TEXTURAL CLASSIFICATION OF MULTIPLE SCLEROSIS LESIONS IN MULTIMODAL MRI VOLUMES

by

Ian Storey

A thesis submitted for the degree of MSc Computer Science by Research

University of Lincoln

College of Science
School of Computer Science

October 2019

Abstract

Background and objectives: Multiple Sclerosis is a common relapsing demyelinating disease causing the significant degradation of cognitive and motor skills and contributes towards a reduced life expectancy of 5 to 10 years. The identification of Multiple Sclerosis Lesions at early stages of a patient's life can play a significant role in the diagnosis, treatment and prognosis for that individual. In recent years the process of disease detection has been aided through the implementation of radiomic pipelines for texture extraction and classification utilising Computer Vision and Machine Learning techniques.

Eight Multiple Sclerosis Patient datasets have been supplied, each containing one standard clinical T2 MRI sequence and four diffusion-weighted sequences (T2, FA, ADC, AD, RD).

This work proposes a Multimodal Multiple Sclerosis Lesion segmentation methodology utilising supervised texture analysis, feature selection and classification. Three Machine Learning models were applied to Multimodal MRI data and tested using unseen patient datasets to evaluate the classification performance of various extracted features, feature selection algorithms and classifiers to MRI volumes uncommonly applied to MS Lesion detection.

Method: First Order Statistics, Haralick Texture Features, Gray-Level Run-Lengths, Histogram of Oriented Gradients and Local Binary Patterns were extracted from MRI volumes which were minimally pre-processed using a skull stripping and background removal algorithm. mRMR and LASSO feature selection algorithms were applied to identify a subset of rankings for use in Machine Learning using Support Vector Machine, Random Forests and Extreme Learning Machine classification.

Results: ELM achieved a top slice classification accuracy of 85% while SVM achieved 79% and RF 78%. It was found that combining information from all MRI sequences increased the classification performance when analysing unseen T2 scans in almost all cases. LASSO and mRMR feature selection methods failed to increase accuracy, and the highest-scoring group of features were Haralick Texture Features, derived from Grey-Level Co-occurrence matrices.

Acknowledgement

Throughout this research, I have received a great deal of support and assistance. I would like to thank my supervisor Dr Tryphon Lambrou, who has provided continuous guidance and inspiration throughout this research. His expertise and advice in Computer Vision and Medical Imaging have been instrumental in my development as a researcher.

I would also like to thank my secondary supervisor Professor Xujiong Ye who had a continuous presence throughout my degree in both listening to and advising on the work which had been done. Having a second expert in Computer Vision and Medical Imaging supervise this research elevated it to a level of detail I could never have envisioned.

I would finally like to thank my friends and family who have had to put up with me for the last two years and have never failed to provide support and encouragement throughout this degree. This research would not have been possible without them.

Ian Storey.

Contents

List of Figures

List of Tables

List of Abbreviations

AR	Auto regression
AUC	Area Under Curve
BCR	Balanced Classification Rate
BER	Balanced Error Rate
CDMS	Clinically Definite Multiple Sclerosis
CIS	Clinically Isolated Syndrome
CNN	Convolutional Neural Network
CNS	Central Nervous System
CRF	Conditional Random Field
CSF	Cerebral Spinal Fluid
CT	Computed Tomography
DDP	Decimal Descriptor Pattern
DICE	Dice-sorensen coefficient
DSC	Dice Similarity Coefficient
DSCE	Dynamic Susceptibility Contrast-Enhanced Perfusion
DWI	Diffusion Weighted Imaging
ELM	Extreme Learning Machine
FLAIR	Fluid-attenuated Inversion Recovery
FN	False Negative
FP	False Positive
GM	Grey Matter
GMM	Gaussian Mixture Model
HoG	Histogram of Oriented Gradients
HTF	Haralick Texture Feature
IV	Intravenous
k-NN	k-nearest neighbour
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LR	Logistic Regression
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MTIR	Medium Tau Inversion Recovery
NA	Number of Acquisitions
NAWM	Normal Appearing White Matter
NMR	Nuclear Magnetic Resonance
NN	Neural Network
NWM	Normal White Matter
PDw	Proton Density weighted
PLS	Partial Least Squares
PNN	Probabilistic Neural Network
QDA	Quadratic Discriminant Analysis
RF	Random Forest
ROC	Receiver Operator Characteristic
ROI	Region of Interest
rs-fMRI	Resting state functional MRI
SBW	Sampling Band Width
SVM	Support Vector Machine
T1	T1-Weighted Imaging
T1*	Observed T1 reflecting true T1
T2	T2-Weighted Imaging
T2*	Observed T2 reflecting true T2
TE	Echo Time
TN	True Negative
TP	True Positive
TPR	True Positive Rate
TR	Repetition time
WM	White Matter

1 Introduction

1.1 Multiple Sclerosis

With an estimated 2.5 million cases worldwide, Multiple Sclerosis (MS) is the most frequent chronic demyelinating disease-causing long term, irreversible, disability in young adults. Presentation commonly occurs between ages 20-40 years, with a ratio of 2:1 between female and male diagnosis (Alroughani et al. 2016). Geographically, Multiple Sclerosis occurrences are rare in equatorial regions with around 15 diagnoses per 100,000 compared to 250 per 100,000 in northern Europe (Smirniotopoulos et al. 2009).

Multiple Sclerosis causes deterioration in motor and cognitive skills as well as a reduction in life expectancy of 5 to 10 years, alongside common emotional side effects including irritability and anger, often leading towards depression and anxiety (Laing et al. 2015). Symptoms of Multiple Sclerosis attack include physical (pain, fatigue), emotional (stress and anxiety) and cognitive deterioration (Carrigan et al. 2018).

Patients with MS can enjoy a comfortable quality of life through early diagnosis and consistent medication, however, this has been proven to be beneficial during the early stages of MS only. After a time patients may enter the progressive phase of MS which sees a reduction in the number of Lesions within the brain alongside any inflammation, at which point symptoms heighten and become more challenging to treat (Lassmann et al. 2007). There is no known cure for the disease, and alternative therapy approaches for stalling progression are adopted, such as pain management, physiotherapy, for maintaining motor skills, and medication for the reduction in attack duration.

1.1.1 Pathogenesis

MS is believed to be caused by a breakdown in the myelin sheath surrounding axons, or nerve cell endings as a result of the bodies auto-immune response to perceived foreign bodies entering the brain through the blood-brain barrier. T-cells contained within white blood cells and B lymphocytes attack the fatty myelin sheath surrounding sensitive nerve endings resulting in loss of signal and the build-up of scar tissue (Lubetzki & Stankoff 2014). The destruction of myelin and resultant axonal loss is known as demyelination, and the plaques or Lesions, formed by this scarring are the key pathological symptom of MS.

The criterion for diagnosing MS include the identification of at least two Lesions within white matter, alongside inflammation of the nervous system (Goldenberg 2012). Early diagnosis plays a critical role in therapy options for mitigating and controlling Multiple Sclerosis with extensive amounts of overall axonal damage occurring early on during the development phase of Lesions (Kuhlmann et al. 2002). Furthermore, symptoms occurring longer than six months have little to no chance of regeneration without the introduction of disease-modifying therapy (Rudick et al. 1997).

1.1.2 McDonald Criteria

Identification of MS during early stages is critical for patient therapy options as well as mitigating the deterioration in condition and managing symptoms. Until recently, the McDonald Criteria, the standard for MS diagnosis, stated that diagnosis requires dissemination of disease in both space and time. Dissemination in space requires presentation of at least two Lesions from T2-weighted MRI located within regions associated with MS occurrence (Periventricular, Cortial, Spinal). Dissemination in time requires the identification of enhancing and non-enhancing Lesion, within a region of frequent occurrence, using gadolinium-enhanced T2-weighted imaging (Hartung et al. 2019).

The 2017 revision to the McDonald criteria proposed a set of modifications to speed up the process of diagnosis. These changes included considering both symptomatic or asymptomatic Lesions as dissemination in space or time. Furthermore, Clinically Isolated Syndrome (CIS) can now be considered dissemination in space, CIS referring to the stage of MS during which the initial stages of demyelination occur (Schwenkenbecher et al. 2019).

Multiple Sclerosis disease identification and monitoring requires the application of MRI imaging techniques for non-invasive capturing of Lesions located within a patients brain and spinal column. MRI is the gold standard for MS detection, with T2w and fluid-attenuated inversion recovery (FLAIR) imaging proving invaluable in the identification of white matter Lesions (Filippi et al. 2019).

1.2 Medical Imaging

This work focuses on MRI imaging and the applications of Computer Vision and Classification techniques to the problem of Multiple Sclerosis Lesion detection in Multimodal MRI volumes. The following section provides a brief background to Medical Imaging.

1.2.1 X-ray, CT and Tomography

Before the late 19th century, it was impossible to perform non-invasive medical diagnosis of a patient, and disease classification often required surgical examination. Wilhelm Conrad Roentgen pioneered early medical imaging application through experimentation with Crookes cathode-ray tubes which highlighted the visual differences between soft tissue and bone density, caused by absorption rates of X-rays (Bercovich & Javitt 2018). The original X-ray image from this paper is depicted in fig. 1

The early 20th century saw the introduction of Tomography, the technique of reconstructing images from projections into anatomical image slices using X-ray imaging techniques (Seeram 2015). This work led to the development of Computerised Tomography by Godfrey Hounsfield in 1971, an imaging modality frequently applied to cancer detection and chronic pulmonary diseases (Silva et al. 2018).

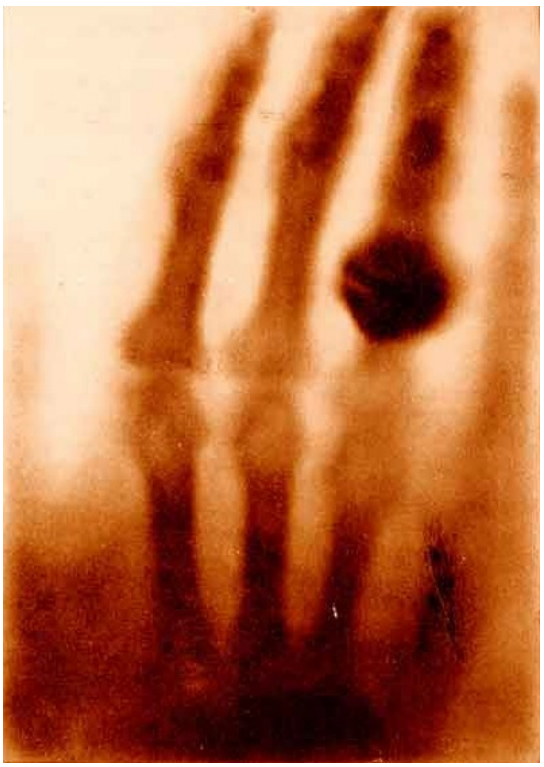


Figure 1: (NLoM 2013) First X-ray Image

The development of CT imaging was made possible through significant advancement in computer technology, which was carried out in parallel during the 20th century. Computerised Tomography performs computation on extracted attenuation coefficient data for use in the construction of anatomical cross-section volumes, obtained utilising X-ray scanning around 360 degrees allowing for a complete visual representation of the examined region (Soto & Nadrljanski 2019).

1.2.2 Magnetic Resonance Imaging

The development of Magnetic Resonance Imaging took several years and combined several fields as well as the contribution from multiple researchers. Physicists Felix Bloch and Edward Mills simultaneously discovered Nuclear Magnetic Resonance in solids and liquids through the displacement of nuclei within stationary magnetic fields for the production of electromagnetic signals. The application of NMR for image production was pioneered by Paul Lauterbur, who created NMR images across three dimensions, utilising signal decay for contrast enhancement. A significant advantage of MRI imaging, unlike X-ray and CT scanning, is that MRI does not use ionising radiation (Bell & Jones 2019).

Medical application of MRI technology was pioneered by (Damadian 1971), who demonstrated that cancer cells exhibited much longer T1 and T2 relaxation times compared to normal cells. The first full-body MRI image was generated by (Damadian et al. 1977) and subsequently applied in cancer detection, requiring a capture time of 5 hours. This first MRI image, displayed in the original paper, is shown in fig. 2 and clearly shows the cross-section of the human chest with the heart at the centre.

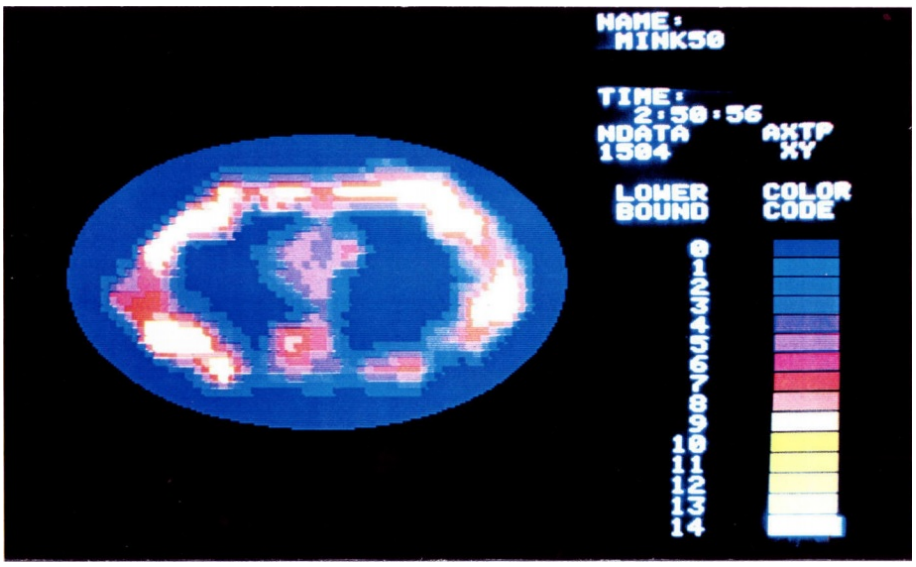


Figure 2: (Damadian et al. 1977) First MRI Image

1.3 MRI Theory

Magnetic Resonance Imaging is a medical imaging modality used for the non-invasive capture of tissue information utilising strong magnetic fields and radio frequencies. An MRI scanner consists of a hollow tube, magnet, radiofrequency coils and gradient coils, and the process of image creation using an MRI scanner concerns the application of magnetic fields and radiofrequency pulses to the body in order to stimulate resonance in hydrogen protons within the body. Protons are found extensively throughout the body at the centre of hydrogen atoms, each having orbits independent of one another. Each proton contains a north and south pole and spins on an independent rotation, similar to Earth's poles.

1.3.1 Acquisition

The MRI scanner generates magnetic field B_0 causing protons at the centre of water molecules to align into parallel and antiparallel longitudinal magnetisation along the direction of the scanning tube. As parallel and antiparallel protons have equal, but opposite magnetic moments, these will cancel out, however, more protons align in the parallel direction, known as low energy, as opposed to antiparallel or high energy. The net magnetisation vector M is the value of remaining spinning protons parallel to B_0. The protons then exhibit precession as described by the Larmor equation.

$$f_0 = \gamma B_0$$

Where f_0 is precession frequency, B_0 is the strength of the external magnetic field, and γ is the gyromagnetic ratio, the ratio between the observed angular frequency of the proton rotating and the strength of the magnetic field. Radio waves equalling Larmor frequency of hydrogen atoms are fired at the now uniform nuclei, causing resonation. Switching off the RF pulse field cause the return of protons to equilibrium within magnetic field B_0, known as relaxation, emitting small

radio signals caused by resonation which are collected and used in the formulation of tomographic grayscale images.

Depending upon settings, or sequences, chosen by a Radiologist, different tissues will exhibit wide ranges of intensity, allowing for clear differentiation between fluids, soft tissues and bone. Different chains of settings are called sequences and the sequence selected is dependent upon the scanned region and the type of disease sought.

1.3.2 MRI Sequences and Multiple Sclerosis

With regards to Multiple Sclerosis imaging, MRI acquisition is superior compared to alternative imaging modalities. T1w imaging reveals regions of severe CNS damage, known as hypointense Lesions, whereas the T2w and ADC sequences reveal regions of Lesion hyperintensity, which represent axonal loss. FLAIR imaging is very sensitive to MS Lesions and represents these regions as high signal intensity (Wilson & Smirniotopoulos 2019), proving useful in Lesion identification.

T1w, T1* and T2w images are the universal standard in MS diagnosis and monitoring owing to their ability to identify demyelinating Lesions within all regions of the brain. FLAIR MRI scanning is a standard clinical sequence in the diagnosis of Multiple Sclerosis owing to its sensitivity to Lesions close to the cortex (Juxtacortical) and around the Lateral Ventricle (Periventricular) regions (Wattjes et al. 2015).

Diffusion-weighted imaging techniques (DWI) are a new avenue of exploration by medical researchers owing to their unique ability to capture microstructural changes inherent with MS plaques (Mustafi et al. 2019). DWI techniques capture the underlying Brownian characteristic of water molecules travelling through the brain, motion, which is influenced by demyelination (Thaler et al. 2018). Concerning the application of DWI scanning, the method itself will not replace contrast-enhanced methods in MS detection. However, the combination of conventional clinical MRI scans alongside DWI techniques provides higher disease classification from multimodal diffusion biomarkers than the analysis of a single MRI sequence (Davoudi et al. 2016).

Classification of MS Lesions is carried out manually through the examination of MRI scans for the dissemination in space and time by a medical expert. This process is both time-consuming and prone to error owing to image quality and misinterpretation (Schmidt et al. 2019). Furthermore, the process of image acquisition and signal quantisation leads to issues regarding the correct boundary location of Lesions even with manual segmentation. The transformation of the analogous signal into image voxels runs the risk of unwanted MRI artefacts. Most notable of these is the partial volume effect in which signals are mixed and averaged with surrounding tissue regions, leading to difficulty in locating the exact boundaries between tissue regions.

The MRI sequences included in this research comprise a standard clinical T2 weighted image and four Diffusion based sequences which are Apparent Diffusion Coefficient, Fractional Anisotropy, Radial Diffusion and Axial Diffusion volumes.

1.3.3 Available Sequences

Within this research, T2 is the only clinical MRI sequence routinely applied in the diagnosis of demyelinating diseases. The remaining four sequences are based upon diffusive properties of tissue.

1.3.3.1 T2

The first MRI sequence used in this research is the standard clinical T2 sequence. The T2 weighted image is generated through the dephasing of protons following a 90° RF pulse, immediately after which all spinning protons are in phase. Owing to paramagnetic interference from neighbouring protons or foreign materials such as Iron, the protons become out of phase with one another exponentially. The measurement of this dephasing is categorised as T2 relaxation (Hacking & Bashir 2019).

1.3.3.2 Apparent Diffusion Coefficient

The ADC volume is produced in multiple steps. Firstly, a clinical T2* image with no diffusion attenuation is captured. This is also referred to as $b = 0$ image and is generally considered a baseline image. The second stage requires a measure of diffusivity, usually captured in three directions, x, y and z by applying gradients to either side of a 180° pulse. The output of this process is a b value, where a higher value indicates stronger diffusion. The T2* $b = 0$ image and three diffusion-weighted images are then combined through averaging the diffusion maps and dividing against the T2* image for production of the ADC image which is a measure of the magnitude of diffusion within tissue (Verma & Bashir 2019).

1.3.3.3 Fractional Anisotropy

Following the calculation of the ADC image, it is possible to quantify the asymmetry within a voxel as described by eigenvalues which are calculated from ADC diffusion maps captured in the x, y and z directions. Fractional Anisotropy represents diffusion as a value between 0 and 1 where a greater value can indicate higher parallel, and reduced perpendicular diffusion (Masdeu & Belen 2019).

1.3.3.4 Axial and Radial Diffusion

As previous mentioned, apparent diffusion coefficient represents the mean diffusivity in each direction, and Fractional Anisotropy indicates the orientation of diffusion used to identify axonal integrity. Axial diffusion describes mean diffusivity of molecules travelling parallel to the tract within regions of interest. Whilst Radial diffusion represents the magnitude of water diffusion perpendicular to this tract (Winklewski et al. 2018).

1.3.4 MRI sequence texture

Displayed in fig. 3 is a comparison between healthy and unhealthy MS Lesion tissue regions in all available MRI sequences used within this research. The brain region depicts the clinical T2w image, whilst the four diffusion based sequences depict enhanced MS Lesion areas.

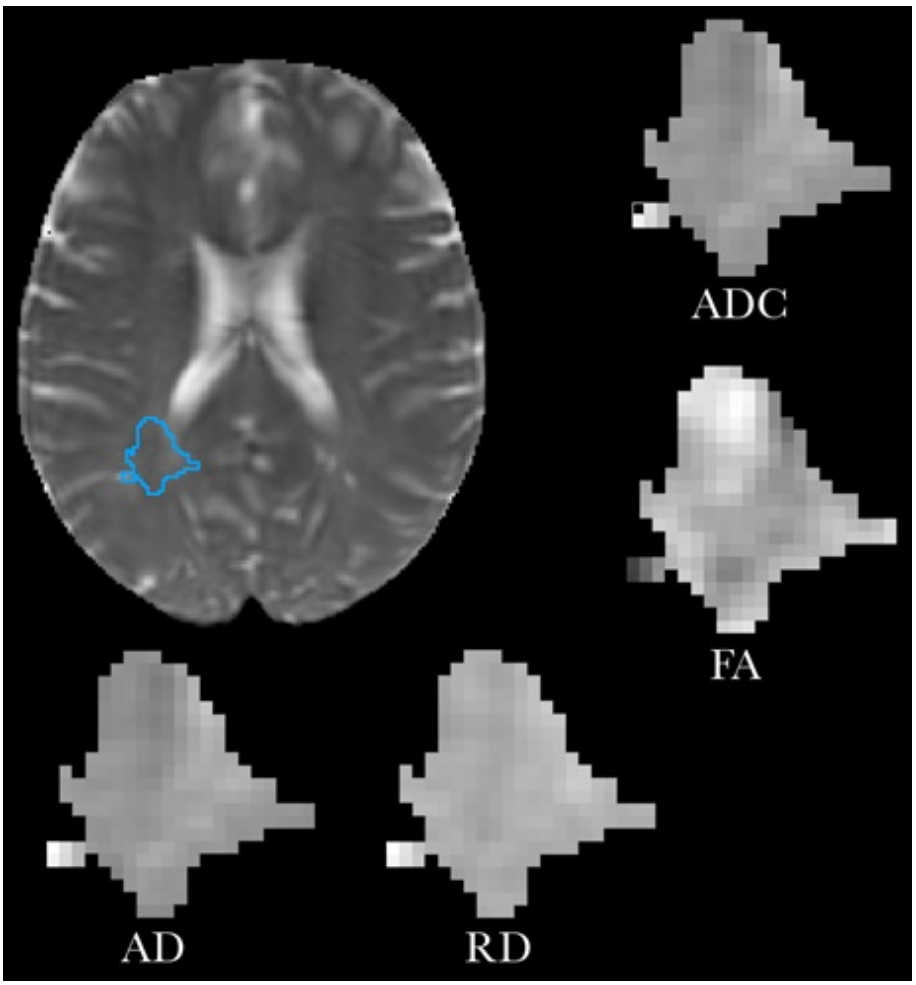


Figure 3: Example of Lesion and Brain Texture shown in dataset p01c1 slice 100 for each available MRI protocol.

1.4 Radiomics

Computer Vision and Machine Learning have been applied to Medical Imaging problems to assist in the identification of a wide range of diseases represented in patient images. Information can be captured from anatomical image slices and then classified through a machine learning model trained on existing medical data for the prediction of whether a particular tissue sample contains healthy or unhealthy qualities. The process of information extraction from medical images is Radiomics, and the process of categorising this data as healthy or diseased is Classification.

The qualitative and quantitative information elicited from medical data through Computer Vision and Image processing techniques has potential in the application of diagnosis, analysis and prognosis of disease and illness. These features can describe key characteristics of a disease, often invisible to human vision, made possible through the use of modern Computer Processing power (Gillies et al. 2015). Machine Learning models trained on databases of feature samples can be used in the Classification of unknown images, as well as sub tissue regions to assist in the prognosis and therapy of patients.

In the case of Multiple Sclerosis, correctly selected feature extraction methods have the potential to describe heterogeneous characteristics such as macroscopic textural and structural information

from MRI volumes containing MS Lesions.

1.5 Proposed Work and Motivation

This research seeks to apply modern radiomic techniques in feature extraction and selection to the problem of automatic semi-supervised Classification and Segmentation of Multiple Sclerosis Lesions in Multimodal MRI patient volumes. The available dataset consists of several DWI sequences, uncommonly applied in clinical settings to Multiple Sclerosis diagnosis. As previously discussed, the McDonald criteria require T2 Lesion in two or more areas of the central nervous system, one of which should be gadolinium-enhanced. Gadolinium is an intravenous contrast agent which displays as high-intensity regions under MRI scanning. Gadolinium-enhanced images highlight areas of normal and abnormal flow, leading to clearer imaging of damaged regions. Contained within the available dataset are a single clinically standard sequence, T2w, and four diffusion sequences.

Diffusion imaging techniques can greatly assist clinical sequences owing to their prevalence in capturing underlying water molecules travelling through axons and out of damaged regions. However, diffusion acquisition is not mandatory to MS diagnosis, and a careful balance must be found between image acquisition for expert diagnosis and patient comfort, greatly affected by the number and complexity of sequences applied.

A major issue relating to the correct identification of Multiple Sclerosis in MRI data is the time-consuming task of manual Lesion identification in slice volumes. Patient datasets containing multiple MRI sequences must be thoroughly examined by professional clinicians to ensure correct diagnosis, therapy and prognosis. In the analysis of brain volumes, as is commonplace, this can require hundreds of manual slice inspections and in some cases second opinions on a diagnosis. The application of Computer Vision and Medical Imaging techniques helps to support Radiologists and Medical experts in the task of MS Lesion identification for efficient and effective treatment.

This research applies a novel set of textural feature descriptors which have yet to be combined and applied to the problem of Multiple Sclerosis in the available Multimodal MRI sequences. This work aims to provide a comparison between three classification models, Support Vector Machine (SVM), Random Forest (RF), and Extreme Learning Machine (ELM). Described below are the aims and objectives of this work.

Aim: To develop a Multiple Sclerosis Segmentation Tool for radiomic biomarker extraction and classification in Multimodal MRI patient volumes containing Multiple Sclerosis Lesions.
Objectives:

- Minimalistic pre-processing for skull stripping and background removal, leaving greyscale brain tissue untouched.
- Extraction of First Order, Haralick, Grey Level Run Length, Histogram of Oriented Gradient, and Local Binary Pattern features from healthy and unhealthy tissue regions.

- Feature Selection utilising mRMR and LASSO feature selection algorithms for dimensionality reduction.
- Machine Learning Training through 5-fold cross-validation utilising Support Vector Machine, Random Forest, and Extreme Learning Machine.
- Evaluation and comparison of machine learning models, feature extraction and selection techniques and validity and success of Multimodal Multiple Sclerosis classification.

The following section consists of a literature review of up to date and modern approaches towards radiomic techniques for image description and Classification of Multiple Sclerosis. Following the Literature review is the Methodology, section 3, containing a detailed account of the implemented solution including motivation, research methods and software packages. Following this is the Results section, section 5, Discussion section 6 and Conclusion, section 7, to this research.

2 Literature Review

2.1 Image processing approaches

Attempts to solve the problem of Multiple Sclerosis Lesion Segmentation in MR Image data have been made utilising Image processing techniques. (Isoglu et al. 2017) propose a simple implementation of the K-means algorithm for Lesion segmentation in brain MRI containing Multiple Sclerosis. This approach identifies a K-means initialisation parameter from histograms extracted from each image, followed by an implementation of MATLAB's k-means clustering algorithm to group pixels into categories based on the initialised histogram threshold.

This simplistic methodology may work in isolated cases where a single slice containing fewer Lesions is analysed, however, the nature of MS Lesions indicates that the model will not adapt for heterogenous Lesions. The pixels of each Lesion must fall within a similar range of greyscale intensities, allowing the setting of the K-means clustering parameter. However, it is also likely that this approach will fail to correctly classify when using unseen patient samples from a separate dataset where the characteristics of the Lesion information deviates from the information used to train the original model.

A second Image Processing approach is implemented by (Ali & Maher 2016). Their methodology consists of Skull Stripping utilising a seed point at the centre of each image slice and iterating around each brain region until a binary mask is obtained. A comparison of Canny and Marr-Hildreth edge detection algorithms is carried out followed by an analysis of each extracted region using a seed point for contour analysis to satisfy the shape and size of each Lesion.

Similar to (Isoglu et al. 2017) this approach implements fundamental Image Processing techniques and makes use of edge detectors in the identification of Lesion like regions for use as input into contour analysis for shape description of these possible Lesions. This implementation is more thorough, and likely to transfer to certain unseen cases.

The nature of MR imaging and its susceptibility for bias field error and noise indicates that image processing methods such as these are not robust enough to scale from smaller domains to larger datasets. Instead, approaches implementing pattern analysis for feature extraction and evaluation have been applied owing to their ability to identify more descriptive biomarkers which are transferable to new datasets.

2.2 Radiomic Approaches

(Theocharakis et al. 2009) develop a pattern recognition system for Multiple Sclerosis Lesion segmentation in MRI data. Manual Lesion segmentations were carried out and provided the basis from which feature extraction of grey-level co-occurrence matrices (GLCM), histogram, and grey-level run length (GLRL) statistics was carried out. These extracted statistics describe the underlying characteristics of grey-level texture and can be used to train classification models in differentiating between healthy brain tissue and unhealthy MS Lesions. Four methods are implemented for

classification, Logistic Regression, Minimum Distance, Probabilistic Neural Network and Linear Discriminant Analysis.

PNN performed to a higher standard out of all classification methods in best overall accuracy with a score of 88.46%. However, the remaining methods fell just short of this score, indicating the transferable qualities of the proposed feature descriptors to a range of classification methods. The study found that each Multiple Sclerosis Region of Interest (ROI) contained high entropy and sum entropy values indicating that these areas contain coarse texture compared to healthy regions in the brain. This demonstrates a clear biological link with a disruption in MR signal intensity owing to inflammation and demyelination of tissue regions.

This approach utilised textural description techniques such as grey-level run-length and grey-level co-occurrence statistics to generate information describing comprising tissue regions. Furthermore, the extracted information is used as input into classification models which learn from a set of training data and classify unseen points into categories relating to the learned parameters.

Another implementation of classification from extracted radiomic descriptors is proposed by (Zhang et al. 2008) who present a method of Multiple Sclerosis Lesion detection utilising texture features derived from grey-level co-occurrence matrices, Run-length matrices, Gradient matrices, AR Models and Wavelet transforms. Following the feature extraction stage, a statistical analysis is carried out utilising Raw Data Analysis, Principal Component Analysis and Nonlinear discriminant analysis. Fisher coefficient is then applied to draw correlation between seperate tissue data. K-Nearest Neighbour and Artificial Neural Network are chosen as the classification algorithms for discrimination between Lesion and a range of brain tissue matters (white matter, dark matter, CSF) as well as between brain regions.

During the statistical analysis of features, it was found that a combination of all features produced the highest classification accuracy than the use of a single set of features. With combined methods producing 100% accuracy and GLCM alone achieving 92.67% accuracy between Multiple Sclerosis Lesion and Normal White Matter. It was noted that the writers found it difficult to obtain the desired averaged GLCM values from orientation matrices in the software package they were using. Averaging GLCM matrices is a step which is often carried out for rotation invariance against directional changes in MR images.

This work highlights the importance of combining differing approaches to feature extraction when performing a statistical analysis of Multiple Sclerosis Lesions in MRI data. Similar to (Theocharakis et al. 2009) the classification performance was high for GLCM statistics and GLRL statistics. Furthermore, in the work of (Zhang et al. 2008), combinations of GLCM and GLRL with Gradient matrices, AR Models and Wavelet transform features provided significant performance improvement, despite GLCM alone producing good results.

Further implementation of GLCM and GLRL features is carried out by (Tozer et al. 2009) who perform texture feature extraction and classification of Multiple Sclerosis Lesions in Magnetization

Transfer Images. MTIR scans were used to create Lesion masks which then formed the basis of Lesion region extraction from T2 weighted scans. The images were pre-processed utilising histogram normalisation at 256 grey-levels. The images were then used in the calculation of texture descriptors derived from 8 grey-level co-occurrence matrices which are then averaged to create a mean GLCM from which 14 Haralick Texture Features are calculated.

Voxels were only used in feature extraction if the pixels were present in the extracted masks. Furthermore, the voxels were only used if they were surrounded by eight pixels, excluding boundary regions surrounding the brain or Lesion. The extracted features were used in two statistical analysis tests, ANOVA and two-tailed t-test for examination of the correlation between two types of MS condition, Clinically Isolated Syndrome (CIS) and Clinically Definite Multiple Sclerosis (CDMS).

It was found that there were differences in the texture of White and grey Matter between CIS and CDMS when applying features derived from GLCM. It was also discovered that whilst it is possible to differentiate between groups, the overlap between features is not clear enough for complete group classification. Furthermore, when combined with data such as sex and gender, there is little to no dependencies between the Haralick Texture Feature statistics and these categorical metrics.

Unlike previous literature, this research focuses on the relationship between Clinically Isolated Syndrome and Clinically Definite Multiple Sclerosis images. Whilst this isn't a Lesion classification methodology, it does present key information which transfers to Lesion Classification and Segmentation methodologies. Firstly, the statistics used, Haralick Texture Features derived from grey-level co-occurrence statistics provide the ability to differentiate between the two types of Multiple Sclerosis syndrome. Furthermore, an averaging of grey-level co-occurrence matrices is carried out which intends to reduce the directional effect of each GLCM.

(Mayerhoefer et al. 2009) carry out a thorough investigation into the performance of various texture analysis techniques with regard to varying parameters in MRI acquisition settings such as repetition time, echo time and spatial resolutions of the acquired MRI volumes. The textural approaches investigated were statistics derived from co-occurrence matrices, run-length matrices, absolute gradient, autoregressive model and the wavelet transform.

Motivation for this work focuses on the susceptibility of texture feature sensitivity towards selected MRI acquisition settings. Four MRI acquisition experiments were carried out, one each for number of acquisitions (NA), repetition time (TR), echo time (TE), and sampling band width (SBW), applying varying settings to the parameter being examined whilst applying a default standard clinical value for the remaining parameters.

Following the acquisition of images, a second stage of feature extraction was performed on each set of acquired scans in order to capture the proposed texture statistics. Classification was carried out utilising the K-nearest neighbour classifier trained on a subset of features identified through linear discriminant analysis for dimensionality reduction.

The results of this study found that co-occurrence and run length features are affected most when analysing varying ranges of NA TR, TE and SBW, however, all features sensitivity to settings decreases as the spatial resolution of MR images reduces. Furthermore, if a dataset has varying spatial resolution between volumes, co-occurrence matrix textural descriptors are found to be the only approach to retain a high accuracy. It was noted by the authors that co-occurrence matrix descriptors appear to be the best all-around texture feature examined.

