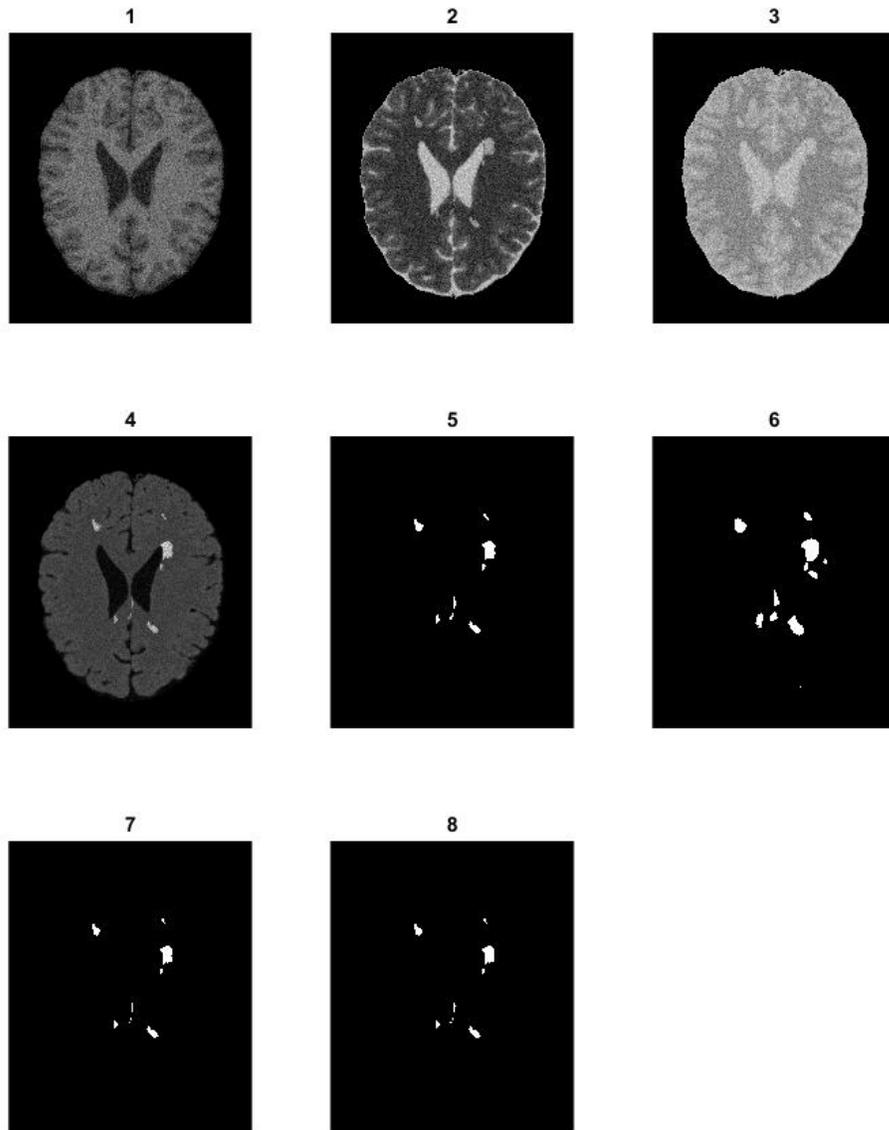
Figure 5.14: Visual look at each method for severe lesion load with 9% noise: 1) T1w, 2) T2w, 3) PDw, 4) FLAIR, 5) Ground truth lesion segmentation, 6) Large patch predicted segmentation, 7) Single voxel predicted segmentation, 8) Single voxel and σ predicted segmentation

5.3.4 Segmentation with Noise Overall Performance

It is in the evaluation of all three lesion loads overall that we can view the performance of these algorithms in the case of distorted data sets. Table 5.20 gives the average lesion map performance across the three scan simulations. We can see that the single voxel and σ method out performs the other two with the most consistent standard deviation. It is also observed in Table 5.21 that the single voxel and σ method stands above the rest in the presence of 9% noise. Figure 5.15 is a graphing of the DSC of each method with respect to noise level added to the images.