Different approaches to the task of Multiple Sclerosis Lesion segmentation are evaluated by (Lladó et al. 2012) who provide a review of existing work in the field of Medical Imaging on Multiple Sclerosis. A critical point in their discussion concerns the choice of supervised or unsupervised methodologies. Often the selection of a particular approach is dependent on the dataset itself owing to human error and fuzzy boundary detection as a result of the partial volume effect in Medical Imaging problems.

Furthermore, labelling which is required for supervised learning is time consuming and requires one or more radiologist experts to confirm and outline the presence of a disease in large numbers of scans. It is also found that the selection of a particular framework is dependent on whether the methodology requires Lesion identification in single MR volumes, or whether longitudinal analysis is desired, where Lesion detection and change monitoring is performed within the same algorithm.

The work examined so far has focused on varying approaches of Radiomic feature extraction from MRI images containing MS. This information is only one step in a Medical Imaging methodology, with classification forming the key link between extracted information and disease identification.

2.3 Approaches to Classification

(Karimaghaloo et al. 2012) propose an automatic method for the detection of Multiple Sclerosis Lesions in gadolinium enhanced MR images utilising posterior probabilities calculated at each pixel for classification as identification of enhancing Lesion and background. The proposed methodology applies a conditional random field which is shown to exhibit greater classification accuracy than Linear Regression, Support Vector Machine and Markov Random field in the case of gadolinium enhanced T2 images.

The CRF is evaluated on 80 multimodal MRI scans from patients with Relapse Remitting Multiple Sclerosis, considering local and more global information from pixel neighbourhoods. The average accuracy of this methodology was 98% successful Lesion prediction with an average number of False Positive values of 2.43.

It is identified that a key problem regarding gadolinium enhancement relates to healthy tissue region enhancement alongside the target disease, leading to greater risk of false positive classification when implementing feature extraction techniques based on texture and intensity. The authors also highlight key corrections which may be required for MRI data extraction in which bias field inhomogeneity and misaligned spatial coordinates present unwanted artefacts which much be addressed before the validity of a study can be confirmed.

(Roy et al. 2013) examine the effectiveness of SVM classification of MS Lesions using texture features obtained from multimodal T1w, T2w and FLAIR MRI data. For pre-processing steps, intensity normalisation and contrast enhancement are applied for greyscale thresholding. Image regions are further developed through the application of a sobel gradient for contrast enhancement. The second step consists of feature extraction using a 5x5 region window to calculate mean and variance alongside gradient based features, run length features, and co-occurrence statistics captured at 4 orientations (0, 45, 90 135). During feature extraction global information is also extracted around the candidate local pixel regions for calculating the probability of that region containing an MS Lesion.

All pixel information and candidate regions are then used for classification using a Support Vector Machine with a Linear Kernel and soft margin C set to 1. The feature vector consists of three local intensity features from T1w, T2w and FLAIR images using three tissue probability maps from White Matter, Grey Matter and Cerebral Spinal Fluid.

Following classification, a post processing stage consists of morphological dilation applied to regions identified as MS Lesion. The structuring element used in dilation is a 1mm circle applied to expand the Lesion regions. Fuzzy C-means clustering is used to remove unwanted artefacts from CSF regions followed by morphological closing using a circle of radius 10 pixels.

This method of Multiple Sclerosis Lesion segmentation achieves a mean F1 score of 0.5 when applied to 8 sets of T1w, T2w and FLAIR MRI from the Childrens Hospital Boston dataset as part of the MS Lesion Segmentation Challenge 2008. The method outperformed existing state of the art approaches which were submitted as part of the challenge, with the closest score achieving a mean F1 value of 0.44.

The work by (Roy et al. 2013) applies a variety of texture features to Support Vector Machine classification and evaluates the approach on a range of MS datasets with positive results. The application of morphology during post-processing allows for identified regions to be further processed. The post-processing approach assists with previously mentioned problems regarding MRI acquisition in which boundary regions suffer from the partial volume effect, as well as the challenges associated with correctly identifying Lesion edges.

(Sweeney et al. 2014) evaluate the performance of varying classification algorithms and feature vectors in relation to MS Lesion detection in Multimodal MR images. This work does not seek to determine the best features or classification algorithm, instead it evaluates the performance of different metrics in relation to each approach. The six feature vectors examined consist of observed raw data, normalised data, random voxel selection, smoothed MR image, Moments and Smoothed Moments. The classification algorithms chosen were Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Gaussian mixture model, Support Vector Machine, Random Forest, K-Nearest Neighbour and Neural Network.

The experiments carried out analysed classification performance, algorithm agreement, and computational time. In classification performance it was found that more complex models such as Gaussian mixture model, Neural Network and Random Forest performed to a higher standard than the other classifiers examined. It was also found that models classifying normalised information converged at a certain point, leading to similar classification performance across all models providing a rigorous training procedure is adopted. It was also highlighted that a significant number of classification algorithms perform below average on unbalanced training data, a common case with brain diseases where healthy tissue greatly outnumbers unhealthy regions.

It was decided that the selection of an appropriate feature extraction methodology holds more importance than the selection of classification algorithms as differences in DSV, ROC and AUC values were attributed to the various feature vectors implemented rather than classification algorithm performance.

The paper by (Sweeney et al. 2014) provides good argument and guidance on areas in which the most time and effort should be focused in the development of Multiple Sclerosis segmentation methodologies. Particular attention should be paid on the implementation of feature extraction techniques and it was found that the selection of a classification algorithm is less critical when a thorough and reliable radiomic extraction stage is implemented.

(Saccà et al. 2018) evaluate the performance of several machine learning algorithms applied to the prediction of early onset Multiple Sclerosis. The sequences examined were T1w, T2w and resting-state functional MRI. Pre-processing consisted of registration, skull stripping, movement and noise correction and smoothing.

Lesions were manually identified by two experts for Lesion mask creation to be used during the research. The focus of this work was a comparative analysis between five machine learning approaches for Multiple Sclerosis feature classification. Features were extracted utilising independent component analysis to decompose regional information into composing multivariate signals. The approach of extracting information from the images was to create White matter and Cerebral Spinal Fluid masks using T1w scans. The masks were then applied to rs-FMRI for signal extraction, in order to collect a range of rs-FMRI signals, each representing a different tissue type.

The classification algorithms examined were Random Forest, Support Vector Machine, Naïve Bayes, K-nearest neighbours, and Artificial Neural Network. The research also applied feature selection using the GIFT toolbox, for the identification of the feature rankings to be used as input into all machine learning models. Gini index was used for Random Forest feature selection whilst recursive feature elimination generated feature indexes for SVM and k-NN classification.

5-fold cross validation was used with a split of 80% training and 20% testing data to avoid overfitting and bias errors. The authors remark that 5-fold cross-validation performed poorly and the margin between each methods classification performance was extreme owing to different feature selection algorithms being implemented across all classifiers. SVM achieved 63.3% accuracy and the lowest score was Naïve Bayes which achieved 46.6% accuracy.

Following the 5-fold cross-validation test a second experiment was carried out using sensori-motor I features. This second approach using a single set of features achieved 85.7% accuracy for Random Forest and SVM. Whilst Naïve Bayes, k-Nearest Neighbour and Artificial Neural network all achieved an accuracy of 71.42%. It was observed that the best overall classifiers were Random Forest and Support Vector Machines, both of which were successful in Multiple Sclerosis Lesion detection.

One criticism of the work carried out by (Saccà et al. 2018) is the use of different feature selection algorithms on each classifier. Intrinsic feature selection based around the Gini index was used for Random Forest feature selection. Recursive feature elimination was applied to SVM and k-NN classifiers. The absolute value of the t-statistic was used for Naïve Bayes and the absolute values of weights was used for the Artificial Neural Network.

Applying varying feature selection techniques across separate classifier results in a lack of comparative analysis in the examination of how each classifier respond to features inputted into each network. A more robust implementation would be to train a set of all models using each feature selection technique to ascertain the performance of each model alongside the performance of each set of feature descriptors, as well as the performance of each feature selection algorithm. An informed judgement could then be made as to the best classifier, the best feature selection algorithm, and the best set of features.

2.4 Approaches to Feature Extraction

A thorough investigation into various types of radiomic textural descriptors is provided by (Loizou et al. 2015) who implement an evaluative textural methodology for the comparison of different approaches to feature extraction from MS MRI images. Their feature extraction approach captures shape and textural information before performing statistical analysis on the obtained information to identify differences between healthy tissue and MS Lesion.

The shape and textural descriptors consist of area and perimeter information. Statistical features consist of first-order statistics, spatial grey-level dependence matrices such as Haralick's texture features, grey-level difference statistics, neighbourhood grey tone difference matrices, statistical feature matrix, laws texture energy measures, fractal dimension texture analysis and Fourier power spectrum. A sliding window is applied in the collection of standard deviation and median statistics.

The results of the statistical analysis found that skewness, mean and median could be used for Lesion identification across time series data proving their use as universal feature descriptors in MS segmentation. Furthermore, it was found that there was significant variation between texture features in longitudinal data indicating a potential drawback concerning statistical descriptors when presented with differing time points in MR image data. It was also found that first-order statistics such as median and mean could be used to identify Lesions at various points in time such as 0 month, 6-12 months and 2-5 years.

The study also found that combining GLCM features with additional texture descriptors yielded greater accuracy in the differentiation between Multiple Sclerosis Lesion and Normal White Matter, with a classification accuracy greater than 90%, a finding also discussed by (Zhang et al. 2008).

(Michoux et al. 2015) analyse the classification performance of MR Brain images containing Multiple Sclerosis. The approach consisted of grey-level co-occurrence matrix and run length matrix calculation followed by the capturing of Haralick texture features and Galloway's grey-level run-length statistics. Three classification models were tested, Linear Discriminant Analysis, Logistic regression models and partial least squares test.

The results show that in the comparison between healthy white matter and MS Lesions, in both enhanced and unenhanced MR images, the p-values are highly significant. This indicates strong differences between the structures of both tissues. Furthermore, the use of texture features derived from GLRLM's and GLCM's has proven significant in the classification of Multiple Sclerosis Lesions in T2 images and may be relevant to brain inflammation detection in Multiple Sclerosis patients.

Similar to (Loizou et al. 2015), (Michoux et al. 2015) found that texture parameters between White Matter Lesions and normal appearing White Matter were significantly different, leading to the reaffirmation that GLRLM and GLCM statistics are viable candidates for Multiple Sclerosis texture description. It was also found that the ADC parameter did not assist in Lesion identification. The point is again raised that combining texture parameters could lead to greater description in MR image analysis.

(Roura et al. 2015) examine T2-weighted and FLAIR MR images containing Multiple Sclerosis Lesions for the production of a segmentation tool to aid patient's diagnosis and prognosis stages of treatment. Pre-processing consists of Skull stripping, denoising and bias correction followed by co-registration. Skull stripping is carried out using an existing Brain Extraction Tool Software.

The process of Lesion segmentation is two-fold. Initially the brain tissue is segmented into white matter, grey matter, and cerebrospinal fluid. Secondly Lesions are segmented as outliers using a threshold from the grey matter regions identified from T1w tissue to find the grey matter distribution in FLAIR images for Lesion mask creation. Thresholding values are trained iteratively for the reduction in False Positives and increase of True Positive detections.

The algorithm was trained using the Children's Hospital Boston dataset and tested using the MICCAI MS Challenge 2008 dataset. A Dice Similarity Coefficient of 41 was obtained during the first iteration of training, followed by 43 during subsequent iterations. A TPR of 0.51 was also achieved and it was noted that the rate of FP detections increased slightly, although a high number of TP was also captured, contributing towards a high overall success rate.

This approach presents a method of Multiple Sclerosis Lesion segmentation which focuses on outlier observation as opposed to texture analysis between tissue and Lesion. Whilst the overall results are average, the detail used in presentation of classification scores is useful. Furthermore, the

decomposition of MRI data into brain regions such as White Matter, Grey Matter and CSF leads to more targeted analysis of tissue areas and increased computation performance as Multiple Sclerosis will not appear in CSF, leading to the exclusion of these regions.

(Ghribi et al. 2018) propose a method of Multiple Sclerosis Lesion classification in multimodal MR image data. The sequences used are T1w, T1*, T2w and FLAIR. Inhomogeneity correction, skull stripping, anisotropic diffusion filtering and voxel intensity normalisation were applied during pre-processing. A key intention with this step was to scale the intensity of each MRI modality into a uniform range relative to each scanning sequence through inhomogeneity correction. This not only set each sequence to the same grayscale range, but also led to reduced computational complexity when capturing multimodal statistics. Skull stripping was carried out for noise and tissue removal before the feature extraction stage.

The primary approach for capturing feature information utilised statistical feature descriptors derived from grey-level co-occurrence matrices and grey-level run-length matrices. GLCM can be used to extract texture features such as coarseness, regularity and fine details in the spatial intensity distributions of pixels within a region of interest. Grey-level run-length matrices identify the amount of change in pixel intensity between the region of interest and return a matrix of run length information from which more detailed descriptors can be extracted. Both matrices were computed in 3 dimensions using a sliding region window of 4x4x4 pixels, and the matrices were captured in four orientations around each pixel centre.

Classification was carried out using an SVM combined with a function for predicting the validity of each classification probability which was compared against the absolute values of Lesions obtained from a manual segmentation. Dice similarity scored 0.68 indicating acceptable identification of Multiple Sclerosis Lesions in FLAIR sequence MR images. A higher accuracy of 0.83 was achieved using an unseen dataset, indicating the robustness of the approach utilising statistics derived from GLCM and GLRLM for Multiple Sclerosis Lesion identification.

Significant analysis so far has focused upon texture analysis utilising grey-level co-occurrence matrices and grey-level run length matrices. A wider range of texture descriptors have been applied to MRI disease classification outside of the scope of Multiple Sclerosis. (Abbasi & Tajeripour 2017) apply Local Binary Patterns and Histogram of Oriented Gradients for MR image analysis in the detection and segmentation of brain tumours.

Their methodology consists of pre-processing for the removal of noisy regions such as the skull and facial tissues as well as unwanted information collected during image capture such as magnetic field bias, IV lines and noise. Otsu thresholding is then applied in order to extract tumour regions such as necrosis, non-enhancing tumour, enhancing tumour and oedema before feature extraction consisting of Local binary patterns and Histogram of oriented gradients.

The Local binary pattern approach provides uniform analysis of each binary word to increase effectiveness and remove the impact of noise. The final feature vector consists of the uniform local

binary patterns taken in both spatial and time domain of the image for a more detailed descriptor. Histogram of Oriented Gradients is the second feature extraction technique applied in this study which is more widely applied for object detection in Computer Vision. A neighbourhood of nine pixels is selected, and a subsequent 9-dimension feature vector is defined for use in classification. Owing to the use of 3D volumetric image data they proposed the extraction of HOG information in three dimensions around each central voxel.

Random Forest is applied for classification of features, achieving good results with a maximum Dice similarity score of 0.93. The classification accuracies achieved in this approach for Brain Tumour segmentation improves significantly upon existing methodologies developed for analysis of the same dataset.

This work focuses on Brain Tumour detection in Multimodal MRI data, and whilst not wholly relevant to Multiple Sclerosis, the features proposed are transferable to different diseases in Magnetic Resonance Imaging where shape and texture information are highly important in classification. The authors noted that the use of LBP and HOG features with a Random Forest classifier provided very good results whilst maintaining a smaller feature vector. This improves the performance whilst retaining a high level of classification accuracy. Furthermore, the Random Forest classifier performs well on unbalanced datasets, providing a solution to the inherent class imbalance problem in MS datasets.

(Akbarpour et al. 2017) propose a method for Unsupervised Multimodal MRI Segmentation of Multiple Sclerosis Lesions focusing on edge and texture descriptors. The MRI sequences investigated are T1 and T2 weighted images. Pre-processing consists of skull stripping followed by statistical feature extraction utilising a 3x3 region window.

Features are extracted in both the frequency and spatial domains through the application of a stationary wavelet transform for frequency representation. Following this is the calculation of two statistical features, energy and entropy, which indicate whether consecutive pixels belong to the same type of texture. Fuzzy C-Means Clustering is then applied for segmentation into four classes of Grey matter, White matter, Cerebrospinal fluid and Multiple Sclerosis Lesions.

The results of the work produce a Dice similarity score of around 0.9 which improves upon existing research by 16.38%. Whilst this work focuses upon an unsupervised implementation of Multiple Sclerosis Lesion segmentation, it highlights several useful techniques which are transferable to supervised classification. Jaccard and Dice alongside Sensitivity and Specificity metrics provide powerful indication as to the performance of a classifier when producing binary masks of Lesions. Furthermore, textural information such as energy and entropy which can be extracted from grey-level co-occurrence matrices have been found to be viable in T1w and T2w MR images. The authors also discuss that the Fuzzy C-means algorithm is susceptible to noise within MR images indicating the possible application of alternate classification algorithm.

Earlier described work by (Abbasi & Tajeripour 2017) applied local binary patterns and Histogram of Oriented gradients to the task of Multiple Sclerosis detection. A similar set of features are applied by (Samah et al. 2018) who extract local binary patterns alongside grey-level co-occurrence texture features and Decimal Descriptor Pattern texture descriptors from noisy T1w and T2w weighted MR images containing Multiple Sclerosis. A 3x3 pixel region is applied in the extraction of all features and grey-level co-occurrence matrices are calculated at 0, 90 45 and 135 degrees.

The extracted information is inputted into a multi-class SVM for sample categorisation into diseased and healthy tissue types. The main aim of this research was to analyse the performance of different feature descriptors on noisy MRI data using varying amounts of noise for each iteration. The results of the work found that in the case of noisy images the Decimal Descriptor Patterns outperformed grey-level co-occurrence matrix features and Local Binary Patterns. GLCM and LBP features performed well on T2w images and poorly on T1w images. Decimal Descriptor Patterns achieved around 97.12% T1w accuracy and 94.21% T2w accuracy. GLCM achieved 85.83% T1w accuracy and 50.50% T2w accuracy Whilst LBP achieved 70.87% T1w accuracy and 79.07% T2w accuracy.

This work highlights the importance of noise in relation to Multiple Sclerosis Lesion detection. Furthermore, the correct selection of texture features can positively impact a methodologies susceptibility to noise and misclassification. Without noise, all methods performed well, however, Decimal Descriptor Patterns prove to be the most robust descriptor for images exhibiting unwanted artefacts. One argument against the work concerns the inclusion of a portion of simulated MRI volumes for noise generation. Whilst the creation of such scans contributed towards the datasets lack of noisy image examples, there is argument that real world unwanted MRI artefacts are difficult to replicate and as such results are not an accurate representation of real world MRI imperfection.

2.5 Deep Learning approaches

(Verma et al. 2017) present a study examining dynamic texture parameter analysis with regard to contrast enhancing gadolinium injection in time series MRI data. The key focus is in identifying whether perfusion-related differences in texture exist over time during the application of a contrast bolus. The work also sought to identify whether texture parameters are different between Glioblastoma, Multiple Sclerosis Lesion and Cerebral Lymphoma.

The MRI sequences investigated are dynamic susceptibility contrast-enhanced perfusion (DSCE), T1w and T2* DSCE images. Four texture parameter maps are examined including Mean intensity, Standard Deviation, Variance, and Variance of Variance. SPSS was used in analysing the statistical significance of the extracted texture maps and the tests used were WELCH-ANOVA and Games Howell.

It was found that 24 of the 48 extracted texture parameters were statistically significant between all diseases, and the time at which images are analysed determines the successful use of features.

For example, the statistical differences between features is highly significant during earlier phases of contrast application with deterioration in significance overtime inline with contrast enhancement decay.

This work focuses on time series MRI scans examining the application of contrast agents for Multiple Sclerosis Lesion, Glioblastoma and Cerebral Lymphoma, unlike previously examined work focusing on static MRI volumes. The application of first order statistics such as Mean, Standard Deviation and Variance has produced noticeable statistical significance in identifying underlying diseases in time series MRI data, prompting the possibility to include such statistics as part of a larger feature vector of texture focused descriptors.

Another Deep Learning approach is proposed by (Valverde et al. 2017) who implement a Convolutional Neural Network for automatic White Matter Lesion Segmentation in Multimodal MR images of Multiple Sclerosis patients. The sequences examined are FLAIR, T1w, and T2w images. For network training, sample patches are fed into a 7-layer Convolutional Neural Network (CNN) for multimodal training of axial regions. A second CNN architecture is also proposed in which a probability is calculated, indicating the likelihood of Multiple Sclerosis Lesion on the reduced input image from the output of the initial CNN framework. Training and validation datasets are created for independent model training of both classifiers without sharing any information between the two CNN models.

For testing, an unseen image is inputted into the cascading architecture for classification. The first CNN removes all voxels with a low probability of Lesion categorisation. The second model then takes this reduced image as input and classifies regions with high Lesion probability, returning a binary mask of those areas. The dataset used is the 2008 MICCAI Multiple Sclerosis challenge containing 45 scans of T1w, T2w and FLAIR sequences. The proposed methodology ranked in the top three when testing on all image sequences, achieving an overall score of 87.12% classification accuracy.

This approach manages class imbalance between healthy brain and unhealthy Lesion samples which is a key issue in Multiple Sclerosis Lesion identification problems. The class imbalance problem was tackled through under sampling the major class of healthy brain and extracting a random sample equal in size to that of the unhealthy class. In binary segmentation results, the proposed method achieved a Dice similarity score of 53.5 on the MS1 dataset, and 56 on the MS2 dataset indicating a good segmentation accuracy with low False Positive Rates.

(Yoo et al. 2018) propose an unsupervised Deep Learning approach for Multiple Sclerosis Segmentation in T1w and myelin MR images using a 9x9x9 patch applied to normal-appearing brain tissue. The deep learning framework is based around existing literature in neural network implementations of multimodal feature learning. The focus of the architecture is to identify normal appearing tissue regions and use these as input to a deep learning framework for the extraction of a multimodal feature vector. In classification, areas of tissue exhibiting outlier characteristics are then classified as Lesion.

LASSO feature selection is applied to the extracted feature vector for dimensionality reduction before a Random Forest classifier is trained. Classification consisted of 11-fold cross-validation with an averaged performance captured across each fold. Sensitivity and Specificity metrics were gathered as well as area under curve (AUC) value from receiver operating characteristic. Overall a mean value of 73.7% accuracy was achieved for myelin images, dropping to 73.4% accuracy with LASSO feature selection. T1w accuracy was 70.7% which increased by 3% following the application of the LASSO algorithm.

This paper introduces feature reduction through the Least Absolute Shrinkage Selection Operator (LASSO) algorithm, implemented to reduce the dimensionality of feature vectors inputting into the Random Forest classifier. The algorithm removed between 60 - 80% of regional mean features and also increased the accuracy by 3% for feature vectors extracted from T1w images. Contrarily, features obtained from the deep learning framework inputted into LASSO feature selection reduced the accuracy by around 0.3%.

This work shows the importance of feature selection algorithms and their applications. The ability to remove large amounts of information whilst retaining a high classification accuracy can improve the speed of classification methodologies. This highlights the importance of certain extracted features to the histopathology of a disease, assisting in the identification of meaningful biomarkers. It must also be noted that feature selection can reduce the accuracy of a study in certain cases, indicating the necessity for careful application of these techniques.

In recent years significant work has been carried out in Multiple Sclerosis Lesion segmentation using Deep Learning architectures such as Convolutional Neural Networks. (Aslani et al. 2019) propose a multimodal segmentation method for Multiple Sclerosis Lesions in MRI data utilising a CNN. The sequences used are T1w, T2w, PDw and FLAIR from a dataset of 5 patients. 14 patients are then used for the extraction of testing data and information is analysed on a slice by slice basis as opposed to 3D regions.

The basic idea behind the CNN architecture focuses on a modified ResNet50, 50-layer network for pixel segmentation. A unique ResNet is designed for each sequence and images are analysed on the axial, sagittal and coronal planes leading to a three network pipeline for T1w, T2w and FLAIR images. After each block in the Convolutional Neural Network the feature maps can be extracted before the next step of convolutions takes place. Meaning 109x109, 54x54, 27x27, 14x14 and 7x7 feature maps are extracted from each sequence slice.

Concatenating features extracted from each network block produces a set of descriptors representing a particular resolution for all sequences, known as multimodal feature fusion. Following this a multiscale up sampling stage is performed in which lower resolution features are combined with higher resolution features to create a single feature map encompassing all subsampled descriptors.

In network training, slices containing at least a single Lesion were included in the dataset to avoid overfitting with slices containing entirely healthy tissue. The learning rate was set to 0.0001 with a multiplier of 0.95 every 400 steps.

Two tests were performed, the first using the single branch network, and the second using the multi-branch network. It was found that the best performance overall was a multi-branch approach using FLAIR, T1w and T2w sequences, achieving 0.7067 Dice Similarity. The highest score from the single branch network was again FLAIR, T1w and T2w features with a Dice of 0.6712.

This approach using a Convolutional Neural Network produces high Dice similarity scores with low false positive rates in the problem of Multiple Sclerosis Lesion Segmentation. The extraction of features at multiple resolutions is critical to the proposed second branch approach where feature fusion between different resolutions of extracted descriptors leads to the highest classification accuracies. A drawback however is the necessity to use three different MRI sequences in the testing of Multiple Sclerosis in a single patient case. The ideal solution of using a single clinical sequence from any MRI modality for high accuracy is not possible with this architecture.

Having examined a significant amount of modern research into MS Lesion classification and segmentation, it is now possible to compare key works to understand the various methods and approaches to solving this problem.

2.6 Summary

Examining the reviewed literature, it is now possible to understand more thoroughly the topic of Multiple Sclerosis Lesion Segmentation through the insight of those who have undertaken similar challenges. It is clear that simple approaches such as the Image Processing methodologies of (Ali and Maher, 2016) and (Isoglu, Koka and Duru, 2017) lack robustness and the transferability to images outside a small domain of training examples.

Furthermore, features such as Decimal Descriptor Patterns, Local Binary Patterns and Histogram of Oriented gradients perform well in situations where Haralick texture and Gray Level Run Length features achieve poor classification.

Unwanted MRI artefacts such as bias, noise and inhomogeneity require lengthy feature descriptors for Lesion identification. In contrast, Deep Learning methodologies such as Convolutional Neural Networks derive descriptors from the repeated downsampling and extraction of information, performed at each layer of the network. Over time the model applies backwards propagation to train weights and biases, which represent the incremental changes learned by the network to identify the underlying characteristics of Multiple Sclerosis Lesions.

As seen by (Yoo et al. 2018), the Deep Learning approach can be combined with standalone classifiers to produce a framework in which features are identified by a Neural Network, but classified externally. This type of approach allows for the substitution of classifiers as well as feature selection techniques to assist in identifying an optimum solution.

The identification of Multiple Sclerosis Lesions from Multimodal MRI volumes using Computer Vision techniques is a complex challenge requiring a robust and transferable methodology to patients volumes outside of the domain of the training dataset. The previous section provided an up to

date background in state-of-the-art frameworks applied to Multiple Sclerosis Lesion Segmentation. The following section consists of the Methodology, which explains the proposed techniques, as well as justification for their inclusion.

Table 1: Advantages and Disadvantages of key Multiple Sclerosis Segmentation Methods.

Citation	Method	Advantage	Disadvantage	Sequence
(Zhang et al. 2008)	GLCM, GLRLM, Gradient, AR model, Wavelet transforms. Variance feature selection. Classification using Artificial Neural Network	Robust and Highly descriptive texture analysis and Feature extraction. Repeatability.	Computationally Complex. Classification using black-box software. Poor differentiation between NAWM and NWM.	T2w
(Tozer et al. 2009)	Preprocessing: normalisation (256). Haralick Texture Features from averaged GLCM. Combined HTF with categorical patient data.	Good separation between examined classes of CDMS and CIS.	One feature extraction algorithm used. The class overlap between the portion of HTF.	MTR, T2w
(Karimaghaloo et al. 2012)	Conditional Random Field Model	Highly customisable for targeted classification. The approach takes into account spatial information as well as neighbourhood.	System complexity. Small performance gain over existing SVM.	T1w, T2w, PD1, FLAIR
(Roy et al. 2013)	First Order Statistics, GLCM. Classification using Support Vector Machine. Active contour and Morphological post-processing.	Local and Global image analysis. Improved upon existing methods by 6%. Identified Lesions expanded through morphology.	Small dataset. Average F1 score (50%).	T1w, T2w, FLAIR
(Sweeney et al. 2014)	Brain tissue regions extraction and normalized, followed by voxel selection for input into a Super learner classification approach.	Fast feature extraction. Superlearning approach utilises information from all classification algorithms (LR, LDA, QDA, GMM, SVM, RF, k-NN, NN).	Primitive feature descriptors.	T1w, T2w, FLAIR
(Loizou et al. 2015)	ROI investigation including shape, statistical, Haralick, Fourier power spectrum and fractal dimension texture analysis. Statistical analysis of extracted features.	Detailed analysis of existing applications of texture features. The significant difference in MS Lesion over time. Most features can differentiate NWM from NAWM.	No application of features to classification and segmentation.	T2w

Table 1 – continued from previous page

Citation	Method	Advantage	Disadvantage	Sequence
(Michoux et al. 2015)	Haralick Texture Features and Grey Level Run Lengths. Classification using LDA, LR and PLS	All but one texture feature showed a significant difference between WM and NAWM.	High specificity, low Sensitivity. ADC failed to contribute in Lesion classification as predicted.	T1w, T2w, ADC, DWI
(Roura et al. 2015)	Skull stripping and normalisation. GM distribution acquisition followed by outlier masking for Lesion regions.	Robust and transferable to several tested datasets.	Low Classification accuracy on small volume Lesions	T1w, FLAIR
(Akbarpour et al. 2017)	Skull Stripping and Region Segmentation. Feature fusion using Wavelet and Statistical features. FCM clustering into identified regions.	High Specificity and Jaccard metrics. Best performance in specific methodology.	Susceptible to noise. Primitive FCM classification.	T1w, T2w
(Valverde et al. 2017)	Cascading CNN methodology. The first stage identifies candidate regions. The second stage removes FP voxels.	High Classification accuracy. Design of system accounts for unbalanced training data.	Requires separate training and testing for each dataset. Sensitive to changes in MRI resolution and acquisition parameters.	T2w, FLAIR
(Yoo et al. 2018)	Unsupervised deep learning procedure. Feature vector construction from excluded patches. Feature selection using LASSO. Random Forest Classification.	High Classification accuracy using Multimodal data.	Lack of training data. Unclear if the method will generalize. Reliant on preprocessing.	T1w, T2w
(Ghribi et al. 2018)	Skull stripping and preprocessing. GLCM and GLRLM features extracted from a single scan created using Volumetric Wavelet Fusion. Support Vector Machine classification.	Classification performance competes with existing MS CNN architectures. Capable of Lesion identification from noisy images.	Does not generalise well.	T1w, T2w, FLAIR, PDw
(Samah et al. 2018)	GLCM, LBP and DDP feature extraction. SVM classification.	High Classification accuracy in noisy images.	Inclusion of simulated MRI scans.	T1w, T2w, PDw

Table 1 – continued from previous page

Citation	Method	Advantage	Disadvantage	Sequence
(Aslani et al. 2019)	Modified ResNet50 Deep CNN architecture. Three plane analysis of MRI volumes, applying feature fusion from different modalities.	Multimodal approach utilising separate branches for sequence analysis, resulting in improved performance. Capable of identifying the most influential sequence (FLAIR).	High Computation time.	T1w, T2w, PDw, FLAIR

3 Methodology

3.1 Research Question

The focus of this research was to determine whether qualitative multimodal feature descriptors extracted from a database of Multiple Sclerosis Patients could correctly categorise Multiple Sclerosis Lesions from Healthy Brain Tissue regions, irrespective of the MRI sequence selected for classification. The identified biomarkers should be representative of the underlying histopathology of Multiple Sclerosis, transferable to new scans and clinical situations.

Of the methodologies analysed during the Literature Review section, this research is most comparable with those adhering to a Supervised Textural Feature Extraction technique preceding Machine Learning classification for region segmentation into binary masks for use in identifying Multiple Sclerosis Lesions.

Five textural extraction techniques are adopted for this research, each of which exhibits unique approaches in the representation of MRI greyscale texture. Classification is carried out on subsets of features selected through two feature selection algorithms, in a comparative analysis using three machine learning models.

3.2 Research Methodology

3.2.1 Patient Cohort and MRI sequences

The dataset was sourced by the University Hospital of Nancy, and contains forty scans from eight patient datasets, each containing five separate MRI sequences for training and unseen patient validation. Each patient subset consists of a set of all five sequences. No patient is captured over time, longitudinally, as this study is concerned with the analysis of Multiple Sclerosis in experimental sequences. The focus, therefore, is not on Multiple Sclerosis Lesion change, such as (Loizou et al. 2015, Köhler et al. 2019), but instead on the identification of meaningful biomarker descriptors which can be used to classify Lesions in the provided experimental DWI sequences.

Before receiving the full dataset, registration and normalisation were carried out in line with the standard clinical practice of the hospital. The sequences for each patient were fully registered volumes of 1) normalised T2-weighted, 2) Fractional Anisotropy, 3) Apparent Diffusion Coefficient, 4) Axial Diffusion and 5) Radial Diffusion volumes, details of which are outlined in section 1.3.3. Each patient sequence contained White Matter Masks, Grey Matter Masks, Lesion Normalisation Masks, and Lesion Normalisation Binary Masks.

3.2.2 Multimodal Analysis of Multiple Sclerosis

The primary aim of this research was in the development of a Multiple Sclerosis Lesion Segmentation tool for classification and extraction in Multimodal MRI data. Conventional clinical MR imaging of Multiple Sclerosis consists of Fluid-Attenuated Inversion Recovery (FLAIR), T1-weighted (T1w) and T2-weighted (T2w) sequences. Contained within the available dataset was a

single clinical sequence (T2w), whereas the remaining four sequences were more experimental to the problem of MS Lesion detection. It was therefore essential that all sequences are combined to determine correlations between features and radiomic links for the identification of meaningful biomarkers. This multimodal analysis allowed access to information, unique to each MRI sequence, which can help gain an all-around understanding of the disease (Galassi et al. 2019).

The application of multimodal feature extraction is useful in the discovery of links between features extracted from different MRI sequences (Shoshtari et al. 2016). In the proposed methodology, a multimodal analysis must be carried out to determine repeatable and durable textural links between the experimental FA, ADC, AD, RD sequences and the clinical T2w images.

3.2.3 Data pre-processing

There is a strong argument towards the inclusion of pre-processing frameworks in the development of all medical diagnosis software (Andronache et al. 2013). Pre-processing standardises raw data through intensity normalisation, outlier removal, and skull stripping to alleviate the effects of unwanted MRI artefacts and to set a baseline for greyscale ranges between patients (Dworkin et al. 2018).

In line with the intentions of stakeholders at the hospital, pre-processing would be applied sparingly and only where necessary, as opposed to a stringent framework of algorithms intended on standardising scans between patients and sequences. The primary focus of pre-processing was on the analysis of each patient volume across the five sequences to eliminate concerns regarding the presence of unwanted MRI artefacts including noise, bias inhomogeneity and physical objects such as IV lines and medical apparatus.