Table 5.20: Overall performance of each method on lesion tissue map and computational time of each method on images with 3% noise present

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.464	0.197	4821.7s	14476s
Single Voxel	0.654	0.055	23.2s	62.2s
Single Voxel and σ	0.945	0.011	7.4s	87.0s

Table 5.21: Overall performance of each method on lesion tissue map and computational time of each method on images with 9% noise present

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Large patch size	0.491	0.184	4246.7s	16739.3s
Single Voxel	0.393	0.256	26.8s	69.6s
Single Voxel and σ	0.746	0.125	7.5s	85.5s

It is interesting to note that although the larger patch size doesn't perform as impressively as the single voxel and σ method, it is consistent across the noise levels evaluated. This suggested that the larger patch size would be more accurate at gaining an insight into the scan but not for getting a precisely calculated volume. Therefore, the larger patch size would need a post-processing check done by a human operator.

In the end, the new proposed method is the most accurate at segmenting MS lesion from multi-weighted MR imaging. The single voxel and σ feature space has
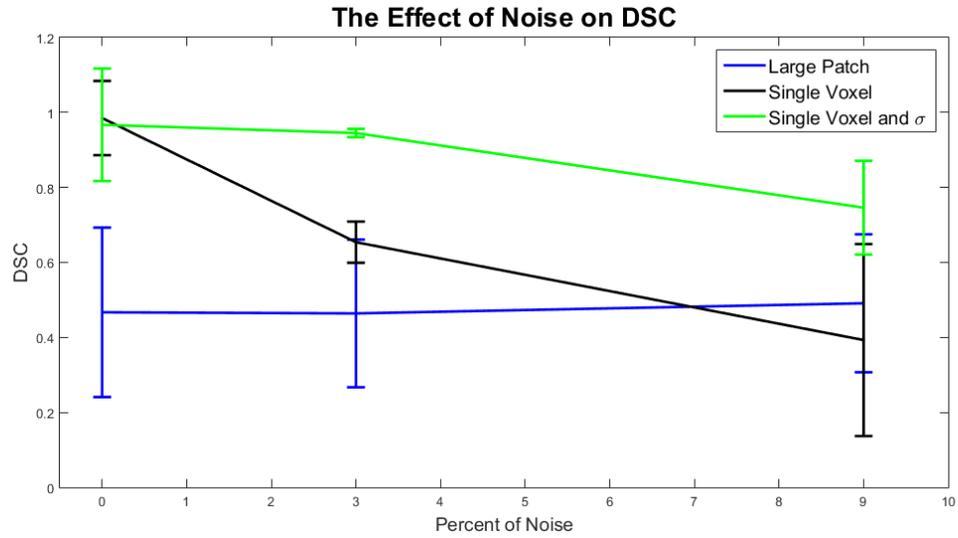
Figure 5.15: The effect of noise on DSC of the three presented methods

shown to adequately segment the lesions of the Brainweb simulated phantom. It even proves to be robust against MR image noise, which in some cases can not be avoided. The following section of this thesis is a discussion of future research objectives in critique of the presented method.

Chapter 6

Discussion

6.1 Healthy Tissue Performance

In examining the results there are two major trends that should be addressed: the efficacy of lesion segmentation versus healthy tissue and the lesion load variation. The DSC for the healthy tissues are given depending on the method in Table 6.1 in the absence of noise.

Table 6.1: Segmentation results of healthy tissues

Method	GM	WM	CSF
Large patch size	0.941	0.907	0.903
Single Voxel	0.896	0.869	0.823
Single Voxel and σ	0.886	0.836	0.854

From these results it becomes apparent that the larger patch size is more efficient than the single voxel and standard deviation classification at identifying healthy tissue. This could be explained by the localized texture contributing to the healthy tissue. The dictionary of healthy tissues in compact feature space do not hold enough texture quality to discriminate the tissues among each other. It could be suggested to increase standard deviation patch size in the third method or explore other statistical moments.

It leads to the suggestion that a varying patch size method may work better for segmentation of different classes. Deshpande [13] [14] had found that varying the amount of m-examples between dictionaries increased discrimination, but the proposed method suggests using different feature space for each dictionary may increase discrimination. Therefore, a future initiative is to evaluate the dictionaries working in different feature spaces.

The second interpretation is based on the varying slices used for testing and their outcomes. The singleton intensity feature space was more effective at identifying small lesions or clarifying that no lesion was present. While the large patch size was more effective at identifying large sclerotic lesions and could be used to evaluate total lesion volume of late stage disease. This can be explained by the ability for the patch size to find a given lesion of a certain size. Large patches find large lesions, small patches identify small lesions.

This leads to the proposed method being more effective at identifying earlier lesions. The method should be tested on patient's with clinically isolated syndrome (CIS). CIS is a pre-clinical disease to MS. Most patients must show up with two CIS cases before being diagnosed with MS [11]. This technique could be useful in early MS diagnosis.

A final interesting idea would be to use both techniques combined in some manner to achieve a more beneficial segmentation algorithm for lesions as well as healthy tissue. In considering all dictionaries in each feature space the residual could be combined in some manner to achieve a more desirable outcome. Future investigations are still needed to perfect this technique.

6.2 Image Texture Analysis

The initial directive of this study was to find an adequate texture quality to use for segmentation [37]. Recent publications have shown an MR image textural difference between MS lesions and WM [38] [24]. Therefore, we investigated three main techniques of image texture analysis and their usage as a new feature space used in DL. We will give a small description of how we used each and give some results that showed their lack of performance. It was by looking at these feature spaces that further investigation could result in a more efficient means of automatic segmentation.

We first explored the interpretation of image texture as frequency-spatial information. We explored multiple methods such as short time FT, short time FT using a Gaussian window [10], and local frequency from the Gaussian windowed short time FT [38]. Unfortunately, the latter two methods were too computationally taxing to acquire results and took much longer than the large patch method and were therefore abandoned. Instead, we evaluated the short time FT by taking a spatial patch of 5x5x5 of the voxel in question and perform the 3D-Fast FT to find the local frequency information. We then would continue to train the dictionaries on this information for segmentation. This method of feature extraction became very time consuming. The intensive computational times for these methods were not beneficial to use given the resulting segmentation. An example of results are given in table 6.2 at the end of this section and were abandoned in the process.

The second attempt was to use the texture quality for the GLCM [20]. In this method we took the 5x5x5 spatial patch and calculated the GLCM at all angles (26 directions) with a distance of considering adjacent voxels. We used a resolution of counting ten gray levels across the spectrum resulting in a 10x10 GLCM. Then this matrix was concatenated into vectors and used for training the dictionary. This quantification technique was competitively quicker than FFT but also did not result

in a more beneficial segmentation algorithm. The results of this are included in Table 6.2.

The final attempt that lead to the proposed method was to use a more simpler method of texture quantification called statistical moments [29]. This method encompasses taking a histogram of the gray levels present in a patch and then calculating statistical moments of the probability distribution in the histogram. We explored using five statistical moments the mean, standard deviation, skewness, kurtosis, and entropy. For these methods we used a 3x3x3 patch, only incorporating adjacent voxels, and trained the dictionary with the varying levels of each. It was then that we found the standard deviation the most beneficial and resulted in the proposed method of this paper.

It is in bringing up these failed attempts that we wish to argue that the texture quantification that is adequate for segmentation has not yet been found. Further investigation of these techniques could adequately result in better segmentation. The method of using dictionaries is robust in that it can handle learning complex systems. By doing so, learning multiple complex systems in a coordinated method could result in a powerful segmentation algorithm. Therefore, further investigation into using texture analysis for segmentation is needed.