Addressing the concerns of (Fartaria et al. 2018) surrounding the reluctance to include automated segmentation methods in clinical scenarios, a critical focus of this research was to maintain a condition similar to that which a radiologist is presented with when examining patient scans. It was therefore essential that extensive pre-processing was removed from the methodology to avoid biasing the dataset and subsequent results over a repeatable, clinically viable solution. The proposed skull stripping algorithm consists of a reliable, fully-automated algorithm for brain extraction (Ali & Maher 2016), with no further modifications or corrections to brain tissue intensities between sequences or patient sets.

The implemented pre-processing framework is in line with the proposed methodology of (Zhong et al. 2012) who perform automatic and reproducible skull stripping through the utilisation of edge detection and morphological processing for binary mask retrieval and filling. Modifications to MRI voxel intensities were applied to background regions surrounding brain areas only, where normalisation and anonymisation were carried out by the University Hospital of Nancy. In doing so, the pre-processing stage consists only of background thresholding and skull stripping in line with the requirements of the stakeholders. The obtained brain tissue regions, now devoid of facial tissue, skull and background, are then used in the acquisition of all texture descriptors.

3.2.4 Feature Extraction

Local feature extraction was carried out utilising a sliding region window applied to all patient scans for the calculation of the proposed texture descriptors. The selection of a sliding region window reflects the concerns raised regarding the exact locations of Multiple Sclerosis boundary regions. By their very nature, MS Lesions can range in size from very small, less than 2mm, to larger volumes of greater than 5mm (Chawla et al. 2018). The ability to classify these Lesions, irrespective of size is critical to clinical assessment and treatment planning (Nair et al. 2018).

The selected size of the region window is 5x5 pixels, applied to both brain and Lesion tissue regions in the axial plane of the MRI volumes. Similar studies to our own have implemented the 5x5 region window for Multiple Sclerosis Lesion segmentation tasks (Roy et al. 2013). Similarly, (Sharma et al. 2019) discuss region window sizes stating that sizes smaller than this fail to capture enough orientation information, whilst sizes above this value (15x15 and 25x25) do not impact tissue orientation, indicating that a 5x5 region window is the smallest viable size to capture orientation information and significant detail from MRI slices.

Smaller region windows of 3x3 pixels do not provide sufficient information from analysed regions, whereas larger region windows, for example, 19x19, decreases performance as separate tissue regions blend (Rampun et al. 2016). It is, therefore, a concern that the five-by-five region window could fail to return enough information. However, in line with existing research and observations of the MS Lesions contained within the available dataset, the window of 5x5 has proven to be sufficiently sized to capture information from both small and large volume Lesions.

When performing feature extraction, the 5x5 pixel window ignores boundary regions between Lesion and Brain tissues, and Brain and background pixels. These areas share information between the classes of healthy and unhealthy and would blend information into a potential third class of hybrid healthy and unhealthy tissue (Tozer et al. 2009, Mobahi et al. 2011).

The feature extraction stage combined various supervised texture analysis techniques to provide the best opportunity of classifying Multiple Sclerosis Lesions. In line with the findings of (Theocharakis et al. 2009, Zhang et al. 2008), the inclusion of multiple statistical and textural feature extraction techniques provides a more robust analysis of regions than implementing a single feature extraction algorithm.

Supervised Radiomic Feature extraction can represent the underlying histopathology of tumour regions such as Multiple Sclerosis. There is, however, a trade-off in striking the correct balance between qualitative feature extraction to correctly identify disease and oversimplification of meaningful MRI volumes into diluted and unrepresentative image descriptors (Zhou et al. 2018).

In total, 302 features were extracted from each non-boundary tissue region of the forty patient scans. The feature extraction stage was carried out utilising the grey and white matter brain

volumes created during pre-processing for healthy brain feature extraction. Lesion feature extraction was carried out using a separate set of masks generated from the application of the provided manually segmented binary masks, to the generated brain volumes.

3.2.4.1 First Order Statistics

First-order statistics, while simple to calculate, make for highly reproducible, fast-to-compute feature descriptors. Owing to their susceptibility to changes in pixel intensity, the combination with second-order statistics such as grey-level co-occurrence matrices provides a more robust set of feature descriptors (Schwartz et al. 2012). The combination with second-order statistics also solves a significant problem with first-order statistics where location information is ignored during the calculation of pixel intensity statistics (Reischauer et al. 2018).

First-order statistics consist of simple metrics such as mean, standard deviation, skewness and kurtosis. The inclusion of standard deviation and variance can provide more significant textural information describing the spread of values within a region of interest (Gonzalez & Woods 2008). This work proposes the use of basic first-order statistics as initial local feature descriptors to be extracted from all MRI volumes.

3.2.4.2 Haralick Texture Features

Haralick proposes a set of 14 texture features derived from grey-level co-occurrence matrices, a method of describing frequency distributions of greyscale pixel occurrences in an image. The GLCM describes the joint probability of a grey-level occurring in a particular texture, or how often a tone will reside in a specific location compared to a second tone (Haralick et al. 1973).

Haralick texture features are found to be sensitive to MRI artefacts, most notably noise (Brynolfsson et al. 2017), and it is suggested that a denoising procedure should be incorporated into any methodology using these features. Furthermore, in the case of ADC images, a standardised resolution must be upheld throughout all scans, and calculated grey-level co-occurrence matrices should be fixed in resolution. The work also proposes a standardisation in grey-level resolution across sequences which conflicts with the aim of the proposed methodology and could present a significant problem with the incorporation of these features in the research.

While GLCM features are susceptible to unwanted MRI artefacts and resolution inconsistencies, significant work in Multiple Sclerosis Lesion Segmentation involves the use of grey-level co-occurrence matrices and Haralick Texture Features. These features have been applied to the classification of diseases outside of MS Lesions, such as Stroke, Glioblastoma and Haemorrhage. Their application also extends to domains outside of medical imaging such as Materials Characterization (Webel et al. 2018) and Crop Classification (Kwak & Park 2019), highlighting the resilience of these texture features to image acquisition modalities and subdivisions within.

The 14 Haralick Texture features are included in the research methodology owing to their significant usage in the field and resilient description of texture characteristics in images. Furthermore, as

highlighted by (Schwartz et al. 2012, Reischauer et al. 2018), the combination of first and second-order statistics may present a set of resilient and generalizable feature descriptors for this work.

3.2.4.3 Grey-Level Run-Length Statistics

(Galloway 1975) expanded upon Haralick's work in grey-level co-occurrence matrices to investigate tonal run lengths and texture descriptors through the analysis of adjacent pixels of the same grey tone of pixel runs within images. Unlike Haralick's GLCM, which describes the probability of occurrence within an image region, Galloway's proposed grey-level run-length matrix examines the patterns of pixel values within a region of interest. The GLRLM describes occurrences of run distances within an image (Gupta et al. 2018).

From the Literature Review section, it is clear that a significant amount of research combines Haralick Texture features with grey-level run-length statistics for texture characterisation. Using GLRLM on its own can lead to lower accuracy (Bharathi & Subashini 2014). Therefore, their inclusion in this work is based not only on their standalone performance, which still achieves a high classification rate, but also the complementary nature of these features when representing texture alongside alternate descriptors such as Haralick's texture features and First Order statistics.

There are no standout disadvantages to using GLRLM features. However, because of the similarities between Galloway's and Haralick's proposed matrices, the constraints associated with the calculation of GLCM features could apply to GLRLM features. These issues, such as requiring a standardised greyscale for optimal performance, noise reduction, and using a fixed resolution, apply to both GLCM and GLRLM calculation.

3.2.4.4 Histogram of Oriented Gradients

Histogram of Oriented Gradients has been developed and built upon by several researchers since the 1980s (McConnell 1986, Freeman & Roth 1995, Dalal & Triggs 2005). HoG features are frequently applied in Computer Vision tasks for object detection utilising intensity gradients and edge directions for shape analysis (Nabizadeh & Kubat 2015).

The Histogram of Oriented gradient separates an image into cells before calculating an orientation histogram for region description. Consequently, the performance of these features is less sensitive to varying grayscale intensities across images owing to an included normalisation step after local feature acquisition during the initial stage of the HoG algorithm (VenkateswarLal et al. 2019).

Significant Computer Vision literature has implemented HoG feature descriptors for classification problems owing to their robust ability to quantify, with a high degree of detail, local shape occurrences within images (Gonzalez-Arias et al. 2019).

Concerning this work, the use of HoG feature descriptors for Multiple Sclerosis classification using the proposed basic pre-processing strategy could prove beneficial to classification accuracy. Owing to normalisation after capturing regional information for the generation of final feature descriptors

(Arunkumar et al. 2018). The inclusion of these descriptors will provide a robust level of detail to compliment the proposed first and second-order methods.

3.2.4.5 Local Binary Patterns

Local Binary Patterns are first proposed by (Ojala et al. 1996) as a method of feature description through the extraction of greyscale invariant texture units described as binary numbers. LBP is further expanded upon by (Ojala et al. 2002) who present the concept of a multiresolution greyscale and rotationally invariant texture classification method utilising local binary patterns for feature extraction and image representation.

A key motivation for the inclusion of Local Binary Patterns in this work is their resilience to illumination changes within images, a problem significant to MRI volumes (Giacalone et al. 2018). A fundamental concern with the Local Binary Pattern descriptor relates to perspective change and scale variations within images (Pietikäinen & Zhao 2015), issues which should not directly impact the proposed research. As the pre-processing stage consists of no changes to greyscale or intensity correction, the LBP operator should prove to be a highly resilient and accurate textural feature descriptor applied in conjunction with HoG and First and second-order statistics.

In line with previous observations, combining LBP features with alternate feature description techniques have yielded greater classification accuracy than using LBP features only (Karimaghaloo et al. 2015). It is also noted that combining Local Binary Patterns with Histogram of Oriented Gradients has increased the accuracy of brain tumour detection in MRI volumes (Abbasi & Tajeripour 2017). This observation, if consistent with Multiple Sclerosis detection, could significantly improve the classification performance of the proposed research.

3.2.4.6 Examples of Texture Descriptors

Examples of texture maps from First Order statistics, Gray Level Run Lengths and Haralick Texture Features can be found in fig. 4. Reading from row one top left to right are 1,1 Lesion and Brain Sample 1,2 Minimum, 1,3 Maximum, 1,4 Mean, 1,5 Median, 2,1 Standard Deviation, 2,2 Short Run Emphasis, 2,3 Long Run Emphasis, 2,4 Low Grey Level Run Emphasis, 2,5 High Grey Level Run Emphasis, 3,1 Short Run Low Grey Level Emphasis, 3,2 Short Run High Grey Level Emphasis, 3,3 Long Run Low Grey Level Emphasis, 3,4 Long Run High Grey Level Emphasis, 3,5 Grey Level Nonuniformity, 4,1 Run Length Nonuniformity, 4,2 Run Percentage, 4,3 Angular Second moment, 4,4 Contrast, 4,5 Correlation, 5,1 Sum of Squares (Variance), 5,2 Inverse Difference Moment, 5,3 Sum Average, 5,4 Sum Variance, 5,5 Sum Entropy, 6,1 Entropy, 6,2 Difference Variance, 6,3 Difference Entropy, 6,4 Information Measure of Correlation I.

3.2.5 Feature Selecction

The implemented feature extraction methodology returns a feature vector of 302 individual descriptors, for every 5x5 brain and Lesion region within each MRI sequence volume. For every piece of brain tissue examined, there are 302 possible ways of describing that region. Feature reduction

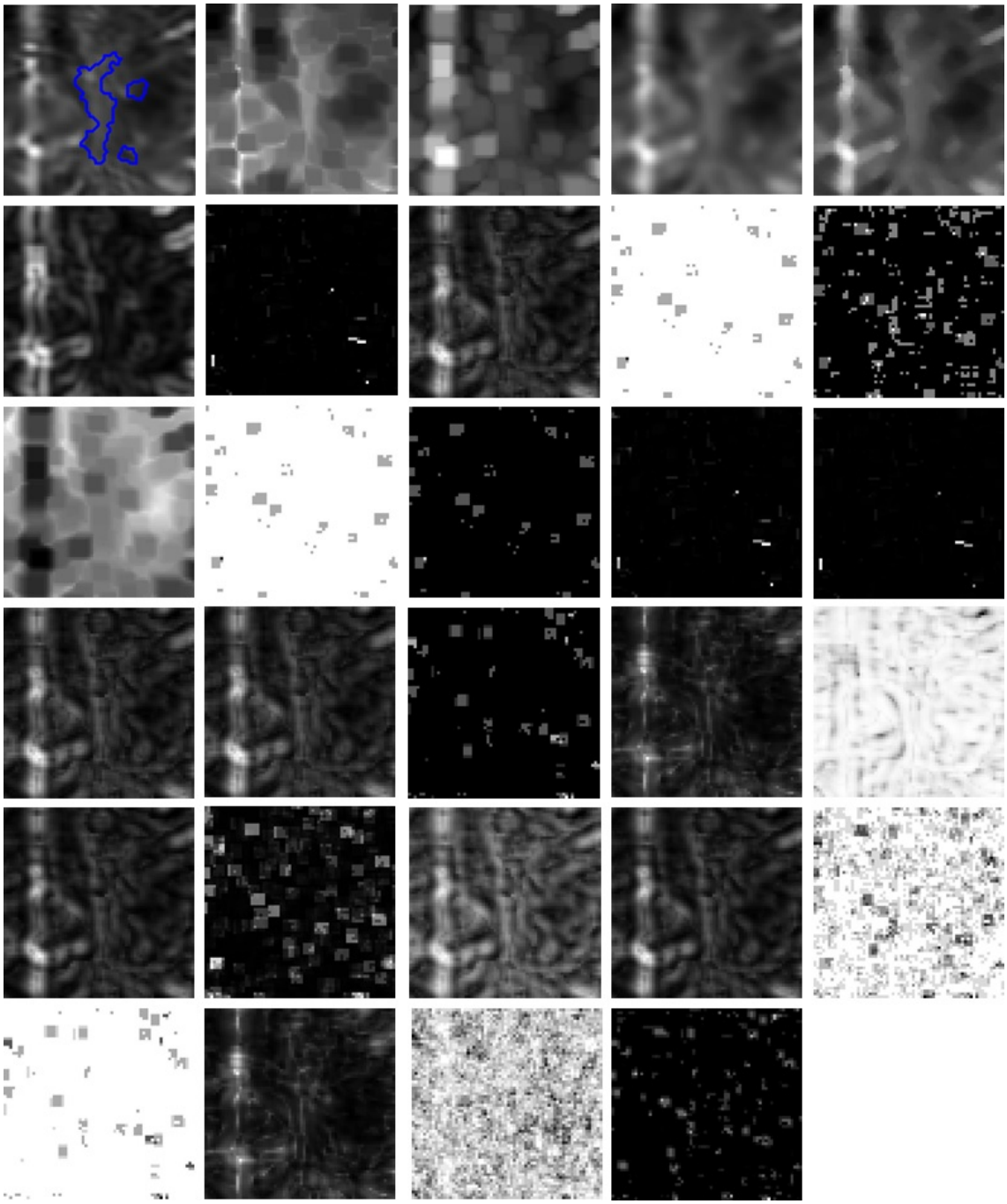


Figure 4: Texture Map of Multiple Sclerosis Lesion surrounded by healthy brain tissue, taken from T2 dataset p06c1 slice 106. Reading from the top row from left to right are the five First Order statistics, followed by 11 Gray Level Run Lengths and 12 Haralick Texture Features.

is an important step for tackling the curse of dimensionality, which is a frequent problem when using large feature vectors (Hastie et al. 2009).

Two feature selection algorithms were applied within this methodology for Multiple Sclerosis feature dimensionality reduction. Least Absolute Shrinkage Selection Operator (LASSO) and Minimum Redundancy Maximum Relevance (mRMR) algorithms.

The theory of feature selection relies upon the identification of a subset of relevant features from the global set. The new subset of features has the potential to improve classification performance and lower training times through the removal of redundant and repeated features, for use in machine learning training and testing (Cai et al. 2018). Furthermore, removing redundant features can negate overfitting training data to a model which impacts the performance of that model on unseen data, leading to better generalisation (Parmar et al. 2015, Li et al. 2017).

Feature selection algorithms fall into three categories, Filter, Wrapper and Embedded methods. Filter methods are applied as a pre-processing step isolated from machine learning algorithms. Instead, statistical characteristics of the data are analysed for feature discrimination before classification (Hancer et al. 2018, Rodriguez-Galiano et al. 2018). Wrapper methods perform variable selection taking into account the predictor performance when used in machine learning to identify a subset of descriptive features (Jiang et al. 2019). Wrapper approaches provide a black-box approach to feature selection, as well as the potential for higher accuracies in final models owing to their direct influence on the feature selection procedure (Jadhav et al. 2018). Embedded methods combine the speed of Filter methods with the Learning capabilities of Wrapper methods for the inclusion of feature selection within model training (Lu 2019).

3.2.5.1 LASSO

LASSO feature selection (Tibshirani 1996) is an embedded method selected as the first algorithm for feature selection of the proposed 302 length feature vector. LASSO can provide high classification accuracy through the removal of variables exhibiting high variance while maintaining a low bias towards particular features (Fonti & Belitser 2017). A drawback of LASSO feature selection surrounds its stepwise nature which arises when the analysing a vast number of features. This presents a large variability in the possible feature subsets, leading to reduced classification performance when uncorrelated features are selected (Morozova et al. 2015).

LASSO feature selection has been applied alongside 5-fold cross-validation using Support Vector Machine for classification of glioblastomas in T2 MRI data resulting in high classification accuracy of around 83% (Qian et al. 2019). Interestingly these authors noted that the inclusion of multimodal MRI sequences could provide even greater accuracy when applying LASSO and SVM algorithms. Similar research in tumour classification again combined LASSO with an SVM classifier using 10-fold cross-validation and grid search for LASSO parameter tuning, which exhibited 100% classification accuracy (Kang et al. 2019).

3.2.5.2 mRMR

mRMR feature selection (Hanchuan Peng et al. 2005) is a filter method implemented as the second feature selection algorithm in this research. mRMR identifies the relevance and redundancy of a variable through the examination of mutual information between multiple variables (Sheng et al. 2018). One significant drawback of mRMR lies in its sensitivity to outliers, which have the potential to be selected as features owing to their differentiable qualities (Kalina & Schlenker 2015).

mRMR has been applied in a range of medical imaging problems such as brain tumour detection (Zhao et al. 2019, Kim et al. 2018) with high classification accuracy generated from training data selected using mRMR feature rankings. Furthermore, combining mRMR with Support Vector Machines has exhibited high classification accuracy, especially sensitivity to disease in the identification of heart disease (Haq et al. 2018).

The rankings from both feature selection algorithms are applied as indexes for training dataset construction into the supervised machine learning stage. These rankings not only provide the possibility of reducing computation time and increasing classification accuracy, but can also help negate the necessity to capture the full range of proposed features upon completion of model training for increased performance in the final system.

3.2.6 Supervised Machine Learning

When designing a machine learning methodology, consideration must be given to the type of classification algorithm implemented. Supervised classification requires labelled data describing the disease location for feature extraction and categorisation. Unsupervised classification does not require labelled information to understand diseased regions and instead identifies features describing the whole image, for example, CNN and Deep Learning architectures. Semi-supervised machine learning infers meaning on missing information extracted from labelled images for classification (Vial et al. 2018).

In the determination of the correct classification approach for this research, consideration was given to the available Multiple Sclerosis Data. Deep Learning methodologies such as Convolutional Neural Networks have been successfully applied to Multiple Sclerosis Segmentation problems, as covered in the literature review section. These methodologies, while robust and successful, present limitations. Most notable is the amount of training data required for learning, often tens of thousands of images for common diseases (Lam et al. 2018). This problem becomes more complicated with diseases such as Multiple Sclerosis where the amount of training data required from the non-clinical DWI sequences investigated in this research does not exist.

3.2.6.1 Support Vector Machine

Support Vector Machine is the first classification algorithm used in this research. SVM has been chosen for its previous success in the classification of Multiple Sclerosis (Abdullah et al. 2012, Zurita et al. 2018) and other degenerative brain diseases, often combining features such as GLRLM

(Öztürk & Akdemir 2018), GLRL (Jafarpour et al. 2012), LBP and HoG descriptors (Amin et al. 2018). The SVM proposed in this methodology utilises the RBF kernel for non-linear classification and is optimised using a grid search and 5-fold cross-validation, similar to the experimental setup of (Sørensen & Nielsen 2018).

SVM (Boser et al. 1992, Cortes & Vapnik 1995) is a classification model which seeks to identify a separating hyperplane which functions as a decision boundary between two classes (e.g. Healthy and Unhealthy). Primitively speaking any line which separates the data classes could function as the hyperplane. However, the SVM aims to find a hyperplane such that the points closest to the line (Support Vectors) for each class are maximised. This region between the hyperplane and support vectors is called the margin. The objective of this region is to provide as much distance between classes as possible for use in classifying future data points.

The nature of SVM's linearly separating hyperplane presents an obvious problem, its inability to classify non-linear data. Support Vector Machine adopts the "kernel trick" to remap data into a third z dimension for linear separation (Yang et al. 2019). A noted disadvantage of the SVM is computational cost and unreliable accuracy when presented with low amounts of training data containing large feature vectors (Sharma et al. 2016)). With regards to SVM and Multiple Sclerosis, the class imbalance problem is a potential risk owing to the susceptibility of medical imaging problems to unhealthy minority samples and majority healthy samples (Zhu, Xia, Jin, Yan, Cai, Yan & Ning 2018).

3.2.6.2 Random Forest

Random Forest is the second machine learning model implemented as part of this research. Random Forests have been applied to various domains such as Human detection (Jalal et al. 2019) and Medical Imaging, including Multiple Sclerosis (Wei et al. 2018, Lötsch et al. 2018). Random Forest was developed through expanding the work of (Tin Kam Ho 1995) who analysed Random Decision Forests. (Breiman 2001) further expanded upon this, suggesting the random selection of features at each tree node as well as out-of-bag estimations deduced from classification results and error rates.

Random Forests are stable algorithms for classification owing to the employment of multiple trees in decision making, often leading to SVM outperformance (Speiser et al. 2019). They are insensitive to large feature vectors and only require a small amount of tuning such as the number of trees in the forest (Xia et al. 2018). Furthermore, setting an incorrect number of trees in a forest does not significantly impact the performance of the classifier (Thanh Noi & Kappas 2018), leading to faster optimisation time and a less complicated tuning procedure.

The significant disadvantage of Random Forest lies in their complexity. The version implemented in this research utilised 200 decision trees within the model, leading to a high level of complexity in both construction and decision making. Owing to the ensemble nature of the RF classifier in which classification is achieved through isolated prediction by each decision tree in the forest before

aggregated voting decides the outcome. The number of trees in that forest directly impacts the performance and complexity of the overall model.

3.2.6.3 Extreme Learning Machine

Extreme Learning Machine (Huang et al. 2006, Huang et al. 2012) is the final classification algorithm implemented in this research. Deep Learning methodologies learn utilising backwards propagation for cost minimisation, performed during each training iteration. Extreme Learning Machines do not employ backwards propagation, and instead, weights and biases are chosen randomly and are trained using a least-squares solution from the linear system (Peng et al. 2013).

The success of ELM stems from the Universal approximation theorem in which a feed-forward network with a single hidden layer and a finite amount of neurons can approximate continuous functions with little information on the activation function. Owing to the random allocation of nodes as opposed to backwards propagation during the learning phase, ELM has much faster classification performance and generalises well with unseen data samples (Yu et al. 2019).

The reasoning for the inclusion of the Extreme Learning Machine in this work is two-fold. Firstly, ELM has been successfully applied to a variety of medical imaging problems such as brain tumour detection (Gumaei et al. 2019) and cancer detection using a modified GLCM algorithm (Zhang et al. 2018), as well as classification using additional texture features (Song et al. 2018).

Secondly, there is a significant lack of research in applications of ELM for textural analysis of Multiple Sclerosis, at the time of writing. Existing work has been carried out by Zhang et al. (2017) who perform Multiple Sclerosis classification utilising the online library OS-ELM to evaluate features extracted from MRI volumes using Canny edge detection followed by fractal dimension analysis. This gap in research presents a significant opportunity to assess the performance of Extreme Learning Machine Classification of Multiple Sclerosis in Multimodal MRI data.

3.3 Tools and Performance Evaluation

3.3.1 Tools

The Software has been developed utilising a Windows 10 Machine running MATLAB 2018a 64bit.

Feature Selection Methods are applied using the Feature Selection Toolbox provided by (Roffo 2016, Roffo & Melzi 2016, Roffo et al. 2017), available from `https://uk.mathworks.com/matlabcentral/fileexchange/56937-feature-selection-library`

Haralick Texture Features are applied using a modified version of the MATLAB package provided by (Monzel 2007).

Grey Level Run Lengths are applied using a modified version of the MATLAB package provided by (Wei 2007).

3.3.2 Performance Evaluation

Binary classification performance can be expressed in terms of a Confusion Matrix where columns represent predicted responses and rows represent the actual labels.

1. *True Positive:* the number of positive samples identified as positive by a binary classifier.
2. *False Positive:* the number of negative samples identified as positive by a binary classifier.
3. *True Negative:* the number of negative samples identified as negative by a binary classifier.
4. *False Negative:* the number of positive samples identified as negative by a binary classifier.

3.3.2.1 Sensitivity

Sensitivity denotes the proportion of positives which are correctly predicted as positive. The sensitivity of a trial indicates the effectiveness of that test in identifying patients having a disease.

$$\frac{TP}{TP + FN} \tag{1}$$

A test with 100% sensitivity identifies all patients as having a disease. However, this outcome is not specific. It may be the case that 20% of those patients were healthy.

3.3.2.2 Specificity

Specificity denotes the proportion of negatives which are correctly predicted as negative. The specificity of a trial indicates the effectiveness of that test in identifying patients without a disease (Lalkhen & McCluskey 2008).

$$\frac{TN}{TN + FP} \tag{2}$$

A test with 100% specificity identifies all patients as being healthy.

3.3.2.3 Accuracy

Accuracy is a measure of the success of a test in the classification of a disease. The average of correct classifications is taken comparing predicted positive and predicted negative values over the total number of samples classified in the study.

$$ACC = \frac{TP + TN}{P + N} \tag{3}$$

3.3.2.4 BCR

Balanced Classification Rate is the second measure of the accuracy of a test and is considered a more reliable metric than ACC. BCR takes into account both positive and negative classes through TPR and TNR. The BCR metric applied the sensitivity and specificity of a test when calculating the overall accuracy, describing the average of the proportion of correct classifications of each class.

This metric is particularly useful for unbalanced datasets.

$$BCR = \frac{1}{2}(TPR + TNR) = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{4}$$

The BCR metric is generally accepted as the better method of representing accuracy owing to the ability to account for unbalanced datasets unlike the ACC metric.

3.3.2.5 BER

Balanced Error Rate is the complement of the BCR score and details the average of the proportion of misclassifications of each class.

$$BER = 1 - BCR \tag{5}$$

3.3.2.6 Precision

Precision describes the proportion of correct positive predictions and is calculated as the number of true positive values over the total number of positive predictions.

$$Precision = \frac{TP}{FP+TP} \tag{6}$$

3.3.2.7 Recall

Recall describes the proportion of actual positives which are correctly identified. It is calculated as the number of true positive values over the summation of true positive and false negative.

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

3.3.2.8 F1 score

F1 score is the harmonic mean between precision and recall. The F1 metric describes the model's accuracy with a higher score indicating low false positives and low false negatives.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{8}$$

3.3.3 Similarity Coefficients

The similarity is an important concept in the determination of the cardinality of a set taking into account the Union and Intersection. Two key metrics are the Sorensen-Dice coefficient and the Jaccard coefficient, which both seek to identify shared membership between two sets. A higher score from these metrics indicates that the similarity between the two is greater, with a score of 100 identifying exactness between the groups (Glen 2016).

Similarity coefficients have been applied in this work to predicted binary masks containing possible MS Lesion locations. These binary masks are compared against manual segmentation masks using Sorensen-Dice, and Jaccard coefficients as similarity measures to indicate the accuracy of trained models.

3.3.3.1 Sorensen-Dice coefficient

$$SDC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{9}$$

3.3.3.2 Jaccard coefficient

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{10}$$

3.4 Proposed System

The implemented Multiple Sclerosis Classification and Segmentation approach consists of dataset generation from raw MRI volumes for construction of brain images devoid of CSF, Skull and anonymous background noise. These volumes are then used for feature extraction utilising a 5x5 region window with a stride of 3 pixels for brain feature extraction to capture First Order statistics, Haralick texture features from GLCM, Galloway's grey-level run-length statistics from GLRLM, Local Binary Patterns and Histogram of Oriented Gradients.

Basic machine learning training and testing are carried out on Individual MRI sequences to assess the performance of each Machine Learning algorithm in a set of less complex experiments. This stage also identified an ideal sample size across each sequence for use in constructing a Multimodal training dataset.

The Multimodal training dataset is then used in a 5-fold cross-validation procedure for feature selection utilising mRMR and LASSO algorithms, with an integrated LASSO tuning stage. The cross-validation stage also includes Support Vector Machine, Random Forest and Extreme Learning Machine tuning followed by a separate stage of final model creation for each of the three classifiers utilising settings identified from the 5-fold cross-validation procedure. The Multimodal training stage also carries out further investigation into the ideal sample size, using the individual sequence size as a baseline, and five times this value as the upper limit.

The final models are used in the evaluation of unseen patient data, which has been separated from the cross-validation and model training procedures. An overview diagram of the entire system is shown in fig. 5.

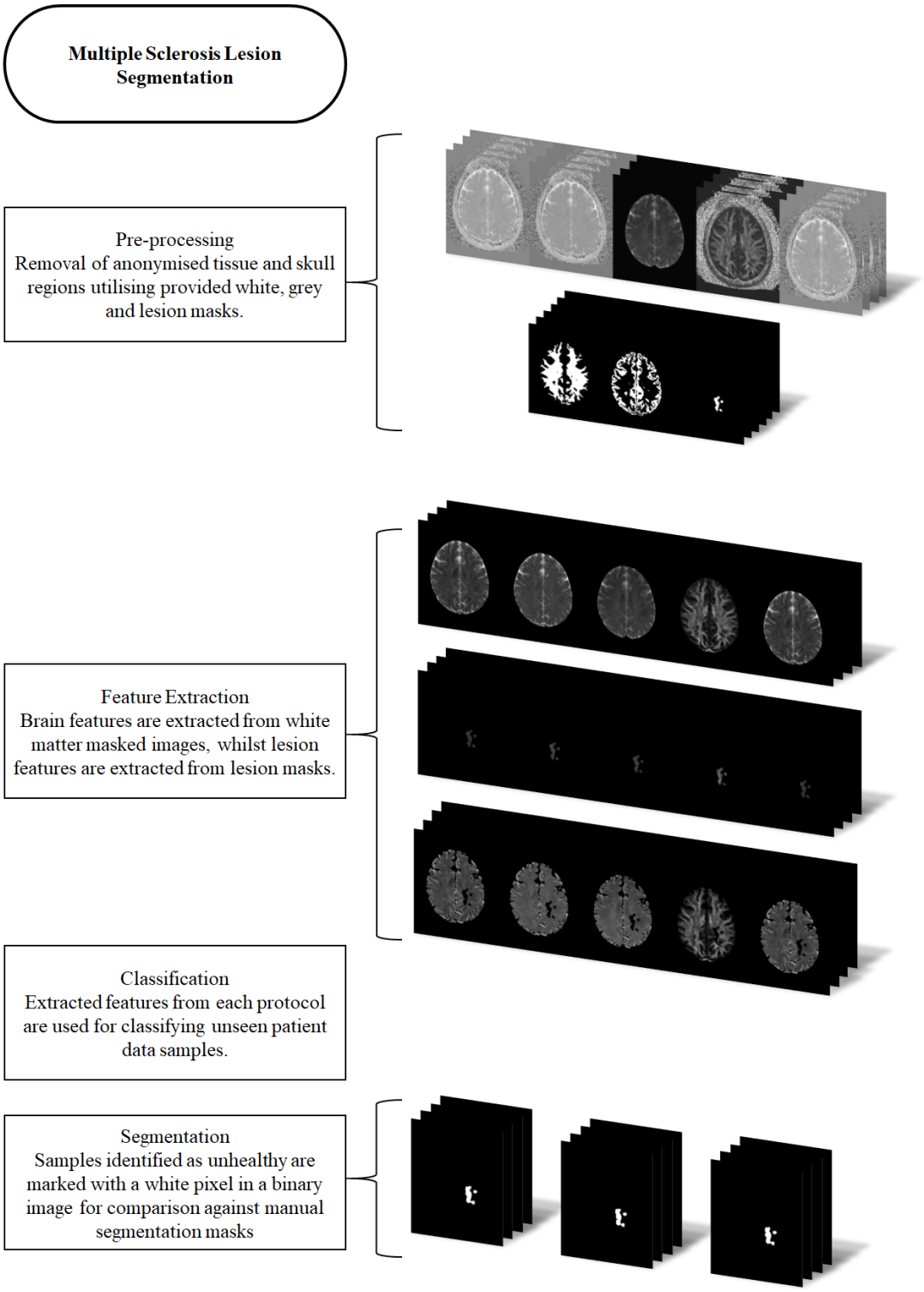


Figure 5: Overview of system

4 Methods

4.1 Data pre-processing

4.1.1 Skull stripping

As suggested by (Abbasi & Tajeripour 2017), the skull stripping algorithm should be applied to an MRI sequence which exhibits clear boundary lines between brain tissue and the inner skull regions. In their work, the T1w scan was selected. The proposed research utilised the clinical T2w modality for skull removal owing to the similarities between the T1w and T2w sequences, and the clarity of boundaries in the available dataset.

Two Brain Tissue Segmentation techniques are implemented and applied to T2w sequences from each patient.

1. The first more straightforward technique combined the provided White Matter, Grey Matter and Lesion masks for each patient set to produce a full set of brain volumes void of Cerebral Spinal Fluid. Volumes depicting only Lesion regions were produced using the Lesion Binary Masks applied to each entire MRI volume. The Lesion Masks were then subtracted from the brain tissue regions for patient volumes void of Multiple Sclerosis Lesions for use in brain sample extraction.

2. A histogram-based, morphological algorithm was developed to remove the skull and facial tissue regions for the segmentation of complete brain volumes including White Matter, Grey Matter, Lesions and Cerebral Spinal Fluid. This approach threshold the image using Otsu-thresholding before eroding each image to seperate tissue from skull regions. The last step identified the largest component in each image and selected this to be the brain region.

During both approaches, the background regions surrounding the brain were set to a threshold of zero to be solid black. This would reduce the complexity when locating the brain areas during feature extraction.

Brain volumes created during the first approach were used for the feature extraction stage in which information was extracted from healthy brain and unhealthy Lesion regions, while the second approach was implemented for manual inspection, viewing and final results presentation as these volumes contained the entire brain volume.