Table 6.2 gives some previous results based on different feature spaces that were evaluated. Each of these methods have $m = 1000$ examples in their dictionaries with a sparsity parameter of $\lambda = 0.95$. These parameters may not be optimal for each method, as we have seen in the results Section 5.1 for the proposed methods. For the evaluation of each of these feature spaces, the parameters are adequate to show their capability of segmenting MS lesions comparatively to each other.

Table 6.2: Texture quantification techniques explored for SM lesion segmentation and their results

Method	Average LES DSC	St.Dev. DSC	Training Time	Testing Time
Short Time FT	0.096	0.146	7359.7s	43.2s
GLCM	0.162	0.174	4794.1s	36605s
Mean	0.311	0.225	1147.5s	17.7s
Standard Deviation	0.231	0.180	1213.6s	30.0s
Skewness	0.060	0.081	1118.6s	28.8s
Kurtosis	0.033	0.047	1157.4s	28.1s
Entropy	0.096	0.146	1160.1s	43.2s
Single voxel and Mean	0.819	0.342	285.8s	23.2s

6.3   Data Sets

The resulting data of the Brainweb set is not normalized to a standard gray scale. The data is comprised of different relaxation times for the given sample at the localized position. For example, T1w images are the measure of T1 relaxation times of the localized molecules. T1 relaxation time is the amount of time a nuclear magnetic vector takes to return to the ground state magnetic field of the MR scanner. Therefore, in the proposed pipeline all T1 relaxation measurements are standardized to a uniform gray scale ranging between 0 and 1.

In medical imaging, a common practice is to limit the amount of visible T1 measurements in the photo and to change the level of visible T1 measurements to increase discrimination of different tissues. This gives rise to choosing the proper gray scale level. In a sense, not using all the T1 measurements maybe the best way to discriminate lesions from healthy tissue. The natural case is to set the lowest T1 to zero and the highest T1 to 1 in the photo, thus taking into consideration all T1 measurements. Future work could look to the possibility of increasing discrimination by changing this range from not being the maximum and minimum T1 measurement.

The difficulty of working with the simulated data set was the discretized tissue maps were not easy to understand. In the lesion maps, there were multiple slices

with a noise in the background and varying gray levels within each lesion to create a realistic depiction of the lesion pathology. MS lesions are not entirely all scarred tissue, but rather a mixture of dead cells, rehabilitating cells, inflammation, and infiltrating immune cells. Making the voxels in those regions somewhat heterogeneous. As well, the immune system can begin attacking an axon and not be a fully developed lesion, this is referred to as normal appearing white matter. Therefore, the Brainweb lesion maps are created to reflect this nature of MS lesions.

In the lack of knowledge and experience of doing manual MS lesion segmentation it was decided to use visibly evident lesions and transplant those relaxation times into a healthy tissue scan, thus eliminating normal appearing white matter and any fuzzy delineation of lesions from white matter. This choice of lesion definition maybe troubling, but the previous method's results were also generated using the same data set as the new method's results. The recreated results are also comparable to the previous papers published using this technique. Future investigation of this method should make sure lesions are clearly defined by an experienced radio-neurologist. This would properly proving the robust efficacy of this method.

The main argument over the data set is ideally this method should be tested on real patient data. This would give more definite understanding of the algorithm efficacy and will improve the robustness. In previous articles [36], the switch from simulated to real data decreased DSC by approximately 50%, thus showing that simulated data results may not be reflective of real world applications.

The difficulty in using real data is to find a set of accurately standardized scans. MR scanners have different acquisition parameters depending on the machine, resulting in different relaxation times for different tissues. As well having four image weightings of one patient is hard to acquire without the patient moving during scanning acquisition. Of course, these movement artifacts may be removed using

de-noising software. Therefore, real data sets need to be meticulously collected for testing.

Given that acquisition data is found the proposed algorithm would be adaptive to the scanner the images were taken on. Since the machine learning classifier is built to learn a set of examples, isolated examples given a certain scanner could be beneficial to the automatic segmentation. Therefore, the algorithm is adaptive to the acquisitions variability in MR imaging.