4.2 Feature Extraction

The feature extraction stage combined various techniques of texture analysis to provide the best opportunity of classifying Multiple Sclerosis Lesions. Evaluating the accuracy in both Multimodal MRI as well as scans containing unwanted artefacts, it was found by (Theocharakis et al. 2009, Zhang et al. 2008) that a combination of feature extraction techniques provides a more robust analysis of regions than implementing a single feature extraction algorithm.

4.2.1 First Order Statistics

The first five elements of the feature vector comprised simple statistics captured from the sliding 5x5 region window of healthy and unhealthy tissue regions.

The minimum is the pixel whose value is lowest in the extracted 5x5 block.

$$Minimum = i|p(i) > 0 and \nexists\ j < i|p(j) > 0 \tag{11}$$

The maximum is the pixel whose value is the highest in the extracted 5x5 block.

$$Maximum = i|p(i) > 0 and \nexists\ j > i|p(j) > 0 \tag{12}$$

The mean corresponds to the average value within the extracted 5x5 block.

$$Mean = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{13}$$

The median corresponds to the middle value within the extracted and sorted 5x5 block.

$$Median = P_{sorted}(N/2) \tag{14}$$

The standard deviation is an indication of the dispersion of values within the 5x5 block from the mean.

$$StandardDeviation = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}} \tag{15}$$

4.2.2 Haralick Texture Features

Grey level co-occurrence matrices are calculated in four directions Θ: 0, 45, 90, 135 degrees with d distance representing the distance between origin and calculation point, e.g. d = 1 for immediate neighbour calculation. An average co-occurrence matrix was then calculated for rotation invariance. The 14 Haralick texture features were then extracted from the mean GLCM for rotationally invariant texture description.

Haralick texture features can be perceived as a spatial distribution of intensities for use in identifying similar regions across different images. The 14 features are explained in greater detail below.

Angular Second Moment

Angular Second Moment is the measure of homogeneity in an image region. For example, a homogenous area has a more uniform texture, whereas a heterogenous area will have more extreme intensity changes.

$$f1 = \sum_{i=1}^{N} \sum_{j=1}^{N} p(i,j)^2 \tag{16}$$

Contrast

Contrast is the measure of variation in the image region. Higher contrast will exhibit greater variations.

$$f2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 p(i,j) \tag{17}$$

Correlation

Correlation is the measure of linear dependence in grey levels of a neighbourhood region. A correlation of 0 indicates that there is no correlation between pixels, whereas a correlation of -1 or 1 indicates negative and positive correlation.

$$f3 = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(i.j)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{18}$$

Sum of Squares (Variance)

The Sum of Squares (Variance) indicates the deviation from the mean of each pixel within the image region, summed together. A low Sum of Squares shows that the region may be uniform as all values reside around the mean.

$$f4 = \sum_{i=1}^{N}\sum_{j=1}^{N}(i.j)(i-\mu)^2 p(i,j) \tag{19}$$

Inverse Difference Moment

Inverse Difference Moment indicates the local homogeneity. The value is high when a pixel region is uniform.

$$f5 = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{p(i,j)}{1+|i-j|} \tag{20}$$

Sum Average

$$f6 = \sum_{k=2}^{2N} k p_{x+y}(k) \tag{21}$$

Sum Variance

$$f7 = \sum_{k=2}^{2N}(k-\mu_{x+y})^2 p_{x+y} \tag{22}$$

Sum Entropy

$$f8 = -\sum_{k=2}^{2N} p_{x+y}(k)\log_{p_{x+y}}(k) \tag{23}$$

Entropy

Entropy is the statistical measure of randomness contained within a pixel region. Entropy will be high in regions exhibiting nonuniformity and sparsity in patterns.

$$f9 = -\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)\log(p(i,j)) \tag{24}$$

Difference Variance

$$f10 = \sum_{k=0}^{N-1}(k-\mu_{x-y})^2 p_{x-y} \tag{25}$$

Difference Entropy

$$f11 = -\sum_{k=0}^{N-1} p_{x-y}(k)\log_{p_{x-y}}(k) \tag{26}$$

Information correlation coefficient refers to the dependency between two random variables and describes the amount of information in one variable compared to another.

Information Measures of Correlation I

$$f12 = \frac{HXY - HXY1}{max(HX, HY)} \tag{27}$$

Information Measures of Correlation II

$$f13 = \sqrt{1 - \exp\left[-2(HXY2 - HXY)\right]} \tag{28}$$

where:

$$HXY1 = -\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)\log\{p_x(i)p_y(j)\} \tag{29}$$

$$HXY2 = -\sum_{i=1}^{N}\sum_{j=1}^{N} p_x(i)p_y(j)log\{p_x(i)p_y(j)\} \tag{30}$$

Maximal Correlation Coefficient

$$f14 = (Second\ largest\ eigenvalue\ of\ Q)^{1/2} \tag{31}$$

where:

$$Q(i,j) = \sum_{k=1}^{N}\frac{p(i,j)p(j,k)}{p_x(i)p_y(k)} \tag{32}$$

4.2.3 Grey Level Run Length Statistics

The grey level run length matrix describes the number of occurrences in which a pixel of a particular greyscale runs and for how pixels this run iterates. A grey level run length matrix is defined such

that the y-axis indicates greyscale, and the x-axis displays the run length. Similar to Haralick's grey level co-occurrence matrix, the grey-level run-length matrix is captured in four directions Θ: 0, 45, 90, 135 degrees. An average GLRLM can then be calculated for rotation invariance.

(Galloway 1975) proposes 11 textural descriptors which are then derived from the mean GLRLM for calculation of rotationally invariant grey level run length statistics.

Short Run Emphasis

Short Run Emphasis indicates the number of shorter runs within an image region. Shorter runs usually indicate finer textures.

$$SRE = \frac{1}{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{GLRLM(i,j)}{j^2} \tag{33}$$

Long Run Emphasis

Long run emphasis indicates the number of longer runs within an image region. Longer runs usually indicate structural regions and coarse textures.

$$LRE = \frac{1}{H}\sum_{i=1}^{N}\sum_{j=1}^{N}GLRLM(i,j)\cdot j^2 \tag{34}$$

Low Grey Level Run Emphasis

Low Grey Level Run Emphasis analyses the spread of low grey level values and returns a higher value for areas containing more low grey level values.

$$LGRE = \frac{1}{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{GLRLM(i,j)}{i^2} \tag{35}$$

High Grey Level Run Emphasis

High Grey Level Run Emphasis analyses the spread of high grey level values and returns a greater value for areas containing more high grey level values.

$$HGRE = \frac{1}{H}\sum_{i=1}^{N}\sum_{j=1}^{N}GLRLM(i,j)\cdot i^2 \tag{36}$$

Short Run Low Grey Level Emphasis

Short Run Low Grey Level Emphasis examines the joint distribution of short runs with low grey-level values.

$$SRLGE = \frac{1}{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{GLRLM(i,j)}{i^2\cdot j^2} \tag{37}$$

Short Run High Grey Level Emphasis

Short Run High Grey Level Emphasis examines the joint distribution of short runs with high grey-level values.

$$SRHGE = \frac{1}{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{GLRLM(i,j)\cdot i^2}{j^2} \tag{38}$$

Long Run Low Grey Level Emphasis

Long Run Low Grey Level Emphasis examines the joint distribution of long runs with low grey-level values.

$$LRLGE = \frac{1}{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{GLRLM(i,j)\cdot j^2}{i^2} \tag{39}$$

Long Run High Grey Level Emphasis

Long Run High Grey Level Emphasis examines the joint distribution of long runs with high grey-level values.

$$LRHGE = \frac{1}{H}\sum_{i=1}^{N}\sum_{j=1}^{N}GLRLM(i,j)\cdot i^2\cdot j^2 \tag{40}$$

Grey Level Nonuniformity

Grey Level Nonuniformity analyses the similarity of pixels within a region and returns a higher value for areas of differing texture.

$$GLNU = \left(\frac{1}{H}\sum_{i=1}^{N}\left(\sum_{j=1}^{N}GLRLM(i,j)\right)^2\right) \tag{41}$$

Run Length Nonuniformity

Run Length Nonuniformity analyses the similarity of run lengths within an image region and returns a higher value for more heterogenous, or differing run lengths.

$$RLNU = \left(\frac{1}{H}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}GLRLM(i,j)\right)^2\right) \tag{42}$$

Run Percentage

Run Percentage analyses the coarseness of image texture and returns a higher value in situations where an image region contains many short runs, indicating a finer texture.

$$RP = \left(\frac{H}{\sum_{i=1}^{N}\sum_{j=1}^{N}(j\cdot GLRLM(i,j))}\right) \tag{43}$$

4.2.4 Local Binary Patterns

Local Binary Patterns are first proposed by (Ojala et al. 1996) as a method of feature extraction through the extraction of greyscale invariant texture units described as binary numbers. The theory behind Local Binary Patterns is that a neighbourhood of pixels, for example, 3x3, are evaluated, beginning at the top-left pixel, against the central point in that neighbourhood. Pixels

of greater or equal intensity to the centroid are represented as 1, whilst those less than are set to 0. The Binary number is then converted to decimal and stored within a second map of Local Binary Pattern values from which histogram binning is applied to produce a 256-length feature vector.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{44}$$

$$s(x) = \begin{cases} 1, & \text{if } a = 1 \\ 0, & \text{otherwise} \end{cases} \tag{45}$$

The local binary pattern is further expanded upon by (Ojala et al. 2002) who present the concept of a multiresolution grey-scale and rotationally invariant texture classification method utilising Local Binary patterns for feature extraction and image representation. The proposed operators build from initial work from the 1996 paper in Local Binary Patterns and seek to tackle the problem of grey-scale and rotation invariance. All proposed methods build from the $LBP_{P,R}$ algorithm.

$$LBP^{rui2}_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c), & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1, & \text{otherwise,} \end{cases} \tag{46}$$

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{P-1} - g_c)| \tag{47}$$

$LBP^{riu2}_{(P,R)}$ builds from this concept to achieve rotation invariance from uniform grey levels through identifying the number of changes in pixel intensity around the central point. Each pixel is calculated as 1 or 0 depending upon the number of uniform changes to obtain rotation invariance with less than two changes leading to the calculation of signed differences for that neighbourhood rotation.

The uniform change $U(LBP_{P,R})$ iterates over each pixel in the neighbourhood and compares the right-hand neighbour. If this neighbour holds the same value as the current starting pixel, they are said to be uniform. Through identifying the number of changes around the binary word, a new word is generated, describing the number of changes within the originally calculated binary pattern. The resultant vector is a rotationally invariant descriptor.

$LBP^{riu2}_{P,R}$ provides spatial and structural description of regional texture, and its combination with $VAR_{P,R}$ was found to elicit both structural and contrast information for excellent feature descriptors. The basic premise of Local Binary Patterns concerns building binary numbers from pixel neighbourhoods through the examination of each circular value about the central point of that region. A sign is created for each point relative to the subtraction of that value from the centre. The binomial factorisation of these binary words generates a unique $LBP_{P,R}$ number representing spatial structure of that neighbourhood.

$$VAR_{P,R} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \tag{48}$$

where:

$$\mu = \frac{1}{P}\sum_{p=0}^{P-1} g_p \tag{49}$$

$VAR_{P,R}$ is invariant against changes in greyscale unlike $LBP^{riu2}_{P,R}$ whose output is unaffected by monotonic transformations in intensity, meaning that contrast is ignored entirely. $VAR_{P,R}$ examined the local variance and presents a set of rotationally invariant descriptors for use in the grey-scale invariant representation of image regions. This operator iterates across the pixel region and calculates the summation of the current neighbourhood pixel minus the neighbourhood mean squared to produce the variance of each rotation.

Owing to the complementary relationship between $LBP^{riu2}_{P,R}$ and $VAR_{P,R}$ a joint distribution of $LBP^{riu2}_{P,R}/VAR_{P,R}$ can be implemented for highly descriptive, rotationally invariant feature extraction of local region texture comprising the whole image.

4.2.5 Histogram of Oriented Gradients

Histogram of Oriented Gradients has been developed and built upon by several researchers since the 1980s. (Freeman & Roth 1995) propose Orientation Histograms for Hand Gesture recognition, which builds on patented work by (McConnell 1986) in the development of a histogram-based algorithm for orientation analysis. Freeman and Roth simplified the histogram comparison section of McConnell's work and proposed the squared error measure as a replacement.

In line with (Dalal & Triggs 2005), the theory behind Histogram of Oriented Gradients begins with gradient detection using a convolution operation and gradient filter such as the Sobel filter to obtain the gradient representation of the original image. Magnitude and direction of gradients are then calculated utilising the following formula.

$$g = \sqrt{g_x^2 + g_y^2} \tag{50}$$

$$\Theta = \arctan\frac{g_y}{g_x} \tag{51}$$

Calculation of magnitude and direction is carried out over subdivisions of the original image, for example, 8x8 blocks. In the calculation of a blocks individual feature vector, the gradient and magnitude matrix of that region are used to store the magnitude values into bins defined by the gradient direction. The extracted feature vector consists of binned values, the length of which is dependent upon the size of the examined pixel region.

It has been found that grey level co-occurrence Statistics and grey level run length statistics can be used effectively in the classification of White Matter Lesions. While Histogram of Oriented Gradients and Local Binary Patterns are less susceptible to bias field inhomogeneities (Nogueira et al. 2017). Therefore, within this research, a combination of feature extraction techniques have been implemented into this methodology for Multiple Sclerosis Lesion Segmentation. The feature

vector was comprised of five First Order Statistics, 14 Haralick Texture Features, 11 Grey Level Run Length Statistics, 9 Histogram of Oriented Gradient Statistics and 256 Local Binary Pattern features.

4.3 Classification

Classification of the feature vectors was carried out using three Machine Learning models. Two models, Support Vector Machine and Random Forest, have been significantly applied to Multiple Sclerosis Segmentation tasks. The third model, an Extreme Learning Machine, has yet to be applied to Texture based Multiple Sclerosis Lesion Segmentation.

4.3.1 Support Vector Machine

Support Vector Machine has two tunable parameters, C or Box Constraint, and γ or Kernel Scale. The implemented Support Vector Machine algorithm was provided within the Statistics and Machine Learning Toolbox provided by MATLAB 2018a. Documentation on the SVM contained within this package can be found at (Mathworks 2019*a*).

4.3.2 Random Forest

The Random Forest algorithm has three parameters for tuning. These are the number of trees in the decision forest, the number of predictors to sample and the number of samples per leaf. The implemented Random Forest algorithm was provided within the Statistics and Machine Learning Toolbox provided by MATLAB 2018a. Documentation on the RF contained within this package can be found at (Mathworks 2019*b*).

4.3.3 Extreme Learning Machine

The Extreme Learning Machine algorithm has three parameters for tuning. These are the N0 or Number of initial training data used during initial stages. The Block or size of data learned at each step, and the number of hidden neurons in the ELM parameter. The implemented Extreme Learning Machine algorithm utilised the OS-ELM library available from (NTU 2013), and was modified for integration into the MS segmentation system.

4.4 Experimental Setup

Two initial experiments are proposed for evaluating the described Multiple Sclerosis Segmentation methodology. The initial approach is designed to evaluate the performance of features extracted from and classified on single MRI sequences. For example, an SVM model trained on T2 features, and used to classify unseen T2 scans. The second approach combines features taken from all MRI sequences for multimodal classification and segmentation of unseen patient datasets. In both experiments, models are trained using features extracted from all patient sets except a single, leave-out patient dataset, which is used as the unseen testing set for Lesion Classification and Segmentation.

As well as these two initial experiments, a set of secondary experiments were designed to understand the classification performance and relationship between groups of features from the implemented feature extraction techniques. This second set of experiments also carries out an Individual MRI sequence and Multimodal set of experiments in order to assess the extent to which different sequences assist or hinder classification performance.

4.4.1 Training samples and data split

In calculating an optimal sample size for use in final model training it was imperative that the chosen number of samples achieved good classification performance whilst remaining unbiased to varying MRI sequences and machine learning models. To achieve this aim a grid search was devised to search the sample size space between 500 and 7000 samples using an increment of 100 samples. In doing so, the results of these experiments could be analysed in order to identify a sample size which performed consistently across MRI sequences without overfitting any machine learning model to the data.

A significant consideration was whether to include different numbers of healthy and unhealthy samples to compensate for the common problem in MS Lesion detection methodologies where datasets are dominated by healthy samples. In the available dataset, only 4.68% of available samples represent MS Lesion, and when factoring in the possibility of noise or other MRI artefacts the chances of all unhealthy samples operating as meaningful MS biomarkers reduces.

Within this research an equal split of healthy and unhealthy samples is used owing to the complex task of identifying an applicable and unbiased number of samples for both healthy and unhealthy classes when attempting to train machine learning models using unbalanced datasets. Furthermore, the implementation of an equally balanced split can lead to increased sensitivity to MS Lesion during testing, however, there is also the risk of increasing False Positive classification through undersampling the healthy class (Valverde et al. 2017).

4.4.2 K-fold Crossvalidation

K-fold cross-validation was carried out during model creation in order to partition the dataset into separate training and validation sets. The model is trained on 80% of the data and tested on the remaining 20%. This process is repeated until all data has been used in both training and validation. Applying cross-validation leads to reduced data variance and a reduction in bias and overfitting as the procedure ensures performance is uniform across all folds (Saccà et al. 2018). Within this research, 5-fold cross-validation is applied to sample size calculation and feature selection in order to calculate sample size, perform feature selection tuning and to avoid model overfitting by leaving out portions of training data.

Pseudocode for the k-fold procedure is shown in algorithm 1 highlighting the sample size loop, k-fold loop and feature selection and model training stages.

Algorithm 1 K-fold cross-validation overview

```
1: procedure K-FOLD(T, f)                ▷ where T is Training Data and f is number of folds
2:     for n = 100 : 100 : length(T) do   ▷ where n is the current sample size of features for training
3:         partitionStart = 1
4:         partitionEnd = n/f
5:         for i = 1 : f do
6:             TrainingSet = T(1:n)
7:             leaveoutSet = TrainingSet(partitionStart:partitionEnd)
8:             TrainingSet(partitionStart:partitionEnd) = [];
9:             Feature Selection;
10:            Model Training;
11:            partitionStart = partitionEnd + 1
12:            partitionEnd = partitionEnd + n/f
13:        end for
14:    end for
15: end procedure
```

4.4.3 Individual Sequence Classification

LASSO and mRMR rankings are calculated using features extracted from a single MRI sequence. These rankings are then used to train Support Vector Machine, Random Forest and Extreme Learning Machine classifiers for each feature selection method. An unseen patient dataset from the same MRI sequence is then provided to each classifier for Multiple Sclerosis classification and Lesion Segmentation.

Individual classification also assists with Multimodal classification by providing a baseline number of samples to include in model training. This stage also indicates basic tuning parameters such as LASSO λ values per sequence. Individual classification also provides critical information to the Multimodal section through the feature selection rankings outputted for each MRI sequence. The top 25 rankings from each feature selection algorithm operating on every MRI sequence are combined into two rankings (LASSO and mRMR) for use in Multimodal classification.

4.4.4 Multimodal Classification

Multimodal classification takes the top 25 LASSO and mRMR rankings from the previous classification stage and performs both feature selection algorithms once more on these feature ranking vectors which are identified as being representative of the underlying histopathology of Multiple Sclerosis within each MRI sequence. A similar training stage is again implemented utilising a sample size calculation step beginning at the optimal size identified for individual sequence classification and increasing this until classification accuracy improvement is insignificant in relation to the time taken for training.

Multimodal classification equally samples from sequences across each patient and each sequence to train SVM RF and ELM classifiers for both mRMR and LASSO features. A leave-out set is then used in the validation of all six models and subsequent generation of binary images containing representative Lesion segmentations.

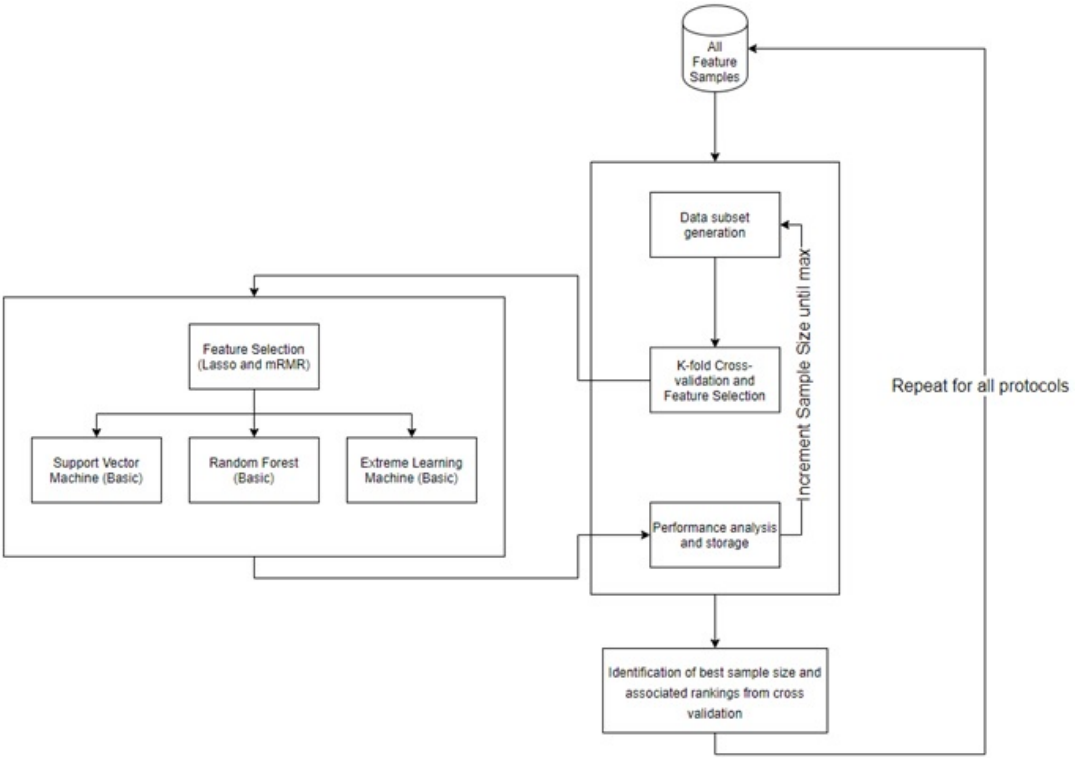


Figure 6: Individual sequence K-fold and feature selection diagram

4.4.5 Statistical Comparisons

Two secondary experiments are proposed for comparative analysis between the performance of patients and sequences with respect to each feature selection technique. Similar studies are conducted to the Individual, and Multimodal analysis highlighted previously, however, feature vectors are split into comprising feature extraction techniques and are applied to each patient dataset in order to extract features from a single method only. The performance of each feature extraction technique is analysed for both Individual and Multimodal classification with respect to patient and sequence performance in order to assess the performance and validity of each method included in this research.

This second analysis examines both Multimodal and Individual sequences. However, during this second experiment, feature selection algorithms are left out of model training. Instead, the performance is evaluated utilising entire feature vectors representing each method.

Diagrams of both systems are shown in the following two pages with fig. 7 displaying Individual and Multimodal classification of all features selected using LASSO and mRMR algorithms. The second diagram, fig. 8, depicts Individual and Multimodal classification of individual groups of features.

This section has thoroughly detailed all methods implemented within this research as well as a thorough description of the proposed Multiple Sclerosis Lesion Segmentation framework. The following section presents the results of this research for initial experiments of Multiple Sclerosis Lesion Segmentation utilising Feature Selection of the 302 length feature vector in Individual and

Multimodal MRI analysis. Following these results is the comparative analysis of each feature selection method with regards to patient datasets, MRI sequences and separate feature selection methods.

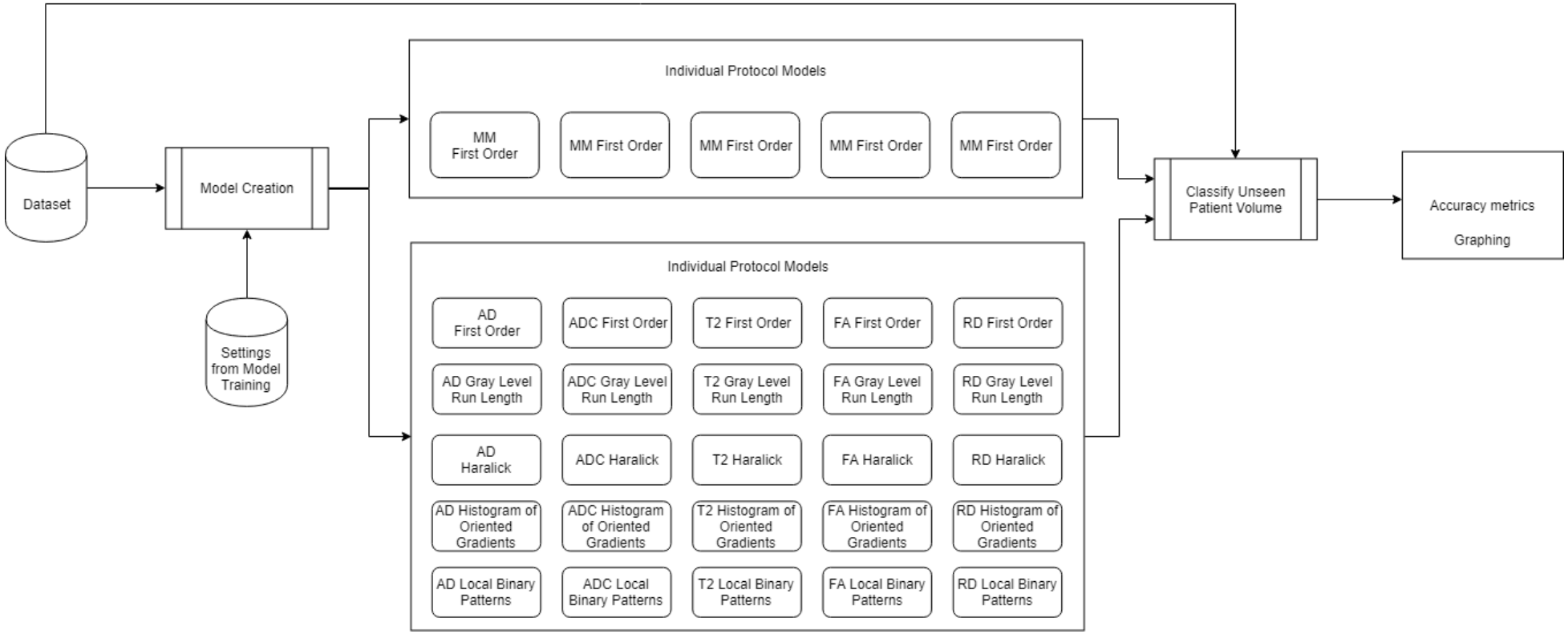


Figure 7: Overview of System for experiment 1 for Individual and Multimodal classification using LASSO and mRMR feature selection methods.

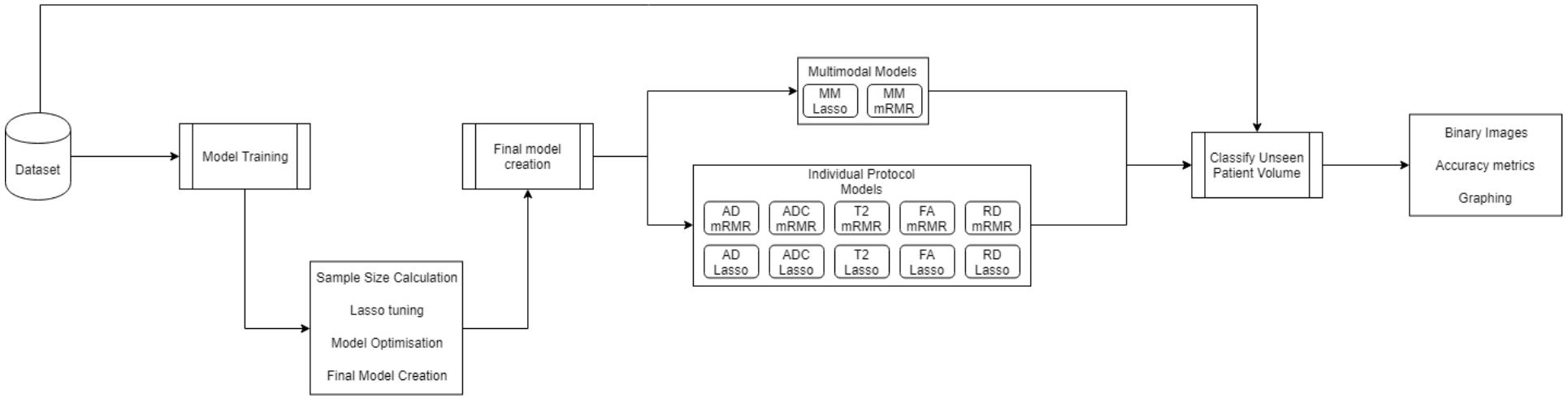


Figure 8: Overview of System for experiment 1 for Individual and Multimodal classification using groups of features from different feature extraction techniques.

5 Results

This section presents the results of this research with an overview of the training procedure provided in section 5.1. The findings of experiment one using feature selection algorithms on Individual MRI data is depicted in section 5.2, while Multimodal is shown in section 5.3. Finally, the results of experiment two in which comparison was made between feature extraction techniques are found in section 5.4 for both Individual and Multimodal Machine Learning models.

As shown in table 2, the number of Lesion samples taken from each patient varies greatly depending upon the severity of Multiple Sclerosis.

Table 2: Number of Healthy and Unhealthy tissue samples per MRI sequence.

Patient	Lesion Samples	Brain Samples	Lesion Load
p01c1	134,829	539,626	19.99%
p02c1	2840	515,014	0.54%
p04c1	2705	638,004	0.42%
p06c1	17,170	532,649	3.12%
p07c1	5125	569,696	0.89%
p08c1	11,670	590,412	1.94%
p09c1	6740	480,218	1.38%

5.1 Training

5.1.1 Model Training - Individual

The first stage of the results section provides an overview of the training procedure alongside justification for parameter selection in the training of each machine learning model.

5.1.1.1 Sample Size Calculation

Sample Size calculation of Machine Learning Models trained on Individual MRI sequences using LASSO and mRMR feature selection is shown in fig. 9 with samples ranging from 500 to 7000. The solid blue line represents the average between individual MRI sequences. Extreme Learning Machine trained using mRMR feature selection is the only model not exhibiting increased classification performance as the sample size increases.

5.1.1.2 LASSO λ tuning

Results of LASSO λ tuning are shown in fig. 10. Support Vector Machine and Random Forest models perform well, achieving an F1 score between 65 and 85. Both of these models exhibit a sharp increase in F1 score as the λ parameter surpasses 60. Extreme Learning Machine plateaus from the beginning and reduces accuracy from 60 to 90.

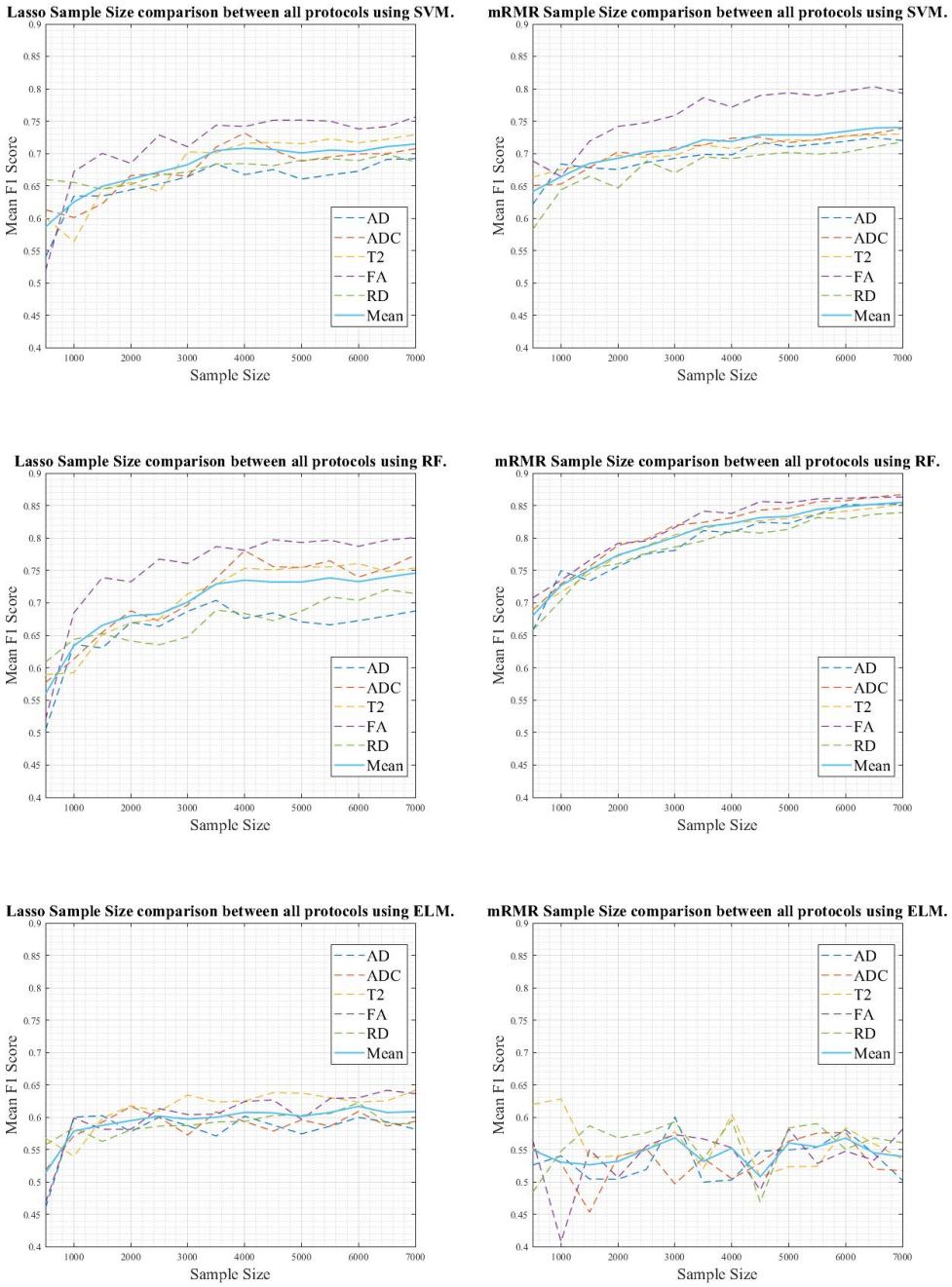


Figure 9: Sample Size calculation of each MRI sequence using mRMR and LASSO Feature Selection and SVM, RF and ELM classifiers.

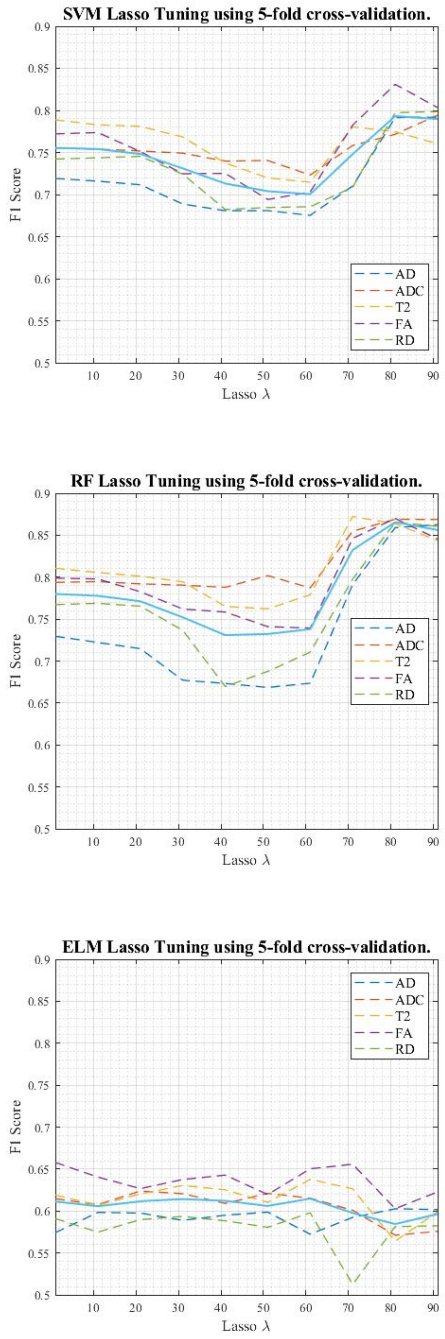


Figure 10: LASSO λ parameter tuning comparison between all three classification models using a sample size of 4000. Showing average F1 score for each MRI sequence.

5.1.1.3 Support Vector Machine Tuning

Support Vector Machine Tuning is shown in fig. 11 for LASSO and fig. 12 for mRMR feature selection techniques. Both LASSO and mRMR rankings exhibit similar trends in classification accuracy, with an increase in γ influencing an increased F1 score until a γ of 2, most notable in mRMR feature selection. After this point, the classification accuracy reduces as γ increases past 2. This is shown across both feature selection methods and all MRI sequences.

LASSO feature selection Box Constraint values generally converge around γ values between 2 and 8, whereas mRMR Box Constraint fails to unanimously converge, with lower values of 0.0313, 0.125 and 0.5 achieving less than desirable scores compared to the remaining Box Constraint parameters shown in fig. 12.

5.1.1.4 Random Forest Tuning

Random Forest number of trees tuning is shown in fig. 13 in which mRMR rankings outperform LASSO rankings by over 5%. In both cases, Fractional Anisotropy achieves the highest F1 score while T2 achieves the lowest scores for LASSO rankings and RD for mRMR rankings.

Both methods follow a similar trend in the correlation between the number of trees and increased classification performance, with a low number of decision trees, around ten, exhibiting a slightly lower classification accuracy compared to 50. However, following this threshold, there is no discernible increase in performance, and the accuracies obtained using 10,000 trees is comparable to using only 50 trees.

Further Random Forest tuning has been carried out for the Minimum Leaf Size and Number of Predictors to Sample parameters for both LASSO fig. 14 and mRMR fig. 15 feature selection methods. Both of these figures show the same trend of increased performance between 2 and 10 Predictors to Sample, with no significant increases in accuracy after this point. Both LASSO and mRMR classification results exhibit the highest accuracy with a minimum leaf size of 1, and a linear decrease is shown as the Minimum Leaf Size parameter increases.

5.1.1.5 Extreme Learning Machine Tuning

ELM tuning of the Number of Hidden Neurons for both LASSO and mRMR rankings is shown in fig. 16. Following a sharp increase in classification accuracy from 1 to 5 hidden neurons, the F1 scores level out around 60% for both feature selection experiments.

ELM Tuning of Block Size and N0 parameter using LASSO feature selection and all MRI sequences is shown in fig. 17, and mRMR feature selection for all MRI sequences in fig. 18. LASSO feature selection achieves a higher F1 score on average, and dispersion in scores between varying block sizes across all MRI sequences is much less than mRMR results which exhibit a higher degree of fluctuation and dispersion between block sizes and N0 values.

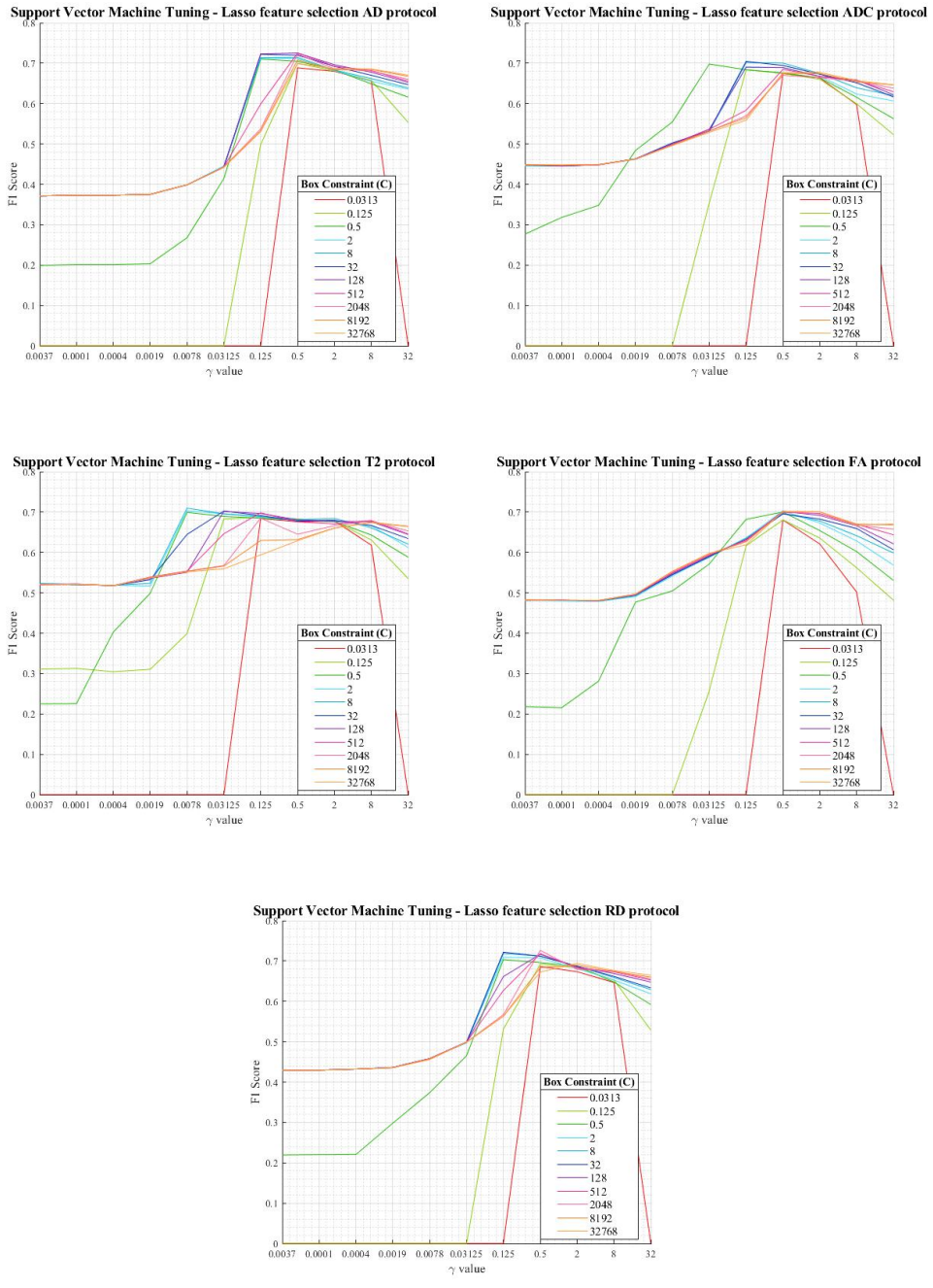


Figure 11: SVM tuning C and γ using grid search and LASSO feature selection.

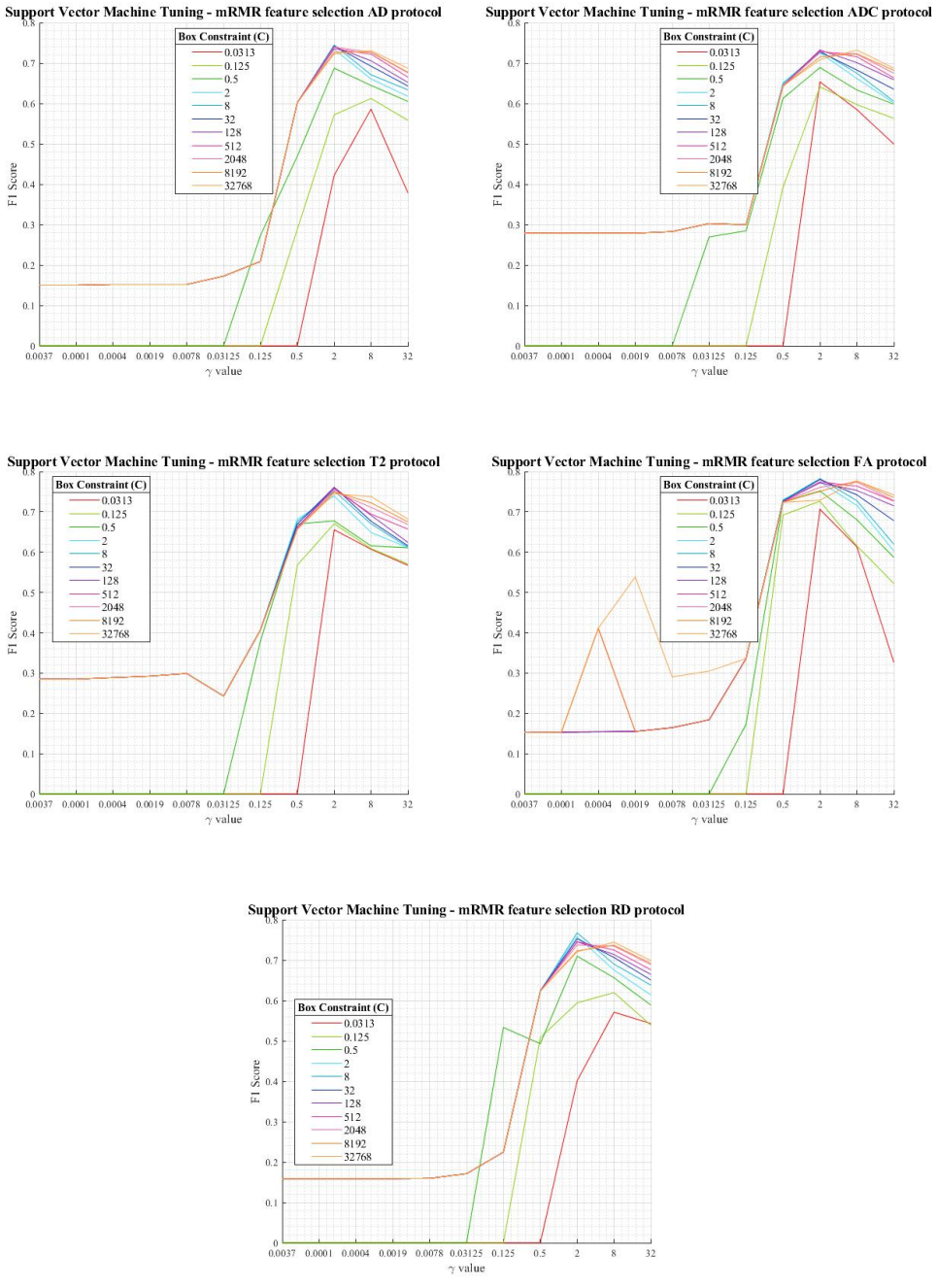


Figure 12: SVM tuning C and γ using grid search and mRMR feature selection.

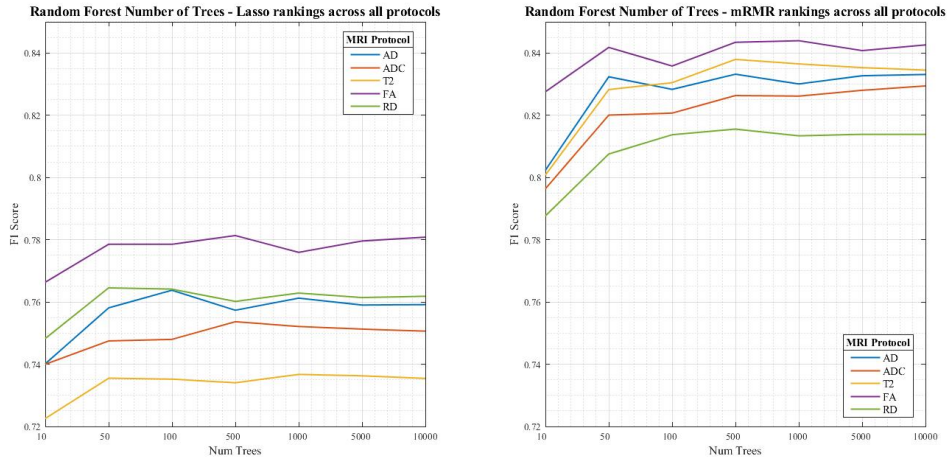


Figure 13: RF tree size tuning for LASSO and mRMR features using grid search.

5.1.2 Model Training - Multimodal

5.1.2.1 Sample Size Calculation

Multimodal Sample Size calculation is carried out for SVM, RF and ELM in the range of 1000 to 30000 fig. 19. The trend in SVM and RF classifiers in both feature selection algorithms, but most prominent in mRMR experiments, is a gradual increase in classification accuracy with an indication of slight levelling towards higher sample sizes. Extreme Learning Machine accuracy fluctuates greatly, with the highest accuracy achieved using 1000 samples.

5.1.2.2 LASSO λ tuning

Results of Multimodal LASSO tuning are shown in fig. 20. Random Forest classification achieves the highest overall set of results with SVM in second place. Extreme Learning Machine scores fluctuate considerably for lambda values between 40 and 90 while lower values between 1 and 30 yield more stable results.

5.1.2.3 Support Vector Machine Tuning

Support Vector Machine Tuning is shown in fig. 21 for LASSO and mRMR feature selection methods. Both methods show an increased F1 score as the γ parameter increments. Following a value of two, a significant reduction in accuracy is shown in both graphs. This is also true of Individual training.

5.1.2.4 Random Forest Tuning

Results for Random Forest number of Trees calculation is shown in fig. 22. Both LASSO and mRMR rankings exhibit the same trend of increased accuracy between 1 and 50 trees despite LASSO scoring consistently higher than mRMR.

Random Forest results of Minimum Leaf Size and Number of Predictors to Sample tuning are shown in fig. 23. Both of these figures show the same trend of increased performance between 2

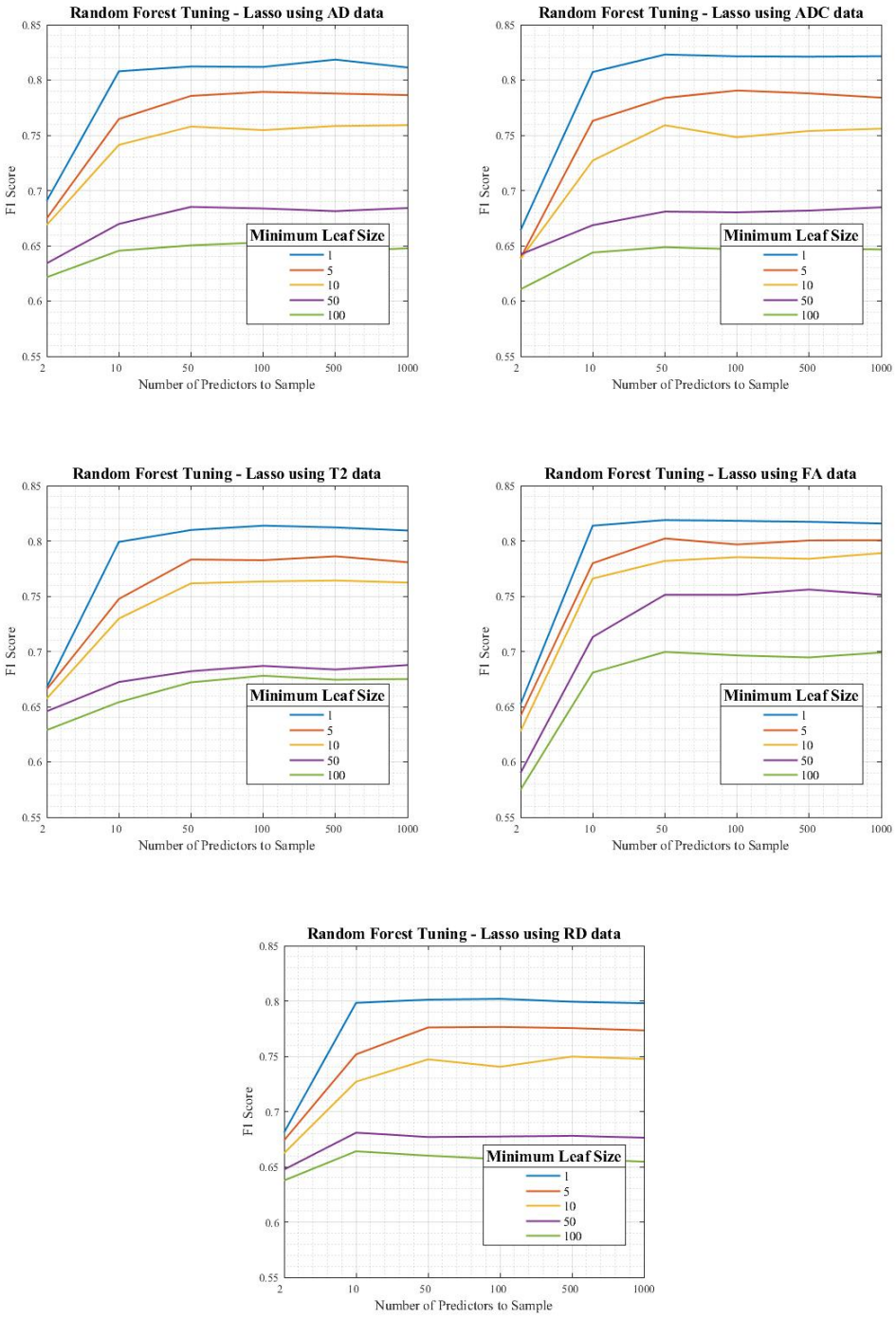


Figure 14: RF Leaf Size against Number of predictors tuning using grid search and LASSO feature selection.

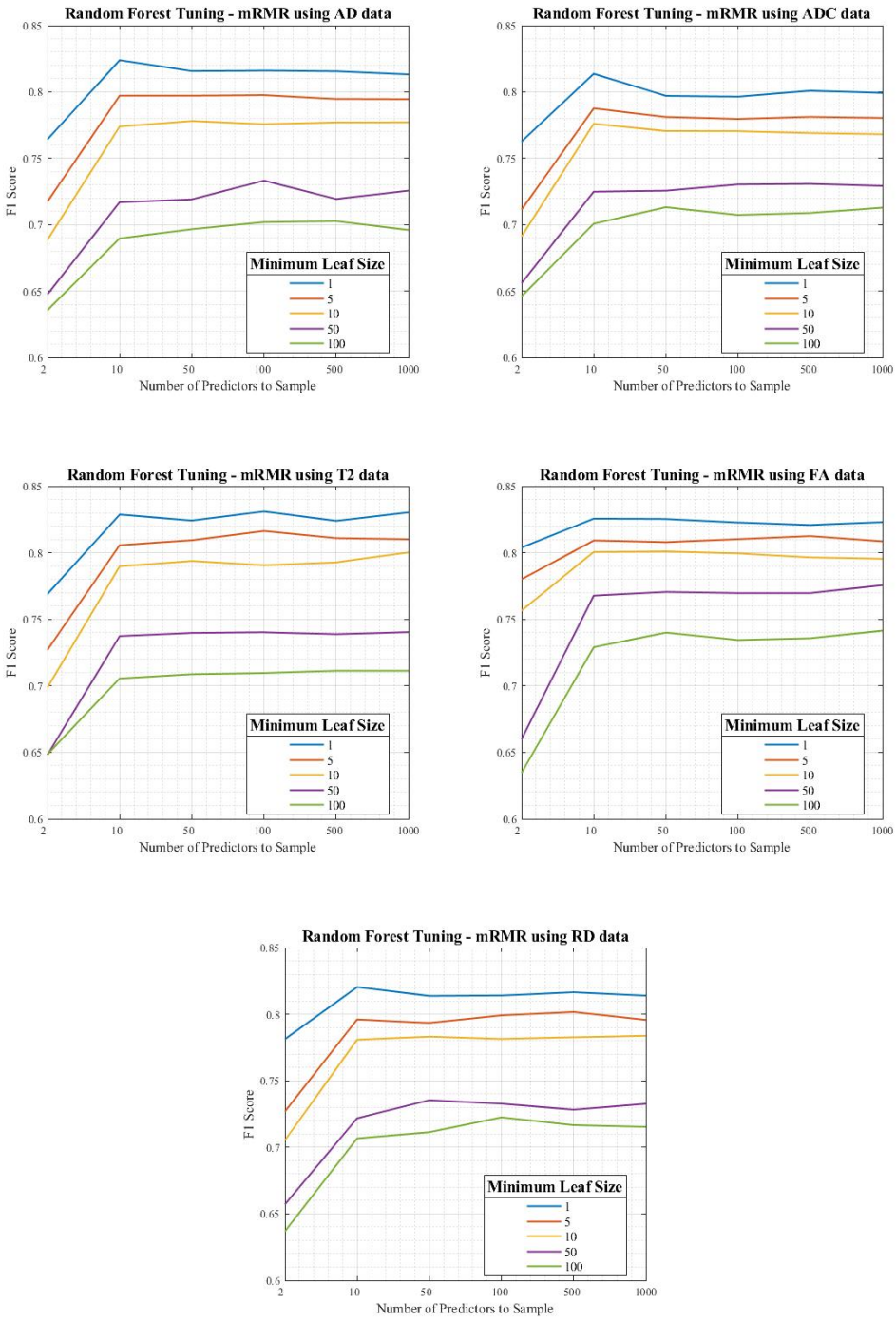


Figure 15: RF Leaf Size against Number of predictors tuning using grid search and mRMR feature selection.

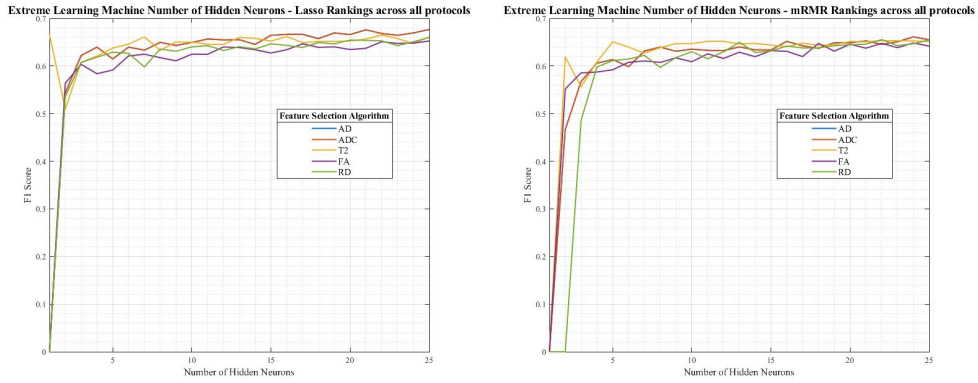


Figure 16: ELM Number of Hidden Neurons tuning for LASSO and mRMR features using grid search.

and 10 Predictors to Sample, with a plateau in accuracy and no significant increases after this point. Both LASSO and mRMR classification results exhibit the highest accuracy with a minimum leaf size of one, and a linear decrease is shown as the Minimum Leaf Size parameter increases, indicated by each line in all plots. The same trend is seen in classification using Individual MRI sequence information.

5.1.2.5 Extreme Learning Machine Tuning

Extreme Learning Machine tuning of the Number of Hidden Neurons for both Multimodal LASSO and mRMR rankings is shown in fig. 24. Both LASSO and mRMR results show a similar trend of a sharp increase in classification accuracy from 1 to 5 hidden neurons and a levelling out of F1 score at around 60%.

Extreme Learning Machine Tuning of Block Size and N0 parameter using LASSO and mRMR feature selection methods on Multimodal data is shown in fig. 25. Both sets of results show similar signs of convergence and sustained accuracy between different parameter values for Block Size and N0 parameters. Noticeably, mRMR rankings begin with a low F1 score for Block Sizes of 10, 50 and 100 when coupled with an N0 parameter of 50 or less. LASSO feature selection does not exhibit this trend and instead hold a continuous trend in accuracy for all values of Block Size and N0.

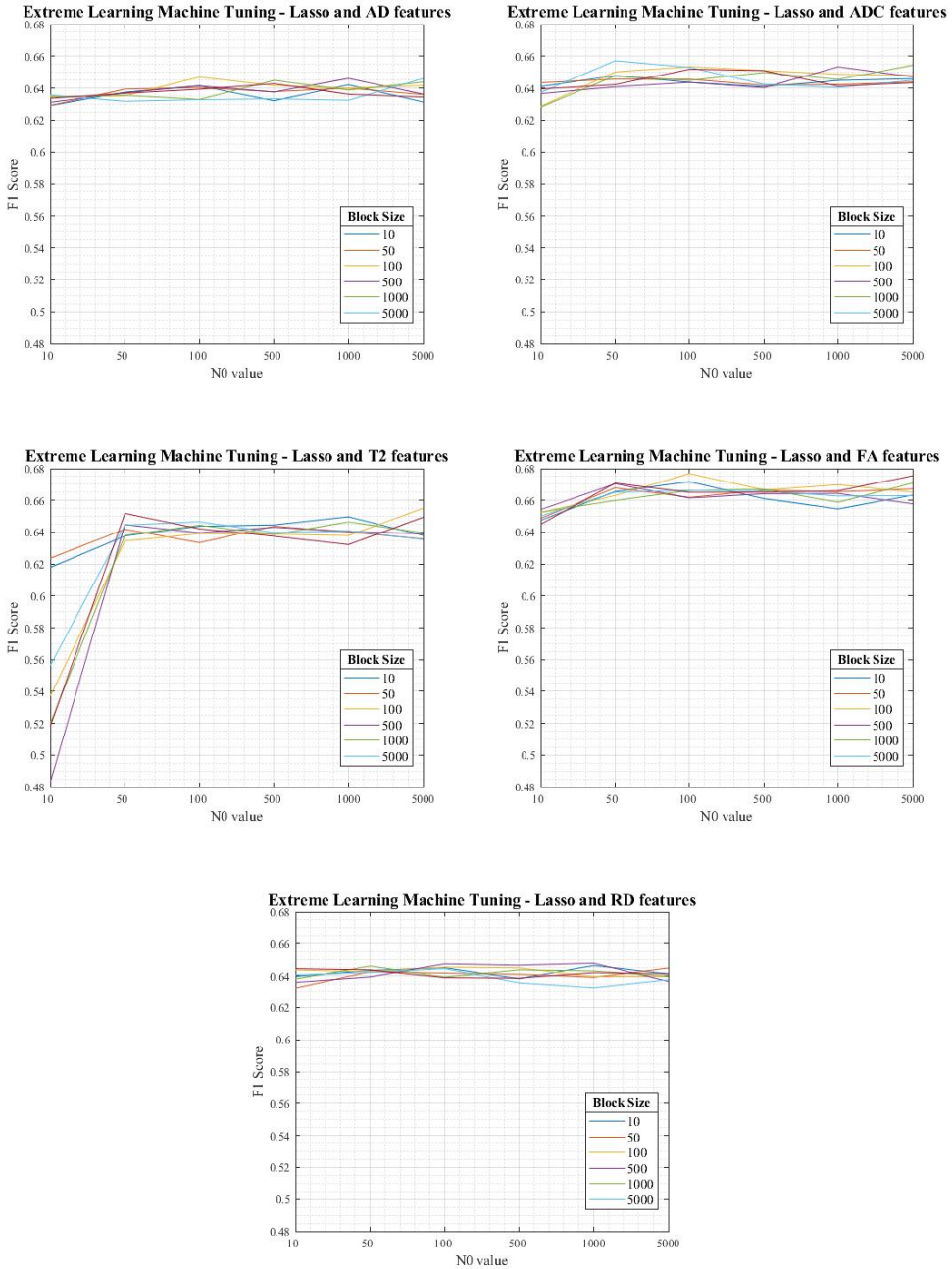


Figure 17: ELM Block Size against N0 parameter tuning using grid search and LASSO feature selection.

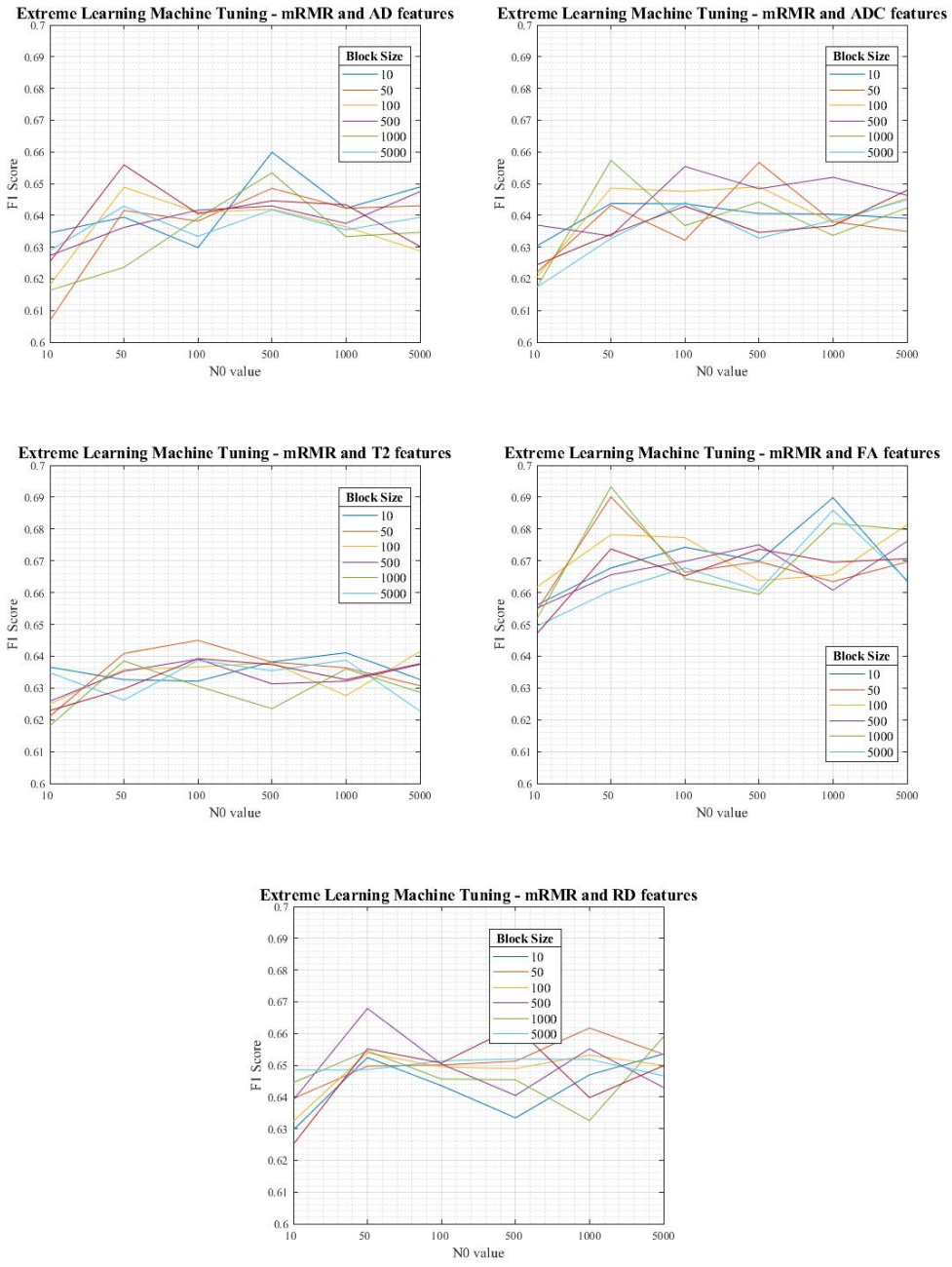


Figure 18: ELM Block Size against N0 parameter tuning using grid search and mRMR feature selection.

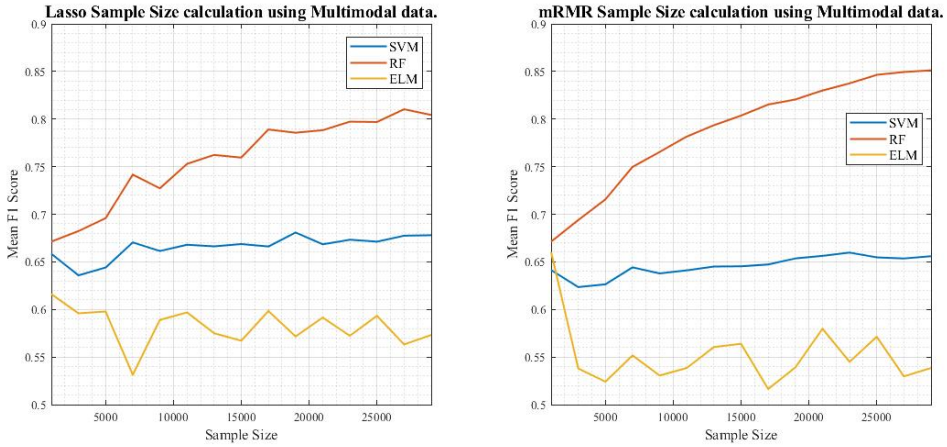


Figure 19: Sample Size calculation of Multimodal MRI data using mRMR Feature Selection and SVM, RF and ELM classifiers.