Chapter 7

Conclusion

This study presents a variation of a supervised MS lesion segmentation algorithm using dictionary learning. We investigated three methods that used varying feature spaces, using localized patch texture, using the voxel intensity in question, and using the voxel intensity with the standard deviation of the adjacent voxels.

We further proved the efficacy of the proposed methods by evaluating the performance of the algorithm in the presence of noise (3% and 9%). It has been concluded that the voxel intensity with local standard deviation is a plausible feature space for lesion segmentation using DL.

Although results were competitive, testing on real patient data is needed to examine the full capability of this method. We also suggested further investigations into the proper texture analysis quantification for segmentation and the variation of patch size for tissue classes.

Bibliography

[1] MAGNETOM Terra. https://www.healthcare.siemens.com/siemens_ hwem-hwem_ssxa_websites-context-root/wcm/idc/resources/hwem_ assets/mri/htmlApps/terra20150528/index.html?device=mobile& language=en, 2015.

[2] P. Anbeek, K. L. Vincken, G. S. Van Bochove, M. J. Van Osch, and J. van der Grond. Probabilistic segmentation of brain tissue in MR imaging. *Neuroimage*, pages 795–804, October 2005.

[3] J. Ashburner and K. Friston. Multimodal image coregistration and partitioninga unified framework. *Neuroimage*, pages 209–217, 1997.

[4] J. Ashburner and K. J Friston. Why voxel-based morphometry should be used. *Neuroimage*, pages 1238–1243, December 2001.

[5] J. Ashburner et al. *SPM12 Manual*. Wellcome Trust Centre for Neuroimaging, 2014.

[6] M. H. Bharati, J. J. Liu, and J. F. MacGregor. Image texture analysis: methods and comparisons.

[7] R. W. Brown, Y. C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley Sons, 2014.

[8] Manning C., Raghavan P., and Schutze H. *An Introduction to Information Retrieval*. Cambridge UP, 2009.

[9] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, and A.C. Evans. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *Neuroimage*, 1997.

[10] L. Cohen. *Time-frequency analysis (Vol. 778)*. Englewood Cliffs, NJ:: Prentice Hall PTR, 1995.

[11] A Compston, C. Confavreux, and H. Lassman. *McAlpine's Multiple Sclerosis*. Elsevier, 2006.

[12] S. Datta, B. R. Sajja, R. He, R. K. Gupta, J. S. Wolinsky, and P. A. Narayana. Segmentation of gadoliniumenhanced lesions on MRI in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, pages 932–937, April 2007.

[13] H. Deshpande, P. Maurel, and C. Barillot. Detection of multiple sclerosis lesions using sparse representations and dictionary learning. *In 2nd International Workshop on Sparsity Techniques in Medical Imaging (STMI), MICCAI*, pages 71–79, September 2014.

[14] H. Deshpande, P. Maurel, and C. Barillot. Classification of multiple sclerosis lesions using adaptive dictionary learning. *Computerized Medical Imaging and Graphics, 46*, pages 2–10, December 2015.

[15] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, pages 407–499, November 2004.

[16] F. Fazekas, F. MD MD Barkhof, M. MD Filippi, R. I. MD Grossman, D. K.B. MD Li, W. I. MD McDonald, H. F. MD McFarland, D. W. MD Paty, J. H. MD Simon, J. S. MD Wolinsky, and D. H. MD Miller. The contribution of magnetic resonance imaging to the diagnosis of multiple sclerosis. *Neurology*, pages 448–56, August 1999.

[17] M. Filippi, M.A. Horsfield, S. Bressi, and V. Martinelli. Intra- and inter-observer agreement of brain MRI lesion volume measurements in multiple sclerosis: A comparison of techniques. *Brain, a Journal of Neurology*, pages 1593–1600, December 1995.