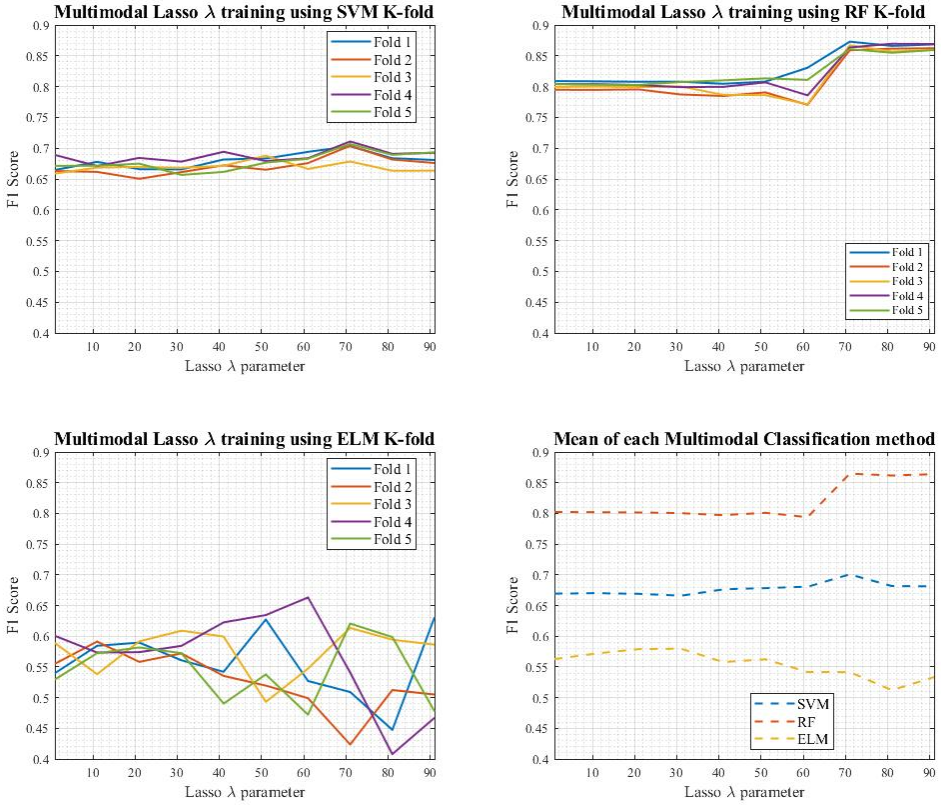


Figure 20: LASSO λ parameter tuning for all classification models using a sample size of 4000. Showing average F1 score for Multimodal MRI data.

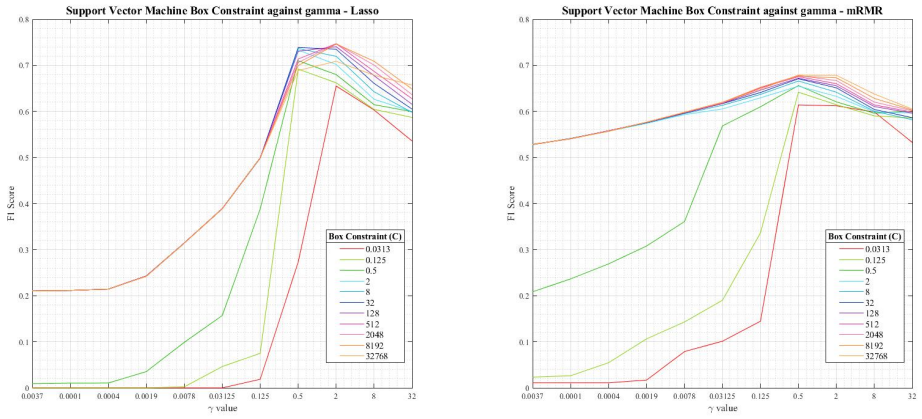


Figure 21: SVM tuning C and γ using grid search for LASSO and mRMR feature selection for Multimodal classification.

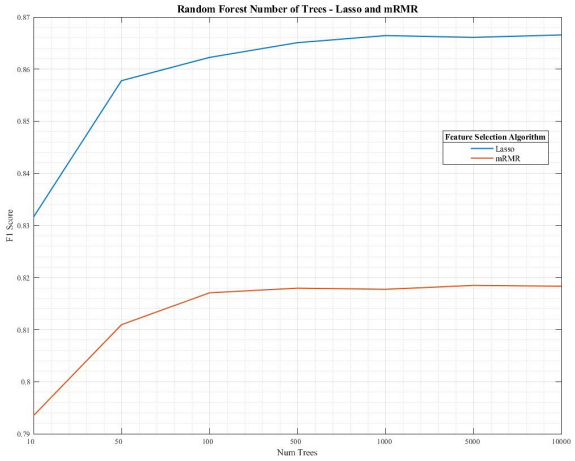


Figure 22: RF tree size tuning for Multimodal classification using LASSO and mRMR features and grid search.

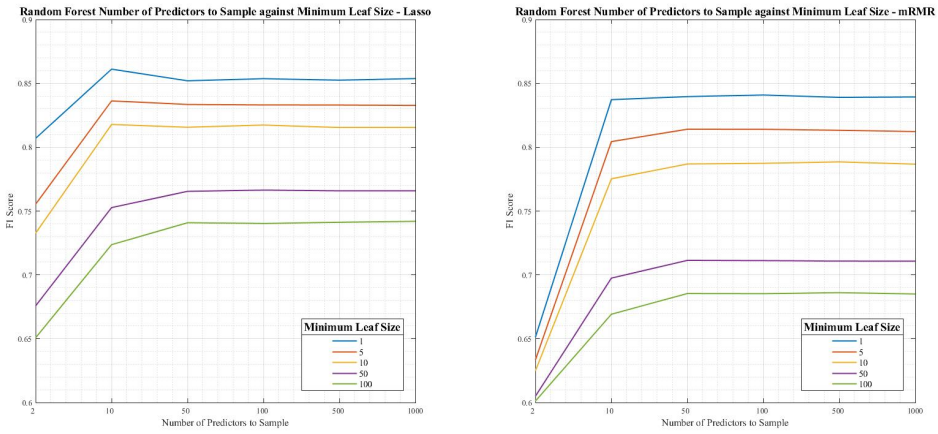


Figure 23: RF Leaf Size against Number of predictors tuning for Multimodal classification using LASSO and mRMR features and grid search.

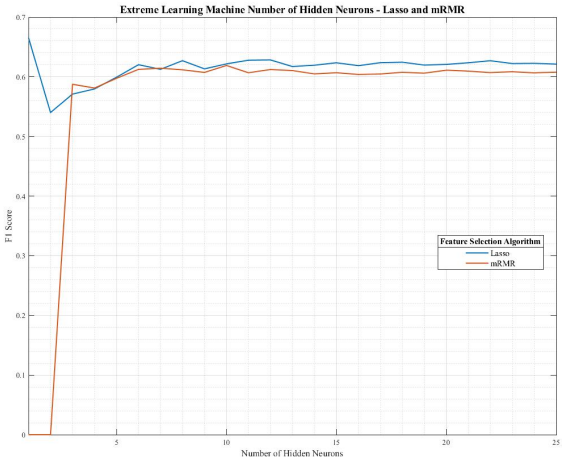


Figure 24: ELM Number of Hidden Neurons tuning for LASSO and mRMR features using grid search for Multimodal classification.

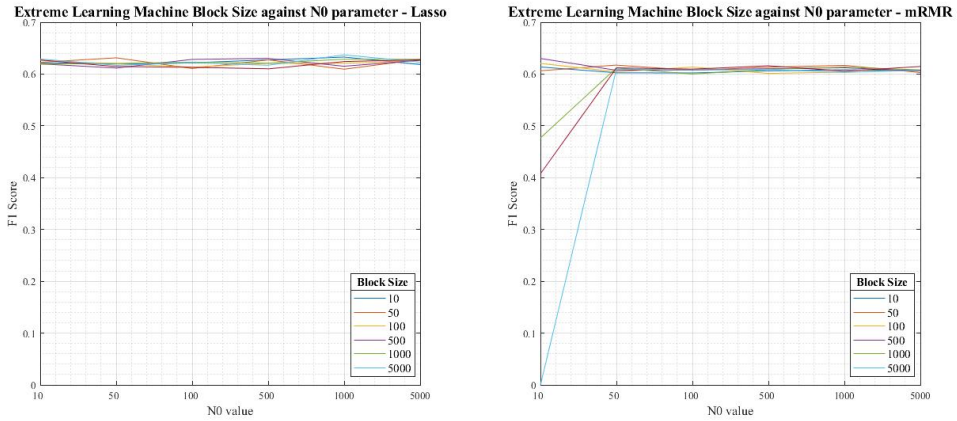


Figure 25: ELM Block Size against N0 parameter tuning using grid search and LASSO feature selection for Multimodal classification.

5.1.3 Tuned parameters

Following the cross-validation and optimisation procedure outlined in section 4, the final parameters shown in table 3 were selected based upon the tuning results shown during this section. This selection was made following careful consideration of performance increase and computational expense, as well as the success of all MRI protocols individually so as to avoid biasing parameters selection.

Table 3: Values of trained parameters following 5-fold cross-validation and optimisation procedure.

Parameter	**Individual**	**Multimodal**
Sample Size	4000	20,000
Lasso λ	70	70
C (SVM)	2	8
γ (SVM)	8	0.5
Trees (RF)	50	500
MLS (RF)	1	1
NPS (RF)	10	10
NhN (ELM)	5	10
Block (ELM)	10	10
N0 (ELM)	5	50

5.1.4 Final Model Performance and Rankings

Average leave out cross-validation accuracies for Individual and Multimodal model training are shown in fig. 26. Throughout all folds in both Individual sequence and Multimodal classification experiments, Extreme Learning Machine achieves the lowest overall accuracy averaging an F1 score of 60%. The Random Forest classifier yields the highest accuracy of 84.52% F1 score across all sequences and Multimodal classification. Support Vector Machine shows a performance between that of Extreme Learning Machine and Random Forest, around 75% F1 score. The final F1 scores, Sensitivity and Specificity of each classifier and feature extraction technique is shown in table 4.

Final Rankings from Individual and Multimodal experiments are shown in fig. 27. HoG and Haralick features are ignored in all experiments, and of all feature extraction techniques, LBP features has the most occurences.

Table 4: Average cross validation accuracy of all 6 classifiers.

Model	**F1 score**	**Sensitivity**	**Specificity**
SVM LASSO	74.80 ± 3.65	70.18 ± 3.81	82.57 ± 3.71 %
RF LASSO	83.24 ± 3.60	78.03 ± 2.53	90.68 ± 3.13%
ELM LASSO	59.76 ± 5.35	56.22 ± 8.75	72.54 ± 6.09%
SVM mRMR	73.70 ± 3.00	67.96 ± 4.76	83.70 ± 1.64%
RF mRMR	84.52 ± 0.68	80.81 ± 1.18	89.60 ± 2.13%
ELM mRMR	55.14 ± 2.11	47.94 ± 4.25	76.93 ± 5.72%

Figure 26: Average accuracy across all 5 cross-validation folds during final model tuning of Models using Individual MRI sequence and Multimodal data.

Individual and Multimodal Rankings using Lasso and mRMR feature selection.

Feature extraction method			AD Lasso	AD mRMR	ADC Lasso	ADC mRMR	T2 Lasso	T2 mRMR	FA Lasso	FA mRMR	RD Lasso	RD mRMR	Multimodal Lasso	Multimodal mRMR
		Local Binary Pattern	**15**	**14**	**10**	**10**	**17**	**10**	**14**	**9**	**15**	**15**	**14**	**20**
		HoG	0	0	0	0	0	0	0	0	0	0	0	0
	Haralick Texture Features	IMoCI	0	0	0	0	0	0	0	0	0	0	0	0
		Difference Entropy	0	0	0	0	0	0	0	0	0	0	0	0
		Difference Variance	0	0	0	0	0	0	0	0	0	0	0	0
		Entropy	0	0	0	0	0	0	0	0	0	0	0	0
		Sum Entropy	0	0	0	0	0	0	0	0	0	0	0	0
		Sum Variance	0	0	0	0	0	0	0	0	0	0	0	0
		Sum Average	0	0	0	0	0	0	0	0	0	0	0	0
		Homogeneity	0	0	0	0	0	0	0	0	0	0	0	0
		Variance	0	0	0	0	0	0	0	0	0	0	0	0
		Correlation	0	0	0	0	0	0	0	0	0	0	0	0
		Contrast	0	0	0	0	0	0	0	0	0	0	0	0
		ASM	0	0	0	0	0	0	0	0	0	0	0	0
	Grey Level Run Length	LRHGE	**1**	0	**1**	0	**1**	**1**	0	0	**1**	0	**1**	**1**
		LRLGE	0	**1**	0	0	0	0	0	0	0	0	**1**	0
		SRHGE	**1**	**1**	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**
		SRLGE	0	0	0	0	0	0	**1**	0	0	0	**1**	**1**
		HGRE	0	0	**1**	**1**	0	**1**	0	0	0	0	**1**	**1**
		LGRE	0	0	0	0	0	**1**	0	0	0	0	**1**	0
		RP	0	0	0	0	0	0	0	0	0	0	0	0
		RLN	0	0	0	0	0	0	0	**1**	0	0	**1**	0
		GLN	0	0	0	0	0	0	0	**1**	0	0	**1**	0
		LRE	0	0	0	0	0	0	0	0	0	0	0	0
		SRE	0	0	**1**	**1**	**1**	**1**	**1**	0	0	0	**1**	**1**
	First Order	Standard Deviation	0	0	0	0	0	0	0	0	0	0	0	0
		Median	0	0	0	0	0	0	0	0	0	0	0	0
		Mean	0	0	0	0	0	0	0	0	0	0	0	0
		Max	0	0	0	0	0	0	0	**1**	0	0	**1**	0
		Min	0	**1**	0	**1**	0	0	**1**	0	0	**1**	**1**	0

Figure 27: Individual and Multimodal Rankings from LASSO and mRMR feature selection methods.

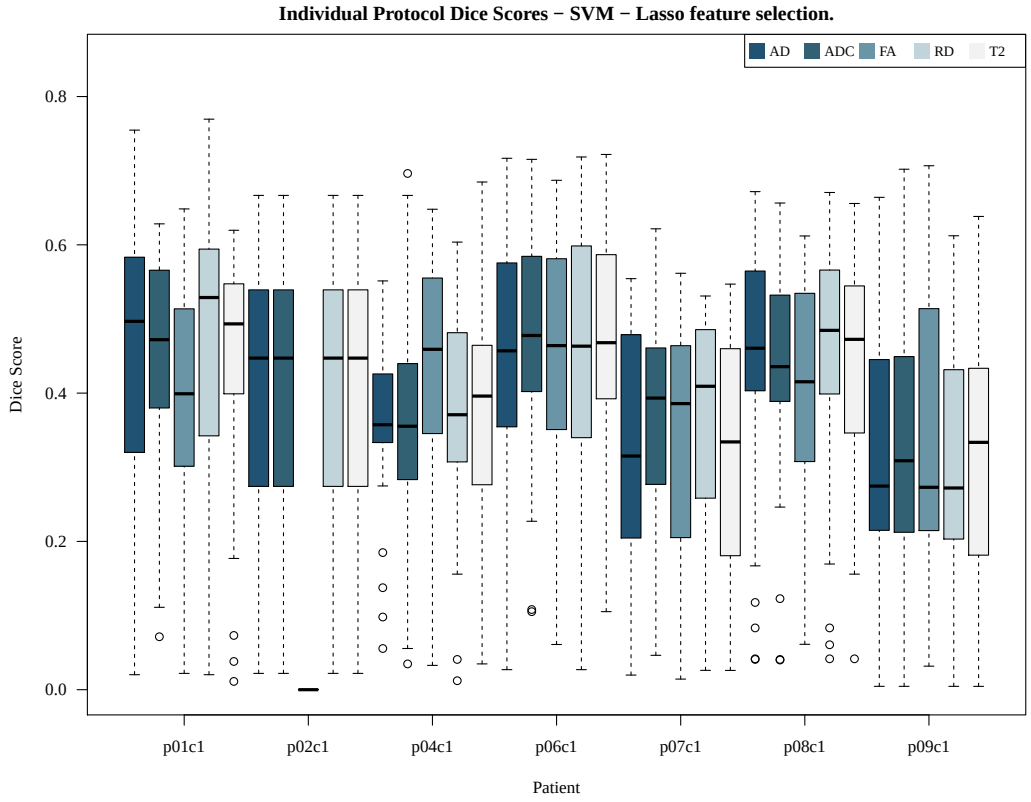


Figure 28: Dice score performance for entire MRI volumes using Support Vector Machine Models trained on Individual MRI sequences and LASSO Feature Selection.

5.2 Unseen Dataset Classification - Individual

The following section provides the results from the first proposed experiment of unseen patient dataset classification using Individual and Multimodal machine learning models and features identified through LASSO and mRMR feature selection algorithms. Lasso graphs have been provided intext, with the mRMR results shown in appendix A. Interpretation of these results is carried out in section 6.1.2.

5.2.1 Support Vector Machine

Individual SVM Dice scores are represented as boxplots in fig. 28 for LASSO and appendix A fig. 63b for mRMR feature selection methods. A significant number of median Dice values fell between 40 and 60 for slices within classified volumes and datasets exhibited similar trends in segmentation accuracy between feature selection methods, with p01c1, p06c1 and p08c1 yielding the highest scores, while sets p04c1, p07c1 and p09c1 score lowest in both methods.

It is also clear that deviations are present such as the inability to classify AD, ADC and RD sequences from patient dataset p02c1 using mRMR feature selection alongside the SVM model. LASSO feature selection exhibits a single instance of misclassification in the case of p02c1 in which the AD MRI sequence fails.

Bar graphs showing the occurrences of slice accuracies are shown in fig. 29 for LASSO and ap-

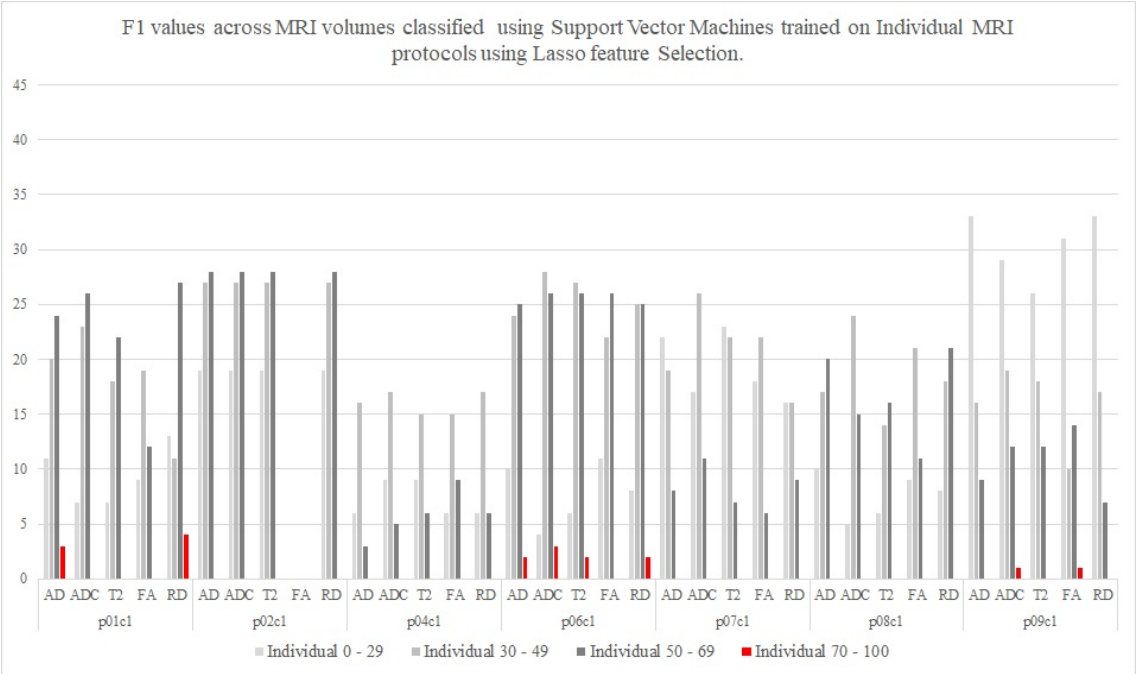


Figure 29: Bar charts displaying binned distributions of slice accuracies for Individual SVM experiments using LASSO feature selection. Red represents the number of slices within a volume scoring greater than 70.

pendix A fig. 64b for mRMR feature selection methods. Significantly more slices were classified between 30 and 69 compared to those with an F1 score greater than 70%.

5.2.2 Random Forest

Individual RF Dice scores are represented as boxplots in fig. 30 for LASSO and appendix A fig. 65b for mRMR feature selection methods. The highest Dice scores were obtained by p01c1, p06c1 and p08c1, with apparent differences between these three patients and the remaining datasets in both the LASSO and mRMR case.

Bar graphs complimenting these boxplots are shown in fig. 31 for LASSO and appendix A fig. 66b for mRMR feature selection methods. Unlike the SVM results, LASSO and mRMR feature selection appear much closer in accuracy across all F1 bins. Similar to the RF boxplot results, the datasets p01c1 and p06c1 show higher levels of slice scores between 50 and 100.

5.2.3 Extreme Learning Machine

Individual ELM Dice scores are shown in fig. 32 for LASSO feature rankings and appendix A fig. 67b for mRMR. Both experiments failed to classify dataset p02c1 FA, and in LASSO feature selection it is possible to identify datasets p01c1 p06c1 and p08c1 as having a higher overall accuracy across MRI sequences.

Bar graphs complimenting these boxplots are shown in fig. 33 for LASSO and appendix A fig. 68b for mRMR feature selection methods. While LASSO feature selection falls behind in maximum slice results between 70 and 100, it is again shown that the two feature selection methods exhibit

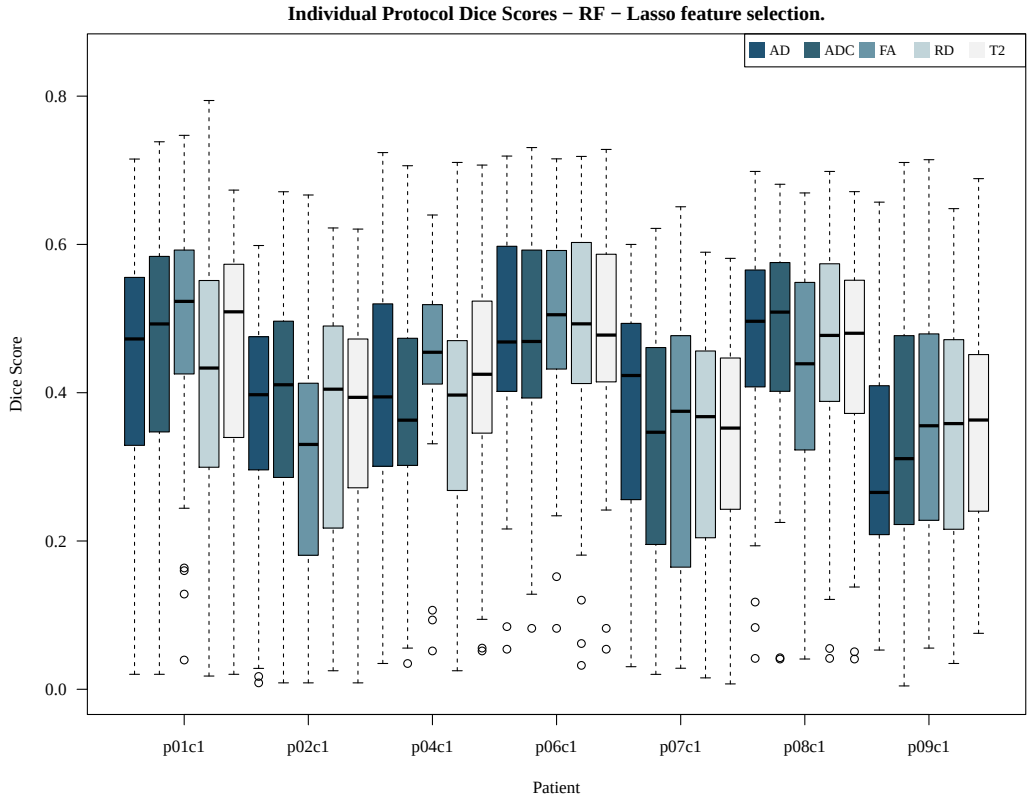


Figure 30: Dice score performance for entire MRI volumes using Random Forest Models trained on Individual MRI sequences and LASSO Feature Selection.

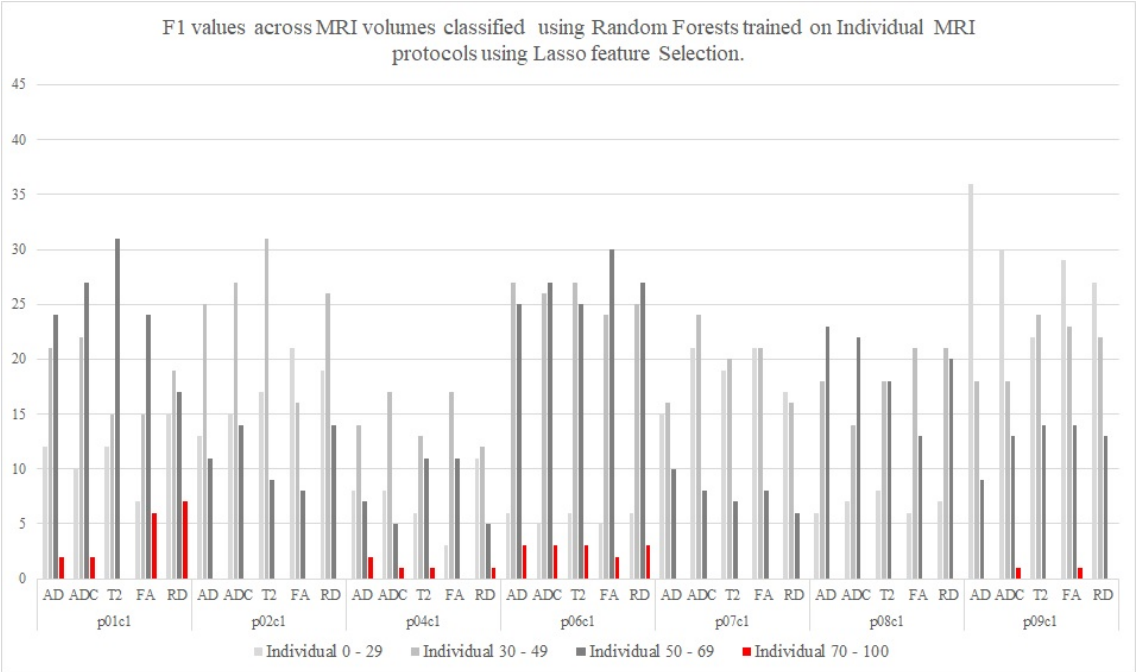


Figure 31: Bar charts displaying binned distributions of slice accuracies for Individual RF experiments using LASSO feature selection. Red represents the number of slices within a volume scoring greater than 70.

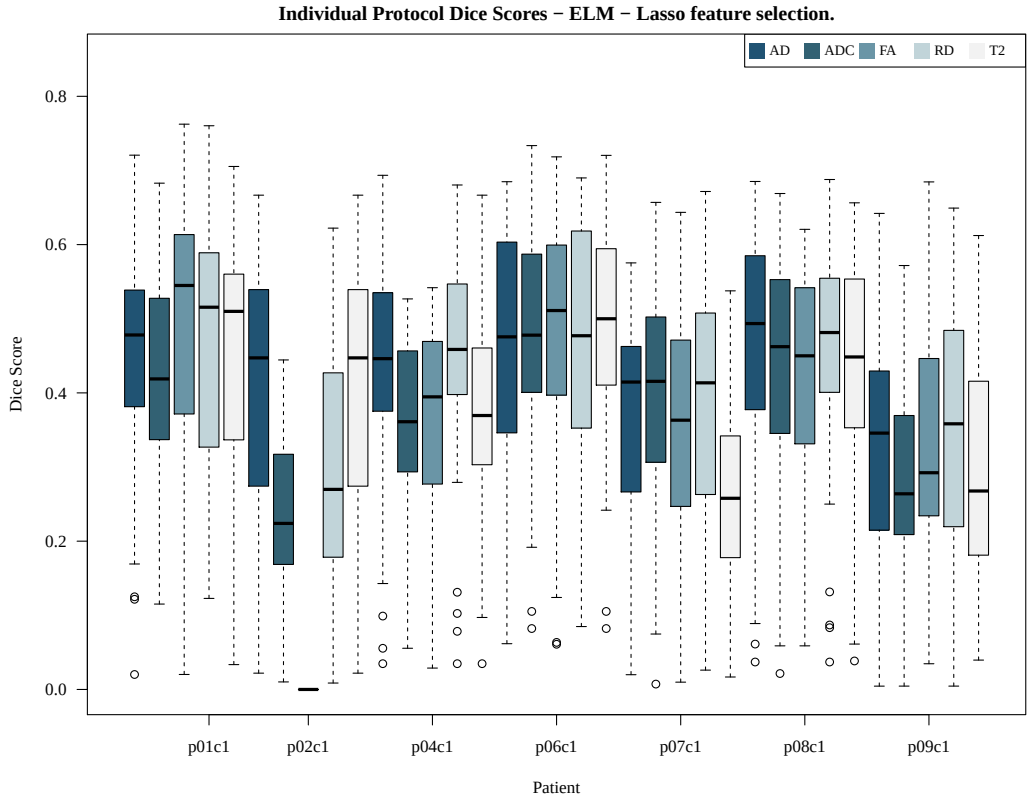


Figure 32: Dice score performance for entire MRI volumes using Extreme Learning Machine Models trained on Individual MRI sequences and LASSO feature Selection.

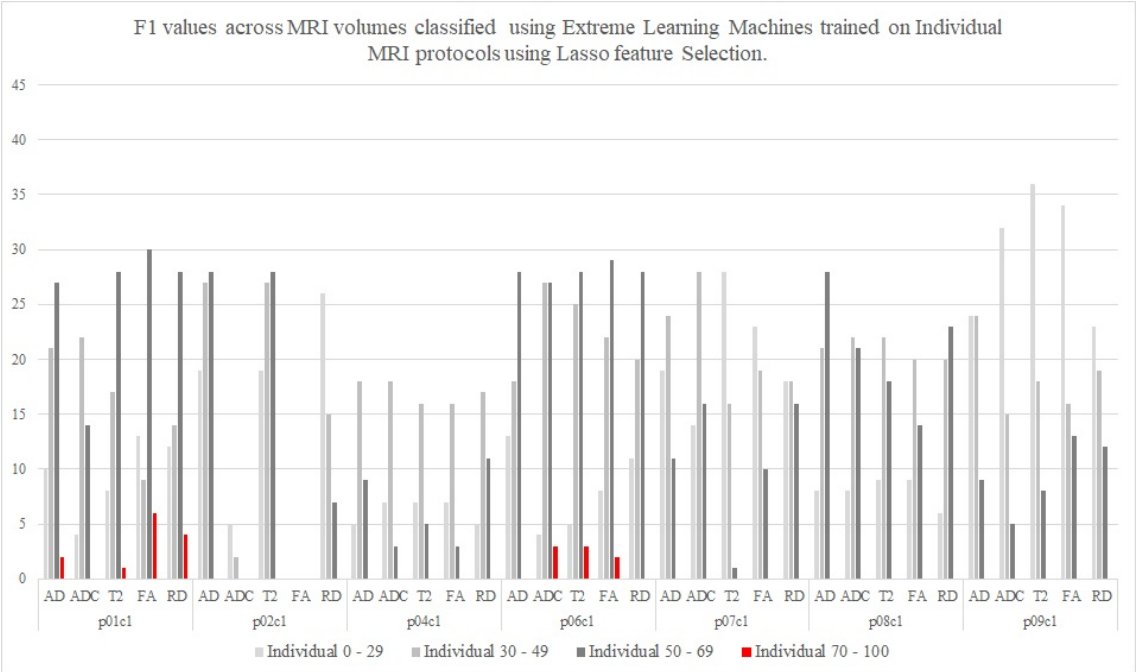


Figure 33: Bar charts displaying binned distributions of slice accuracies for Individual ELM experiments using LASSO feature selection. Red represents the number of slices within a volume scoring greater than 70.

similar trends in classification performance with minimal maximum slice classifications and the majority of slices classifying between 30 and 69.

5.2.4 Multiple Sclerosis Lesion Segmentation Images

The segmentation success of Individual sequence classification using machine learning models trained on Individual MRI sequence data can be seen in fig. 34 and fig. 35. The first figure depicts Axial Diffusion sequence classification on patient p01c1 slice 109 with a clear segmentation of the largest Multiple Sclerosis Lesion regions across all Machine Learning models.

Depicted in fig. 35 is a scan taken from the same dataset as fig. 34, however this time the sequence classified was FA. ELM exhibits the highest F1 score in both instances, which captured the most shape information, with SVM achieving the lowest score.

Individual segmentations of dataset p02c1 AD slice 108 are shown in appendix A fig. 69a and p04c1 ADC 110 in appendix A fig. 69b, both using models trained on their respective sequences. In both cases, smaller Lesions are missed leading to lower slice accuracies and sensitivity.

T2 MRI Individual sequence results are shown in appendix A fig. 70a for mRMR feature selection and appendix A fig. 70b for LASSO feature selection, both of which depict slices from patient dataset p06c1. In both instances, over half of the Lesions are correctly identified as such with a significant portion of both slices identifying MS within larger Lesion regions, while again missing smaller areas.

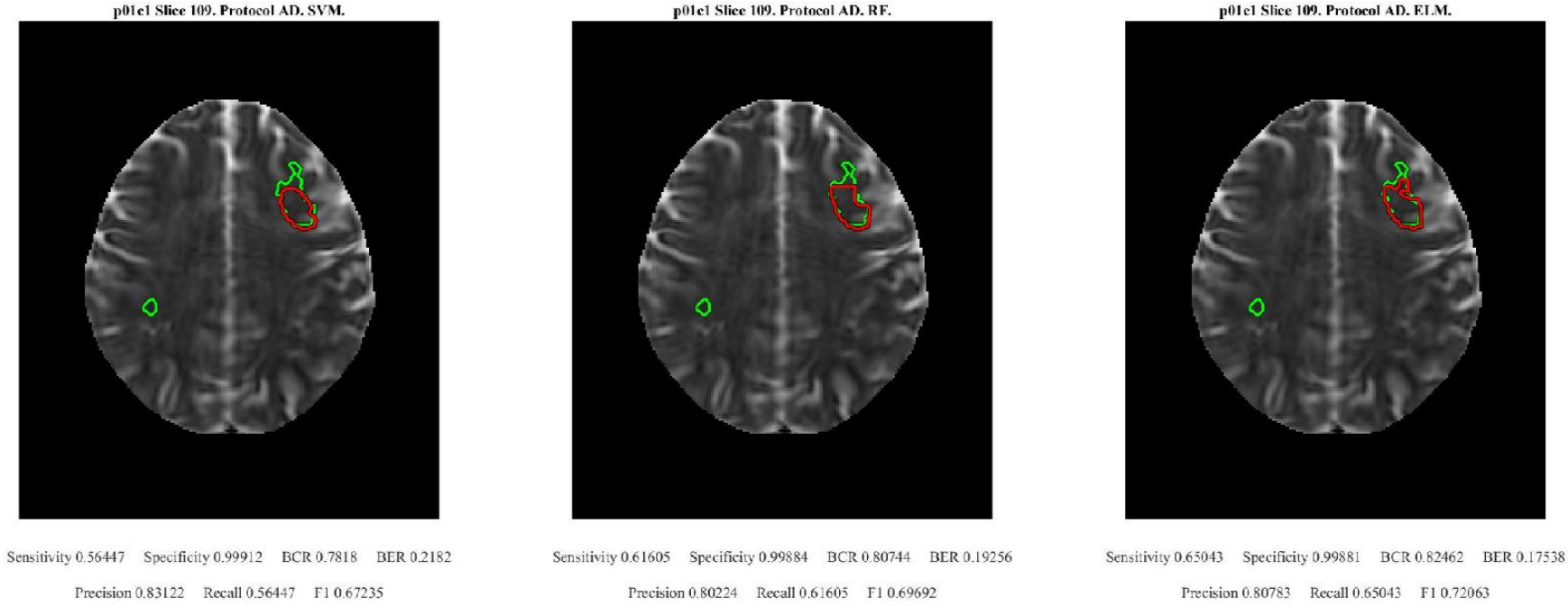


Figure 34: Segmentations using all machine learning classifiers trained on AD MRI sequence and features selected using LASSO feature selection for MS identification of dataset p01c1 AD sequence on slice 109.