[18] D. Garca-Lorenzo, S. Francis, S. Narayanan, D.L. Arnold, and D.L. Collins. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical Image Analysis*, pages 1–18, September 2012.

[19] D. Garca-Lorenzo, S. Prima, D. L. Collins, D. Arnold, S. P. Morrissey, and C. Barillot. Combining robust expectation maximization and mean shift algorithms for multiple sclerosis brain segmentation. *MICCAI workshop on Medical Image Analysis on Multiple Sclerosis (validation and methodological issues) (MI-AMS'2008), Sep 2008, New York, United States*, pages 82–91, September 2008.

[20] R.M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, pages 786–804, May 1979.

[21] Garth James et al. *An Introduction to statistical learning.* Hew York: Springer, 2013.

[22] J. Kittler. *Pattern Recognition: A statistical approach.* Prentice Hall, 1982.

[23] X. Llado, A. Oliver, M. Cabezas, J. Freixenet, J.C. Vilanova, A. Quiles, L. Valls, L. Rami-Torrent, and A. Rovira. Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches.

[24] C.P. Loizou, S. Petroudi, I. Seimenis, M. Pantziaris, and C.S. Pattichis. Quantitative texture analysis of brain white matter lesions derived from T2-weighted

MR images in MS patients with clinically isolated syndrome. *Journal of Radioneurology*, pages 99–114, June 2014.

[25] J. Mairal, F. Bach, and J. Ponce. Sparse modeling for image and vision processing. *Foundations and Trends in Computer Graphics and Vision 8.2-3*, pages 85–283, December 2014.

[26] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online Learning for Matrix Factorization and Sparse Coding. *Journal of Machine Learning Research*, pages 19–60, 2010.

[27] D. H. Miller, M. Filippi, F. Fazekas, J. L. Frederiksen, P. M. Matthews, X. Montalban, and C. H. Polman. Role of magnetic resonance imaging within diagnostic criteria for multiple sclerosis. *Annals of Neurology*, pages 273–278, July 2004.

[28] H. Murase and S. Nayar. Visual Learning and Recognition of 3-D Objects from Appearance. *International Journal of Computer Vision*, pages 5–24, January 1994.

[29] A. Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill, 1965.

[30] D. Pham, C. Xu, and J.L. Prince. Current Methods in Medical Image Segmentation. *Annual Review Biomedical Engineering*, pages 315–37, August 2000.

[31] N. Shiee, P. L. Bazin, A. Ozturk, D. S. Reich, P. A. Calabresi, and D. L. Pham. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *Neuroimage*, pages 1524–1535, 2010.

[32] M. Styner, J. Lee, B. Chin, M. Chin, O. Commowick, H. Tran, and S. ... Warfield. 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. *Midas Journal*, pages 1–6.

[33] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, pages 267–288, January 1996.

[34] J.P. Walters, Z. Liang, W. Shi, and V. Chaudhary. *Wireless sensor network security: A survey*, chapter 17, page 367. Security in Distributed, Grid, Mobile, and Pervasive Computing. Auerbach Publications, 2007.

[35] Bradford Weeks. LDN for MS. `http://weeksmd.com/2009/04/ldn-for-ms/`, April 2009.

[36] N. Weiss, D. Rueckert, and A. Rao. Multiple sclerosis lesion segmentation using dictionary learning and sparse coding. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 735–742, September 2013.

[37] Y. Zhang. MRI Texture Analysis in Multiple Sclerosis. *Journal of Biomedical Imaging*, page 2, January 2012.

[38] Y. Zhang, H. Zhu, J.R. Mitchell, F. Costello, and L.M. Metz. T2 MRI texture analysis is a sensitive measure of tissue injury and recovery resulting from acute inflammatory lesions in multiple sclerosis. *Neuroimage*, pages 107–111, April 2009.

[39] K. H. Zou et al. Statistical validation of image segmentation quality based on a spatial overlap index 1: Scientific reports. *Academic radiology 11.2*, pages 178–189, 2004.