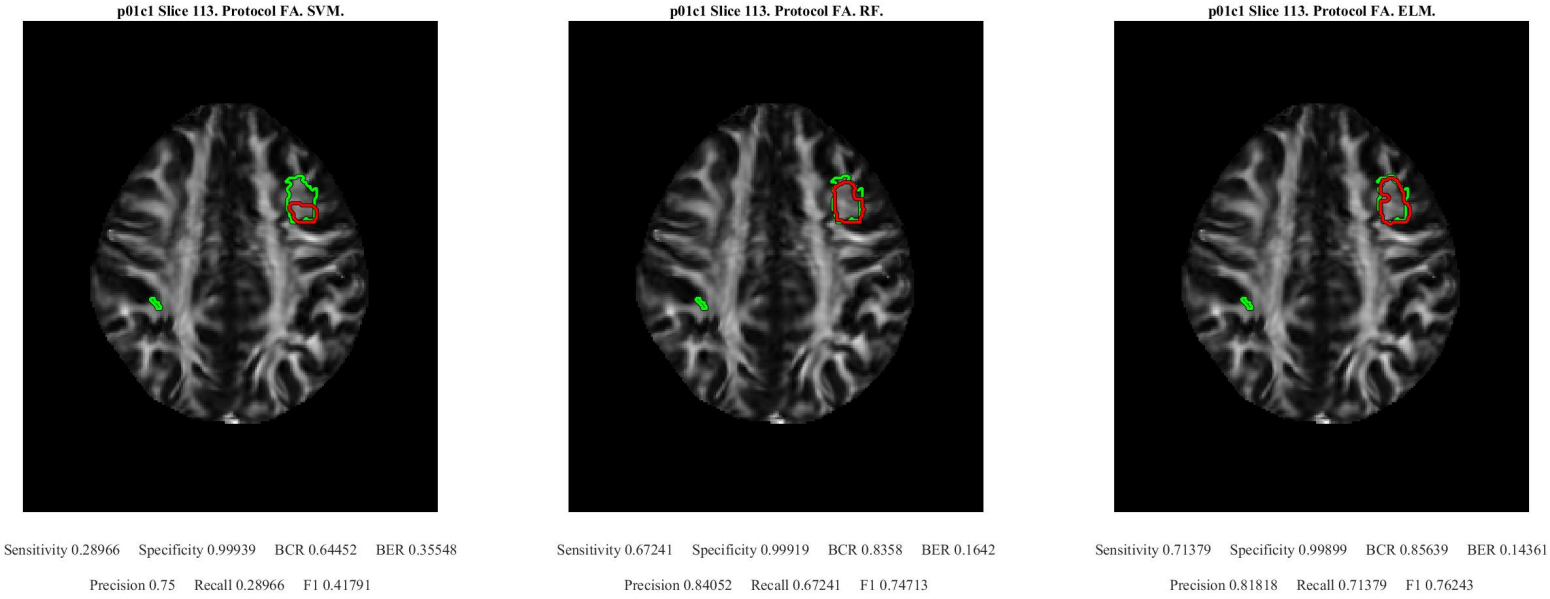


Figure 35: Models trained on FA data only used to predict p01c1 slice 113. Both using LASSO feature selection.

5.2.5 Cross-sectional Heatmap graphs.

This section presents the results for Individual Machine Learning model classification of unseen MRI sequence data, depicted as Heatmap graphs. Cross-sectional graphs of unseen dataset p06c1 are shown in fig. 36 for LASSO and fig. 37 for mRMR feature selection. Predicted slices fall inside the manual Lesion segmentation range, as shown in the first columns. Significant reduction in accuracy is exhibited towards upper and lower edges of the Lesion volume while greatest accuracy is shown in the upper middle section of predicted volumes.

Cross-sectional graphs of unseen dataset p02c1 are shown in fig. 71 for LASSO and fig. 72 for mRMR feature selection methods. Maximum classification accuracy peaks around an F1 score of 60 and a significant number of Machine Learning Models failed for both LASSO and mRMR. Furthermore, in all experiments, significant gaps between slices containing Lesion are present.

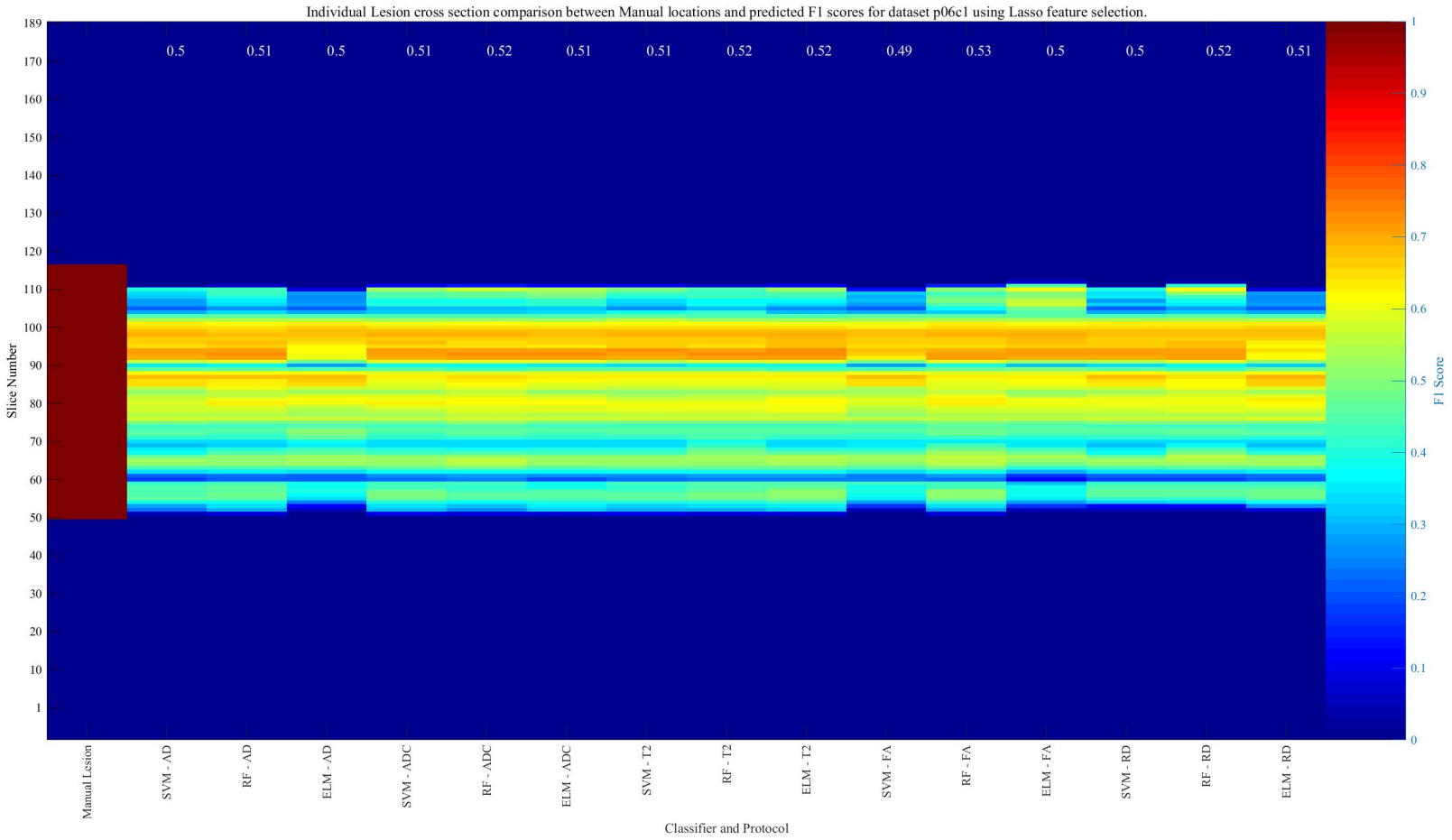


Figure 36: Cross sectional graph for p06c1 using Machine Learning models trained on Individual MRI data and LASSO feature selection.

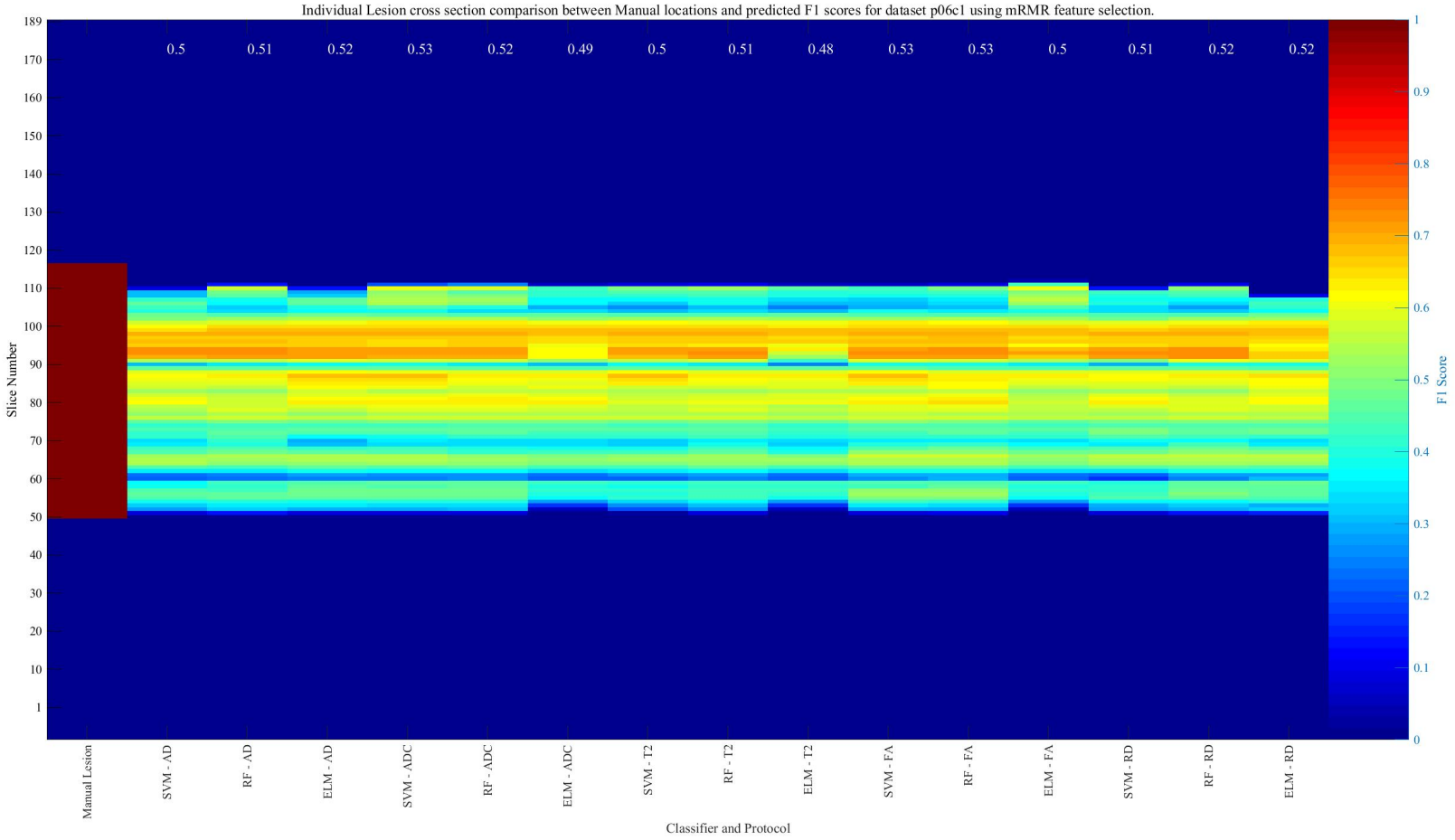


Figure 37: Cross sectional graph for p06c1 using Machine Learning models trained on Individual MRI data and mRMR feature selection.

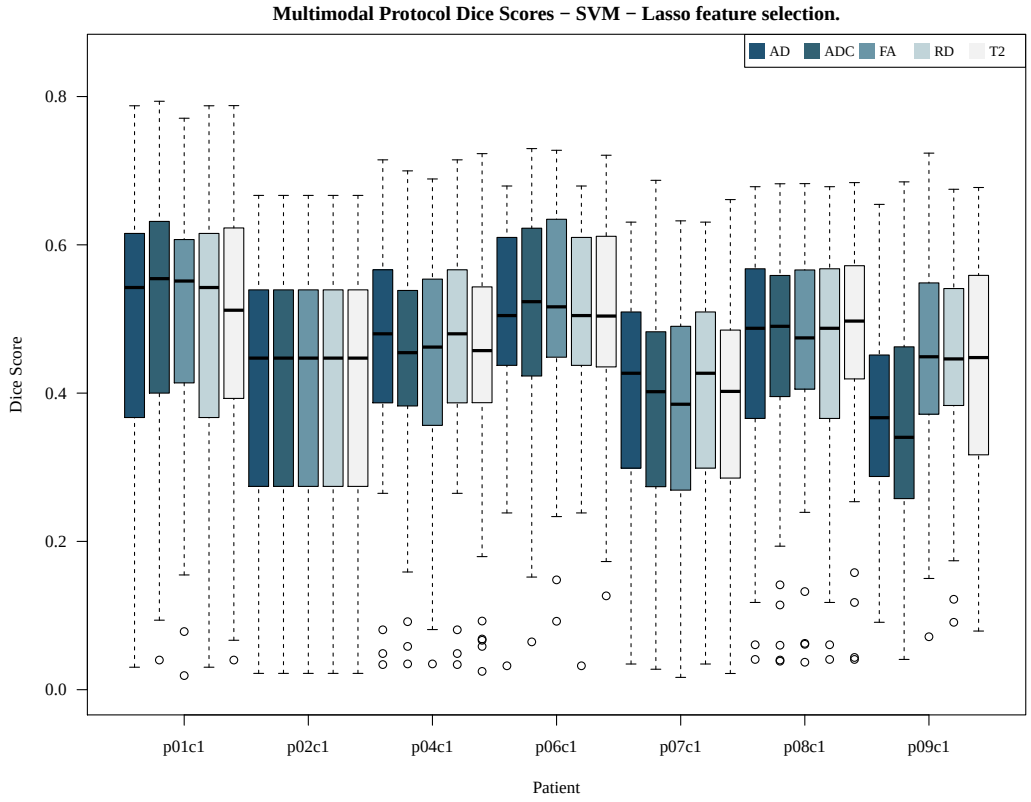


Figure 38: Dice score performance for entire MRI volumes using Support Vector Machine Models trained on Multimodal MRI data and LASSO Feature Selection.

5.3 Unseen Dataset Classification - Multimodal

5.3.1 Support Vector Machine

Multimodal Boxplots showing Dice scores for entire patient volumes are shown in fig. 38 for LASSO and appendix B fig. 73b for mRMR feature selection using Support Vector Machine models. These results group as patient datasets as opposed to MRI sequences between the datasets. A significant portion of these dataset results fall between a Dice score of 0 and 70 for both LASSO and mRMR feature selection methods.

Bar graphs complimenting these boxplots are shown in fig. 39 for LASSO and appendix B fig. 74b for mRMR feature selection methods. LASSO feature selection achieves more top slice classifications compared to mRMR.

5.3.2 Random Forest

Random Forest Dice scores for Multimodal classification volumes are shown in fig. 40 for LASSO and appendix B fig. 75b for mRMR feature selection algorithms. Boxplots group as patient datasets as opposed to MRI sequences and median slice accuracy fluctuates between 0.25 and 0.58. Datasets p01c1, p06c1 and p08c1 and classified slightly more successfully than other unseen datasets for both feature selection methods.

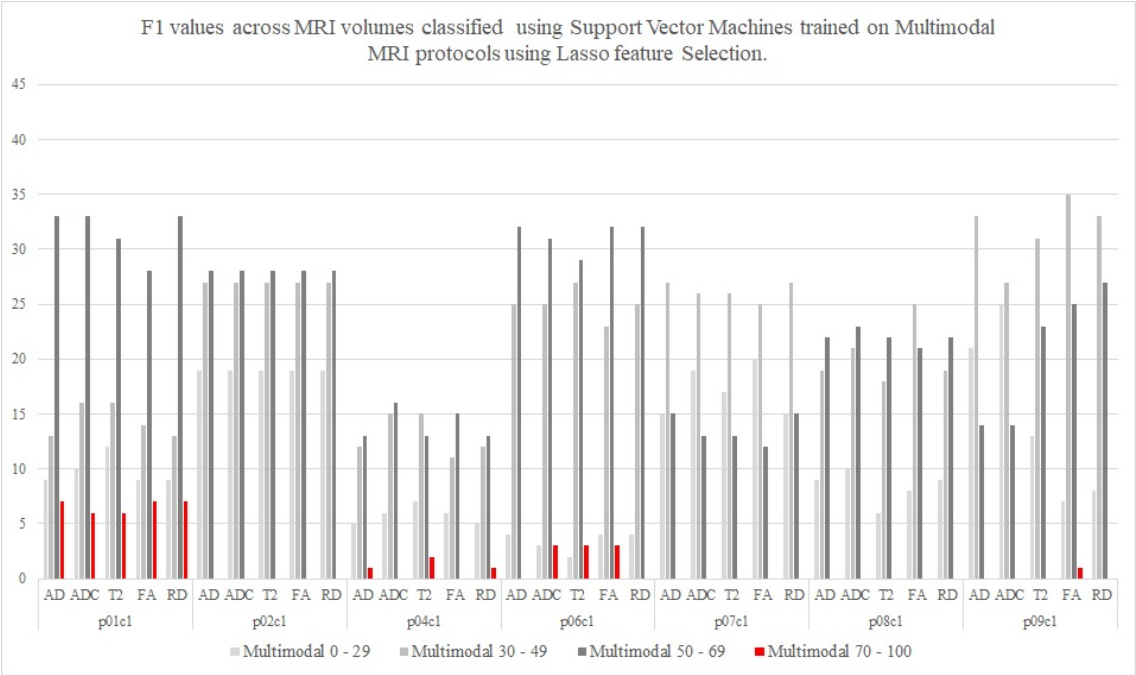


Figure 39: Bar charts displaying binned distributions of slice accuracies for Multimodal SVM experiments using LASSO feature selection. Red represents the number of slices within a volume scoring greater than 70.

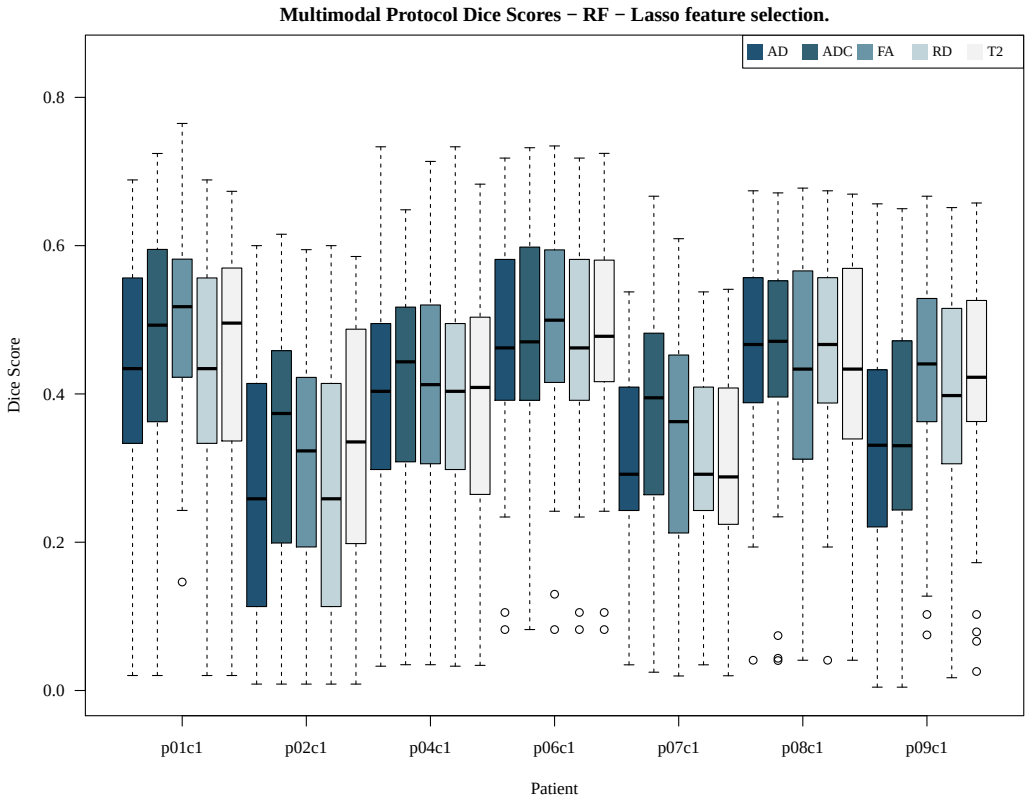


Figure 40: Dice score performance for entire MRI volumes using Random Forest Models trained on Multimodal MRI data and LASSO Feature Selection.

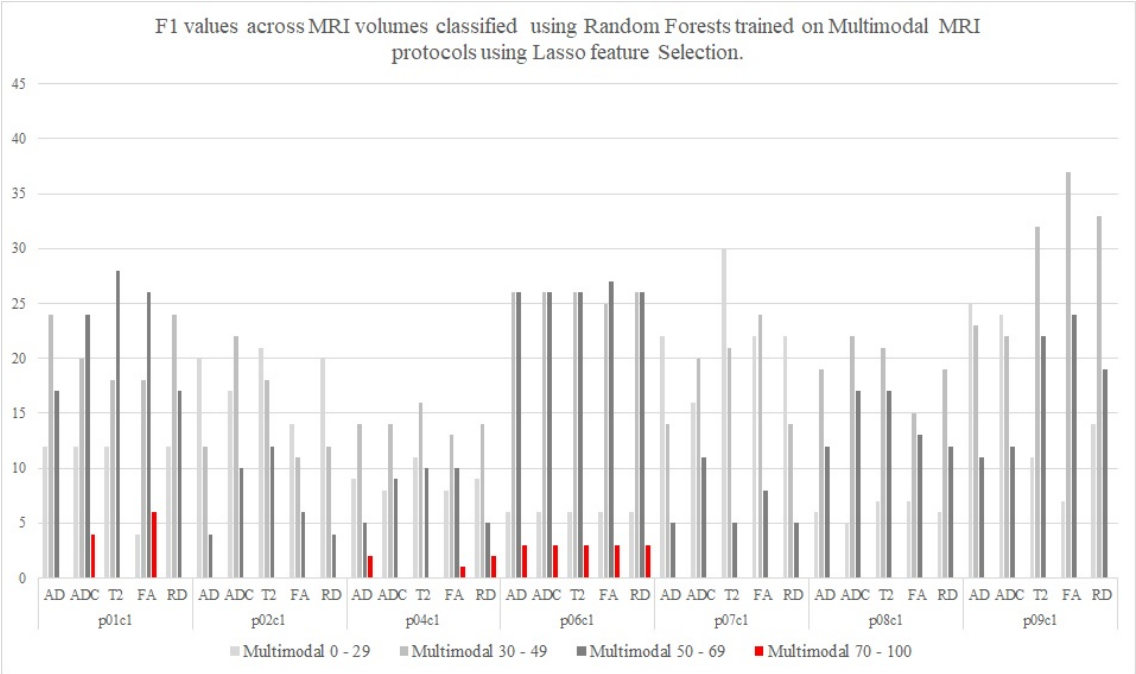


Figure 41: Bar charts displaying binned distributions of slice accuracy for Multimodal RF experiments using LASSO feature selection. Red represents the number of slices within a volume scoring greater than 70.

Bar graphs complimenting these boxplots are shown in fig. 41 for LASSO and appendix B fig. 76b for mRMR feature selection methods.

5.3.3 Extreme Learning Machine

Extreme Learning Machine Dice scores for Multimodal classification volumes are shown in fig. 42 for LASSO and appendix B fig. 77b for mRMR feature selection algorithms. ELM exhibits a higher rate of change between patients and sequences. Median values fluctuate significantly between 0.25 and 0.55, and upper and lower quartile ranges change dramatically between unseen datasets and MRI sequences.

Bar graphs complimenting these boxplots are shown in fig. 43 for LASSO and appendix B fig. 78b for mRMR feature selection methods. A low number of datasets contain slice classifications above 70, and the majority of slices fall between 30 and 69.

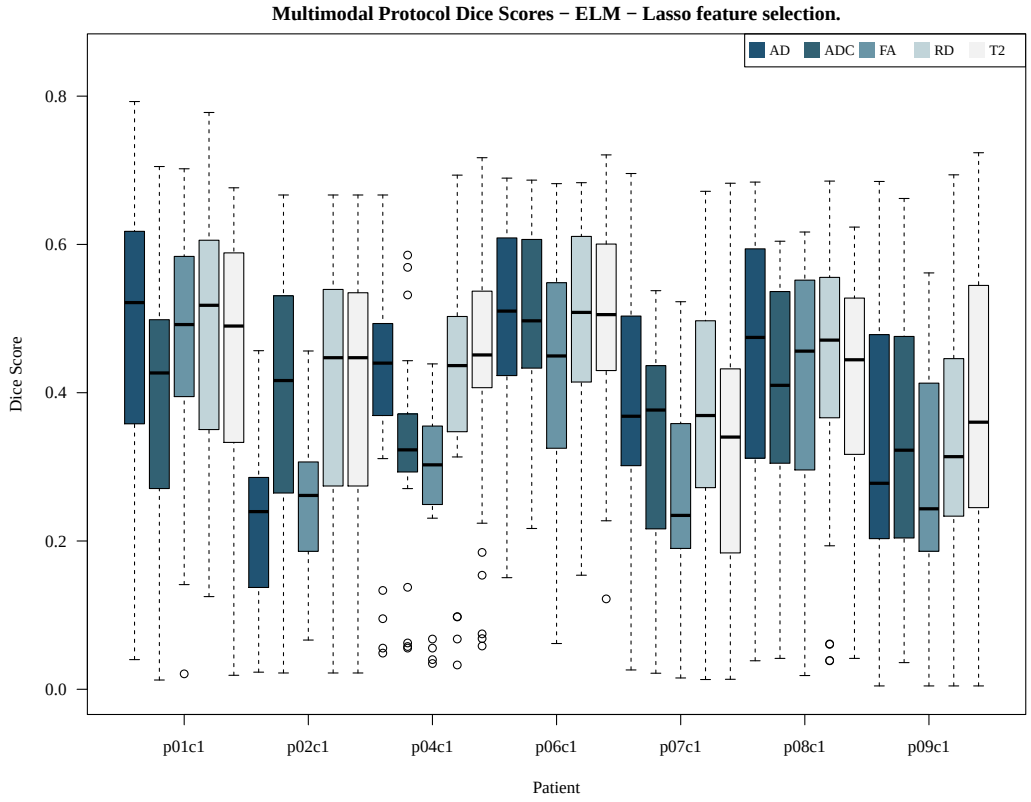


Figure 42: Dice score performance for entire MRI volumes using Extreme Learning Machine Models trained on Multimodal MRI data and LASSO feature Selection.

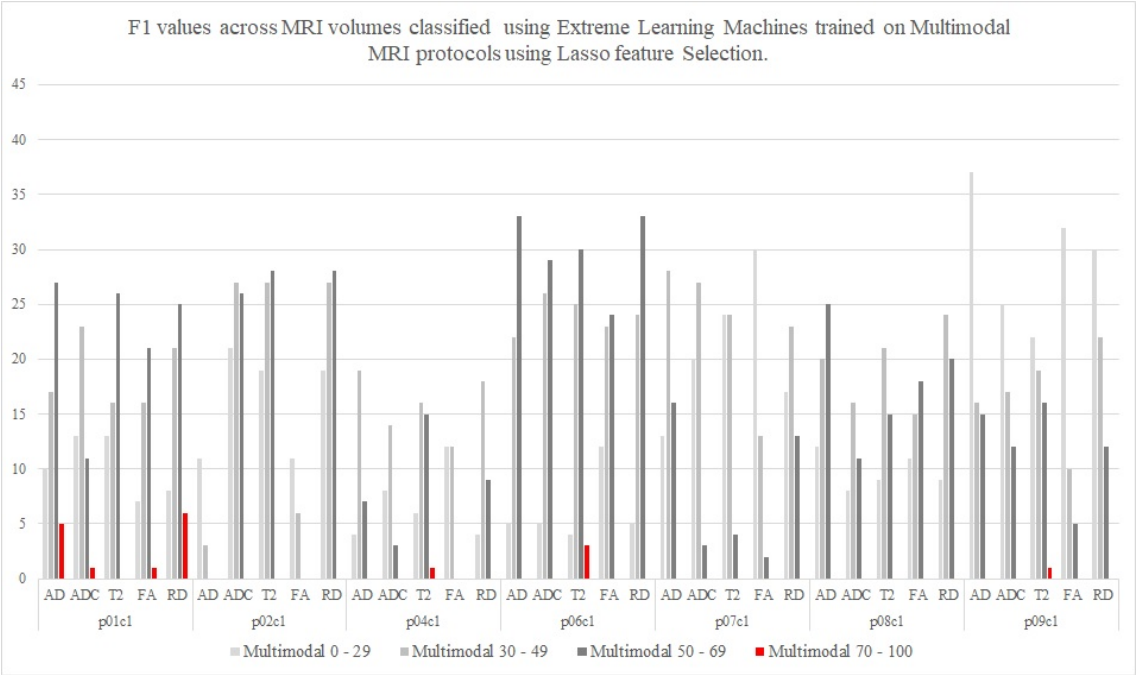


Figure 43: Bar charts displaying binned distributions of slice accuracies for Multimodal ELM experiments using LASSO feature selection. Red represents the number of slices within a volume scoring greater than 70.

5.3.4 Multiple Sclerosis Lesion Segmentation Images

The segmentation success of Multimodal sequence classification using machine learning models trained on all MRI sequences can be seen in fig. 44 and fig. 45. The first figure depicts AD classification of patient p01c1 slice 109 with a clear segmentation of the largest Multiple Sclerosis Lesion regions.

Depicted in fig. 35 is a scan taken from the same dataset as fig. 34, however this time the sequence classified was FA. ELM exhibits the highest F1 score in both instances, which captured the most shape information, with the lowest score being achieved by the SVM.

Multimodal segmentation images of dataset p02c1 AD slice 108 are shown in appendix B fig. 79a and p04c1 ADC 110 in fig. 79b, both using models trained on their respective sequences. In both cases, smaller Lesions are missed leading to lower slice accuracies and sensitivity.

T2 MRI sequence Multimodal results are shown in appendix B fig. 80a for mRMR feature selection and fig. 80b for LASSO feature selection, both of which depict slices from patient dataset p06c1. In both instances, over half of the Lesions are correctly identified as such with a significant portion of both slices identifying MS within larger Lesion regions, while smaller areas are once again missing.

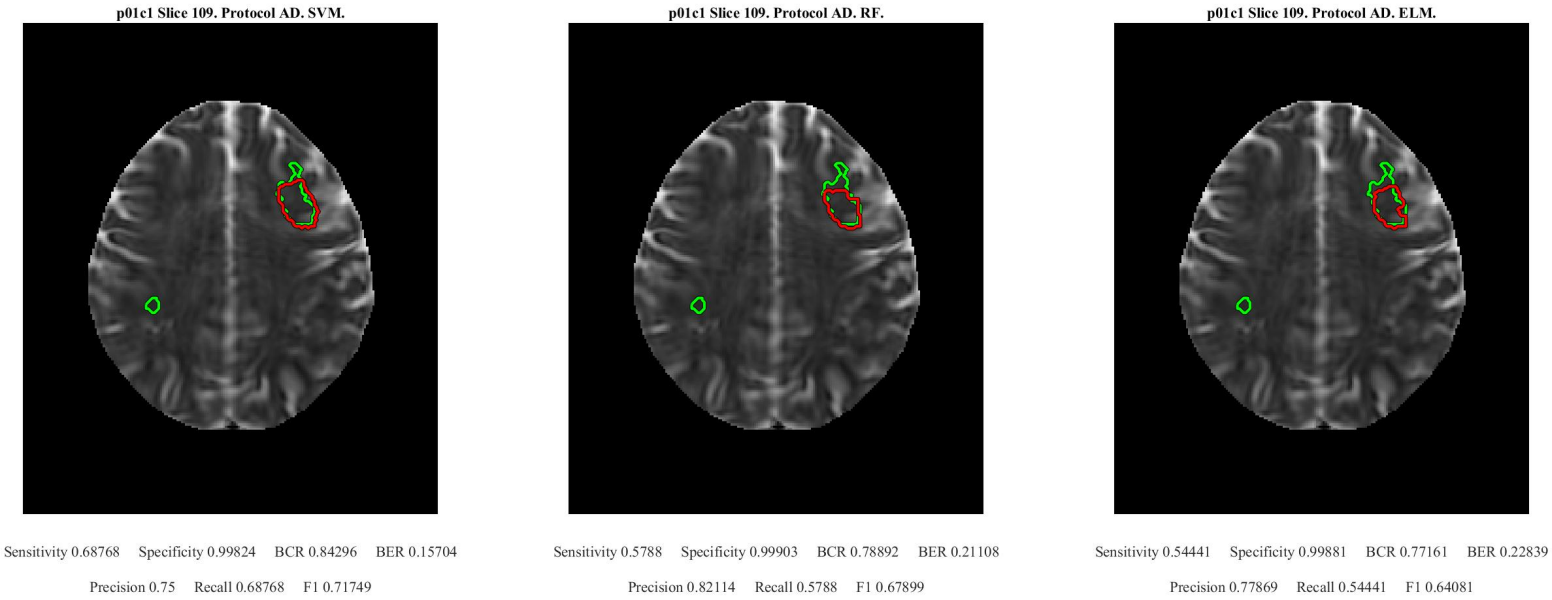


Figure 44: (a) Segmentations using all machine learning classifiers trained on Multimodal MRI data and features selected using LASSO feature selection for MS identification of dataset p01c1 AD sequence on slice 109.

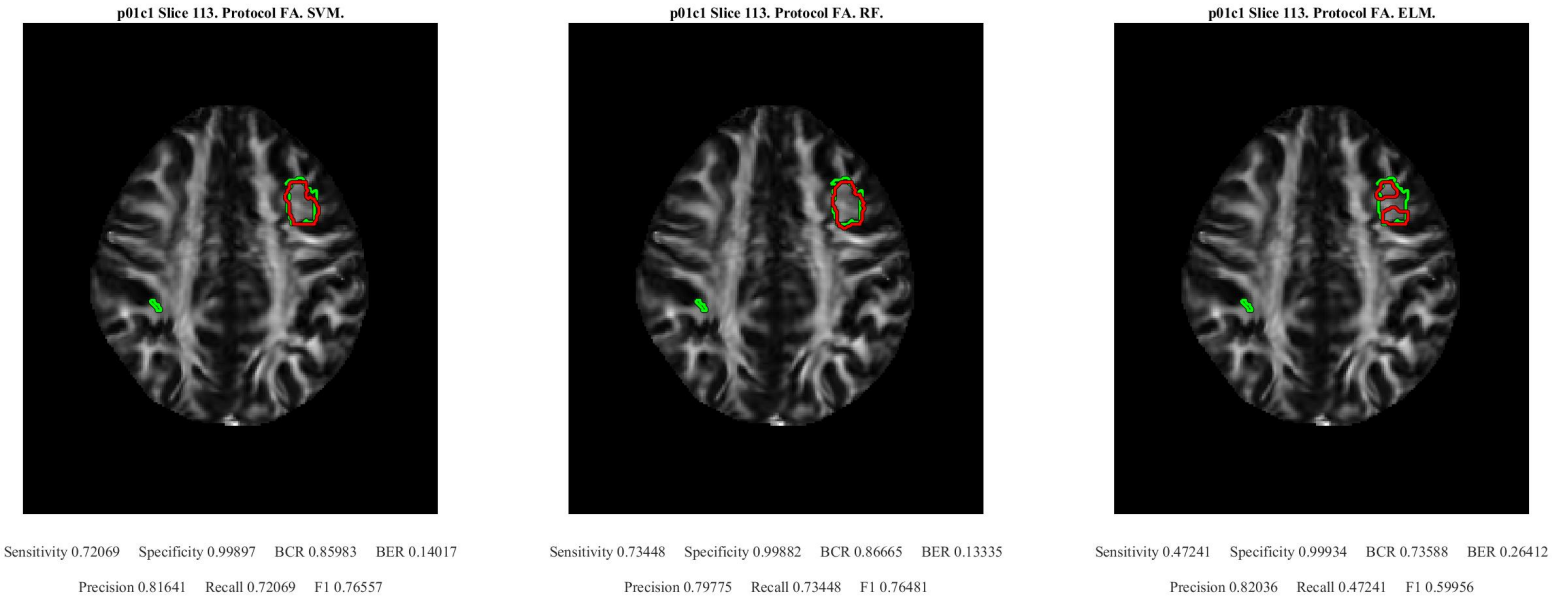


Figure 45: Segmentations using all machine learning classifiers trained on Multimodal MRI data and features selected using LASSO feature selection for MS identification of dataset p01c1 AD sequence on slice 113.

5.3.5 Cross-sectional Heatmap graphs.

This section contains the cross-section heatmap graphs for Multimodal results using LASSO and mRMR feature selection algorithms. Multimodal Cross-sectional graphs of dataset p06c1 are shown in fig. 46 for LASSO and fig. 47 for mRMR feature selection methods. Classification accuracy was greater towards upper-middle regions of the Lesion volume and slices classified less accurately towards edge regions of the Lesion volume in both methods.

Multimodal Cross-sectional graphs of unseen dataset p02c1 are shown in appendix B fig. 81 for LASSO and fig. 82 for mRMR feature selection methods and all machine learning experiments. Maximum classification accuracy peaks around an F1 score of 60 and a significant number of machine learning methods failed for both LASSO and mRMR. Furthermore, in all experiments, significant gaps between slices containing Lesion are present. A significant number of experiments failed consistently across the same slices, most notable around the middle of the brain region.

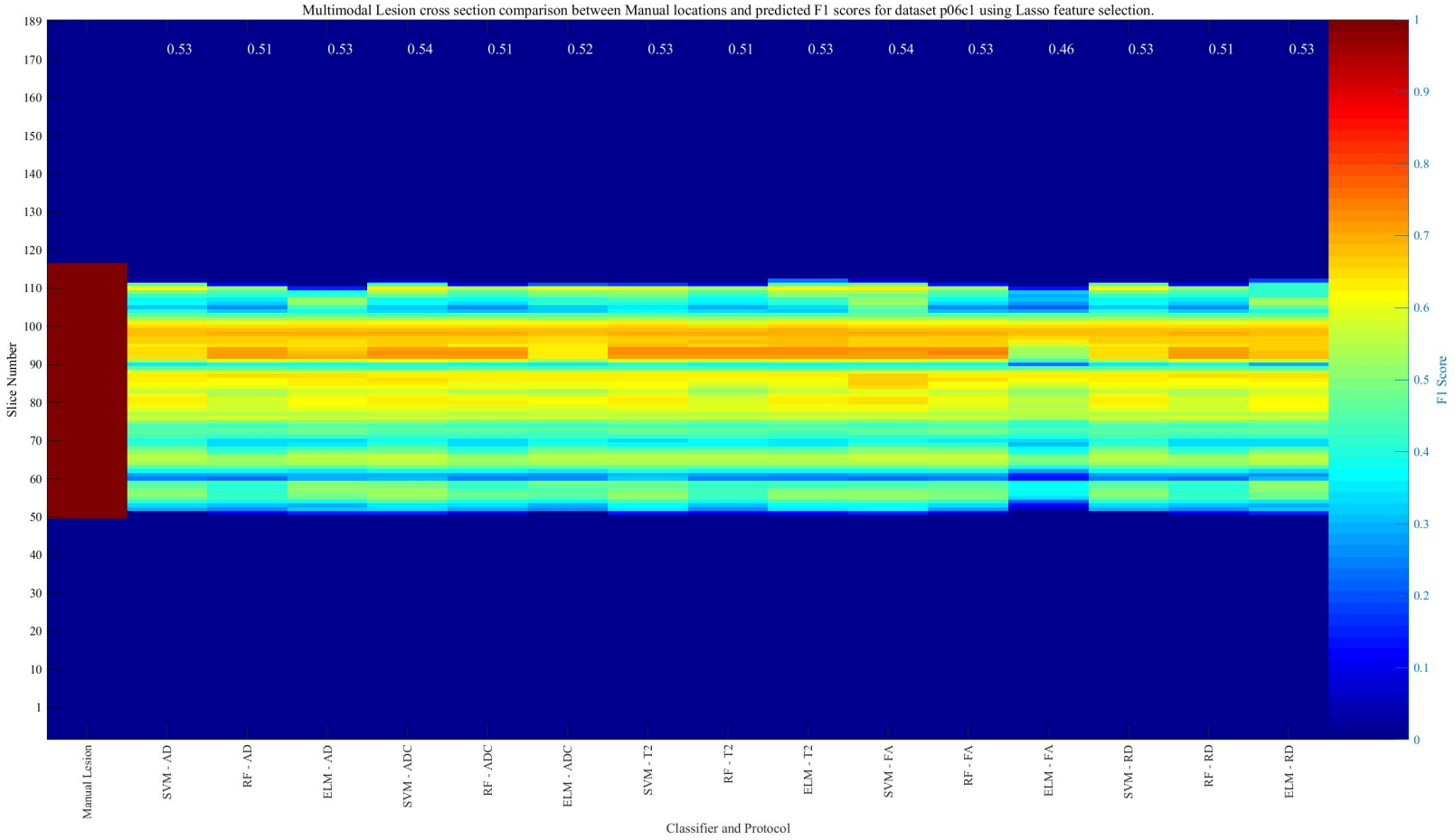


Figure 46: Cross sectional graph for p06c1 using Machine Learning models trained on Multimodal MRI data and LASSO feature selection.

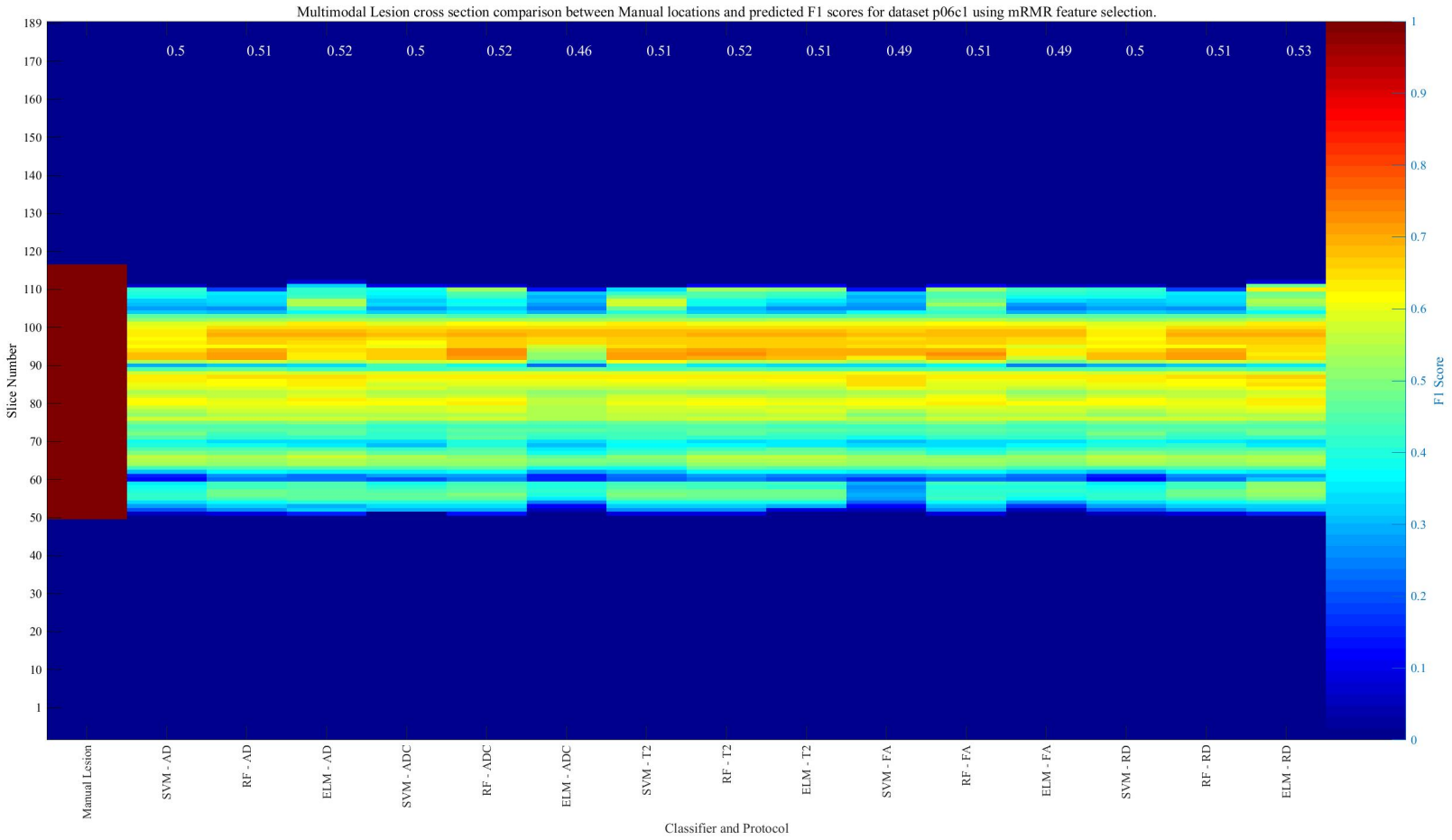


Figure 47: Cross sectional graph for p06c1 using Machine Learning models trained on Multimodal MRI data and mRMR feature selection.

Table 5: Results table for each Feature Selection technique and Machine Learning Model using T2 sequence data.

Method	Model	Data	F1	Sensitivity	Specificity	BER	Precision	Recall
Lasso	SVM	Individual	38.93 ± 7.96	61.58 ± 3.5	99.88 ± 0.04	19.27 ± 1.74	29.09 ± 8.47	61.58 ± 3.5
		Multimodal	46.18 ± 7.04	60.08 ± 3.56	99.9 ± 0.04	20.01 ± 1.77	38.15 ± 8.72	60.08 ± 3.56
	RF	Individual	39.87 ± 9.26	61.41 ± 2.62	99.89 ± 0.04	19.35 ± 1.3	30.37 ± 9.99	61.41 ± 2.62
		Multimodal	39.99 ± 9.84	59.59 ± 3.47	99.89 ± 0.04	20.26 ± 1.73	30.97 ± 10.27	59.59 ± 3.47
	ELM	Individual	38.69 ± 9.97	61.39 ± 4.05	99.88 ± 0.04	19.36 ± 2.02	29.22 ± 10.38	61.39 ± 4.05
		Multimodal	41.26 ± 8.55	61.62 ± 3.17	99.89 ± 0.04	19.24 ± 1.58	31.85 ± 9.42	61.62 ± 3.17
mRMR	SVM	Individual	39.46 ± 7.91	60.8 ± 2.88	99.89 ± 0.04	19.66 ± 1.43	29.83 ± 8.31	60.8 ± 2.88
		Multimodal	44.58 ± 6.95	60.7 ± 3.56	99.89 ± 0.04	19.7 ± 1.77	35.84 ± 8.08	60.7 ± 3.56
	RF	Individual	38.49 ± 9.94	61.05 ± 2.71	99.88 ± 0.04	19.53 ± 1.35	29.01 ± 10.5	61.05 ± 2.71
		Multimodal	41.41 ± 8.46	61.41 ± 2.46	99.89 ± 0.04	19.35 ± 1.23	31.9 ± 9.25	61.41 ± 2.46
	ELM	Individual	40.05 ± 6.36	60.67 ± 3.1	99.89 ± 0.04	19.72 ± 1.54	30.25 ± 6.72	60.67 ± 3.1
		Multimodal	39.39 ± 8.24	62.08 ± 3.13	99.88 ± 0.04	19.02 ± 1.56	29.55 ± 8.9	62.08 ± 3.13

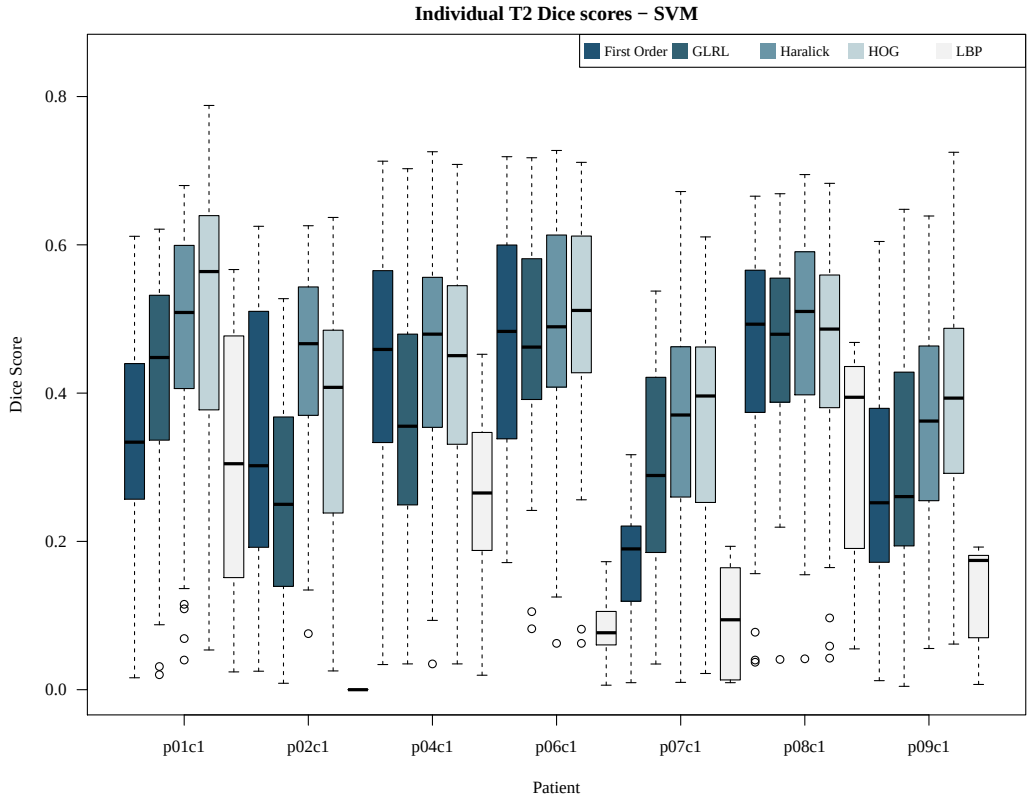


Figure 48: Individual sequence Dice Scores using Support Vector Machine Classifier applied on Individual feature sets using leaveout patient datasets T2 MRI sequence.

5.4 Unseen Dataset Classification - Feature Groups

This section presents the results of the second experiment described in section 4.4.5 for comparison between groups of extracted features. The results depicted in all figures are for experiments classifying T2 MRI scans. Results from other MRI sequences are found in the appendices. The final accuracies across all MRI sequences are provided at the end of the results section.

5.4.1 Support Vector Machine

Results of classification using feature groups are shown in fig. 48 for SVM trained on Individual T2 sequence data, and appendix C fig. 83b for SVM trained on Multimodal MRI data. The interquartile range of dice scores varies greatly between feature extraction techniques and patient datasets. Volumes using LBP features and Multimodal SVM fail and Individual SVM fails for p02c1 LBP features, with low values existing in LBP features for p06c1, p07c1 and p09c1.

F1 distribution charts for Individual fig. 49 and Multimodal appendix C fig. 84b depict the groupings of slice results across each brain volume, and the top slice scores from these volumes are depicted in fig. 50 for both Individual and Multimodal experiments.

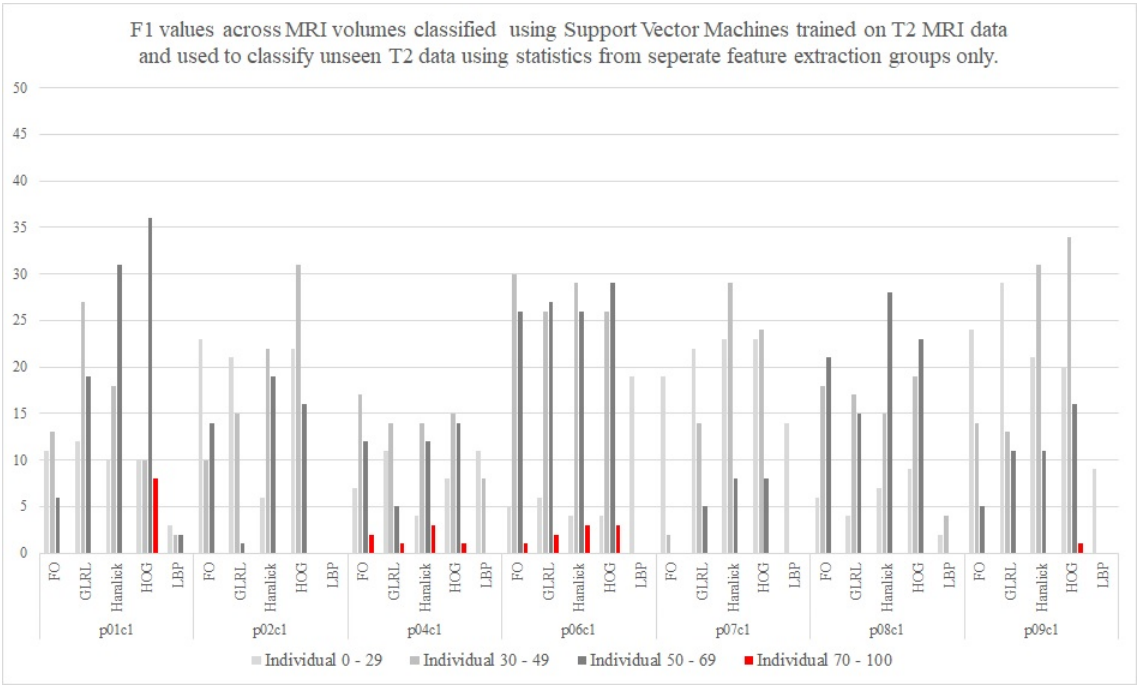


Figure 49: Bar charts displaying binned distributions of slice accuracies for Individual SVM experiments using individual feature extraction techniques. Red represents the number of slices within a volume scoring greater than 70.

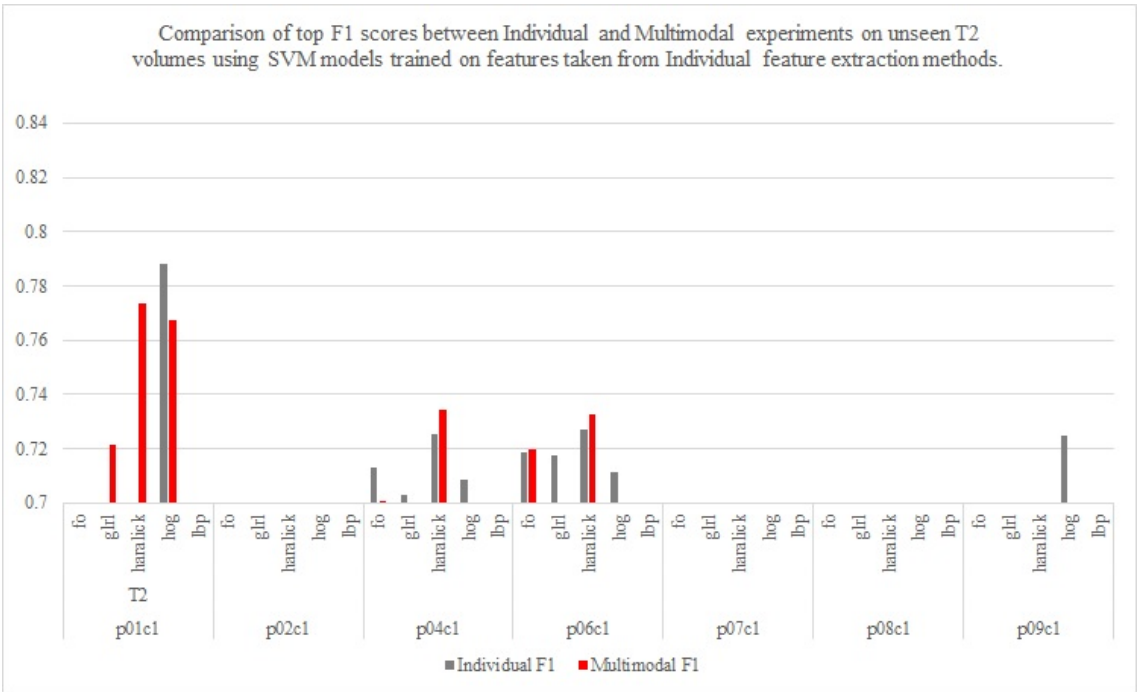


Figure 50: Bar charts displaying top slice scores for Individual and Multimodal SVM experiments using individual sets of features.

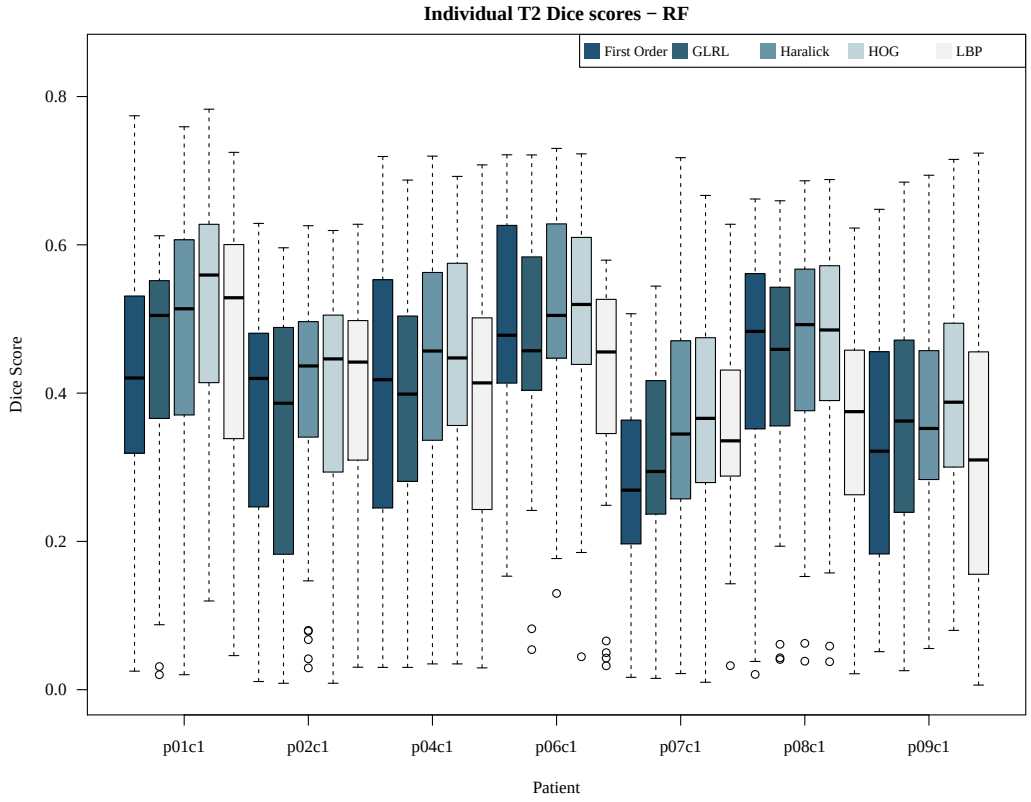


Figure 51: Individual sequence Dice Scores using Random Forest Classifier applied on Individual feature sets using leaveout patient datasets T2 MRI sequence.

5.4.2 Random Forest

Dice Scores for unseen T2 patient data classification using Random Forests trained on T2 data is shown in fig. 51 and Multimodal MRI data in appendix C fig. 85b. Random Forest successfully classified all features, with median values between 30 and 55 and maximum similarities achieving close to 80 in both Individual and Multimodal experiments.

Distribution charts showing the occurrences of F1 scores are shown in fig. 52 and appendix C fig. 86b for Individual and Multimodal experiments. A significant number of slice values were classified with an F1 score between 50 and 70 in both Individual and Multimodal cases, and a comparison between maximum slice values from each MRI volume can be found in fig. 53.

5.4.3 Extreme Learning Machine

Results of classification using feature groups are shown in fig. 54 for ELM trained on T2 sequence data, and appendix C fig. 87b for ELM trained on Multimodal MRI data. Maximum slice scores surpass 80% and median F1 slice scores fall between 0.2 and 0.6. Datasets p01c1, p06c1, and p08c1's interquartile ranges exceed all other leave out datasets.

Distribution charts for ELM are shown in fig. 55 and appendix C fig. 88b for Individual and Multimodal experiments and maximum slice charts are shown in fig. 56.

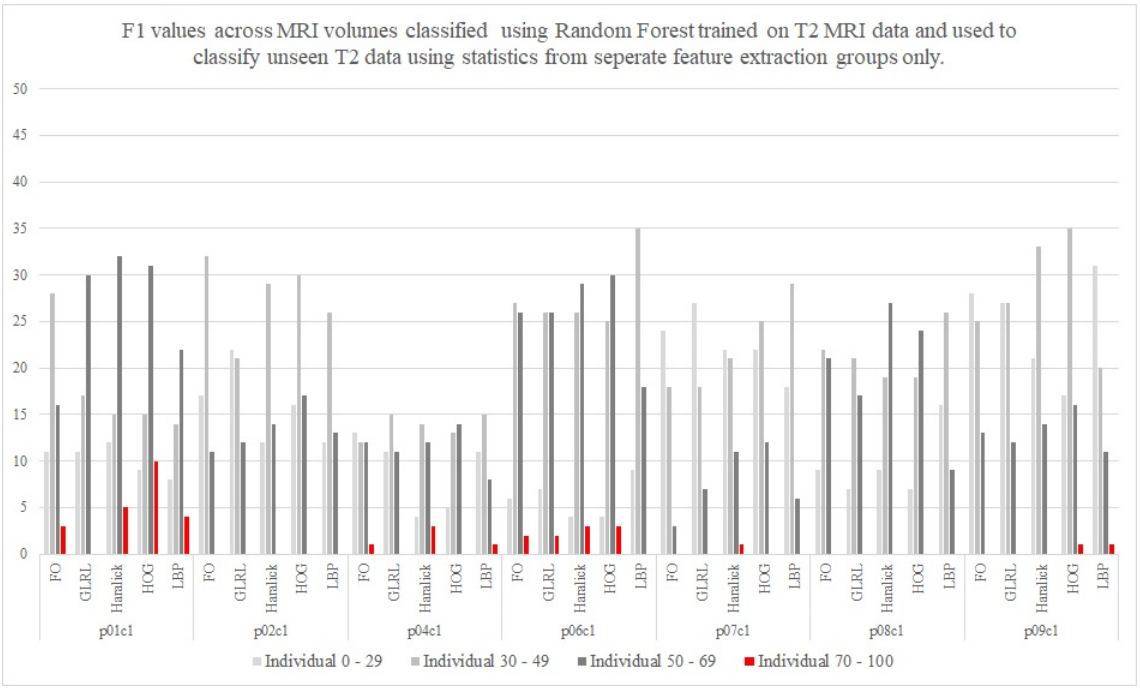


Figure 52: Bar charts displaying binned distributions of slice accuracies for Individual RF experiments using individual feature extraction techniques. Red represents the number of slices within a volume scoring greater than 70.

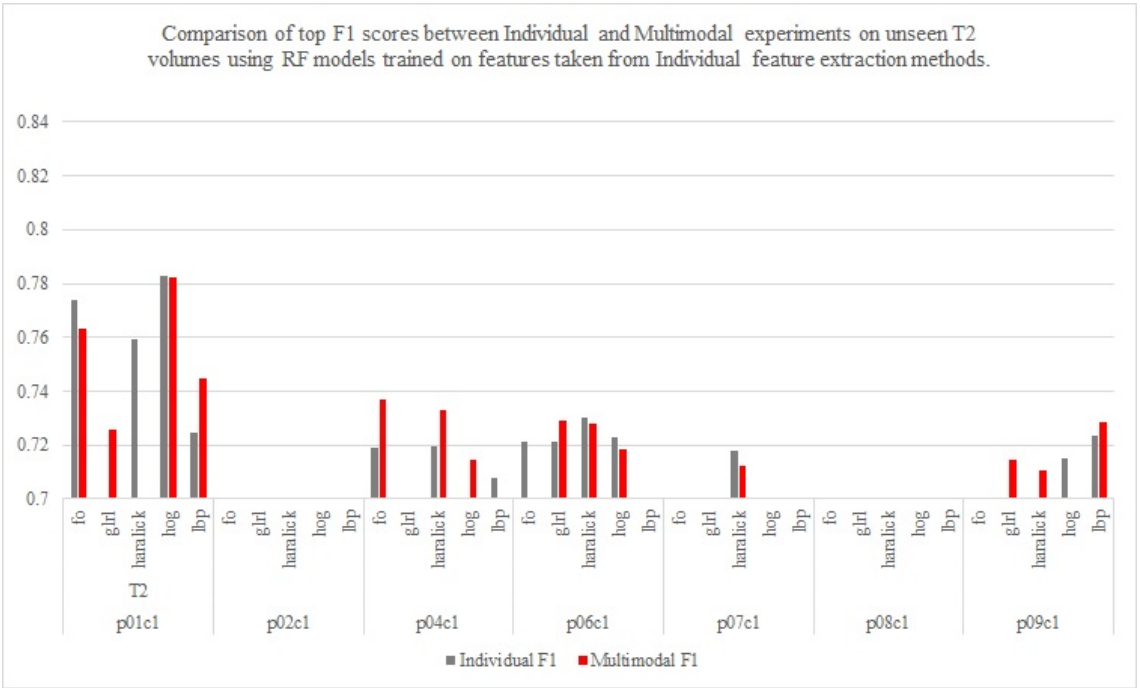


Figure 53: Bar charts displaying top slice scores for Individual and Multimodal RF experiments using individual sets of features.

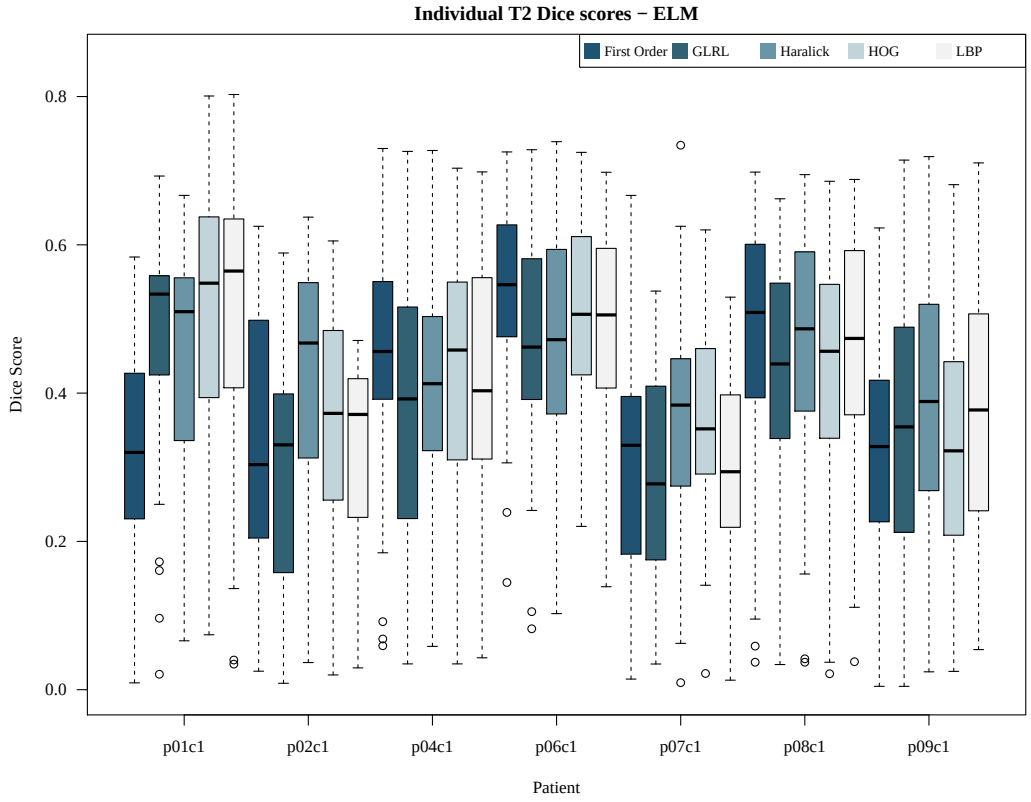


Figure 54: Individual sequence Dice Scores using Extreme Learning Machine Classifier applied on Individual feature sets using leaveout patient datasets T2 MRI sequence.

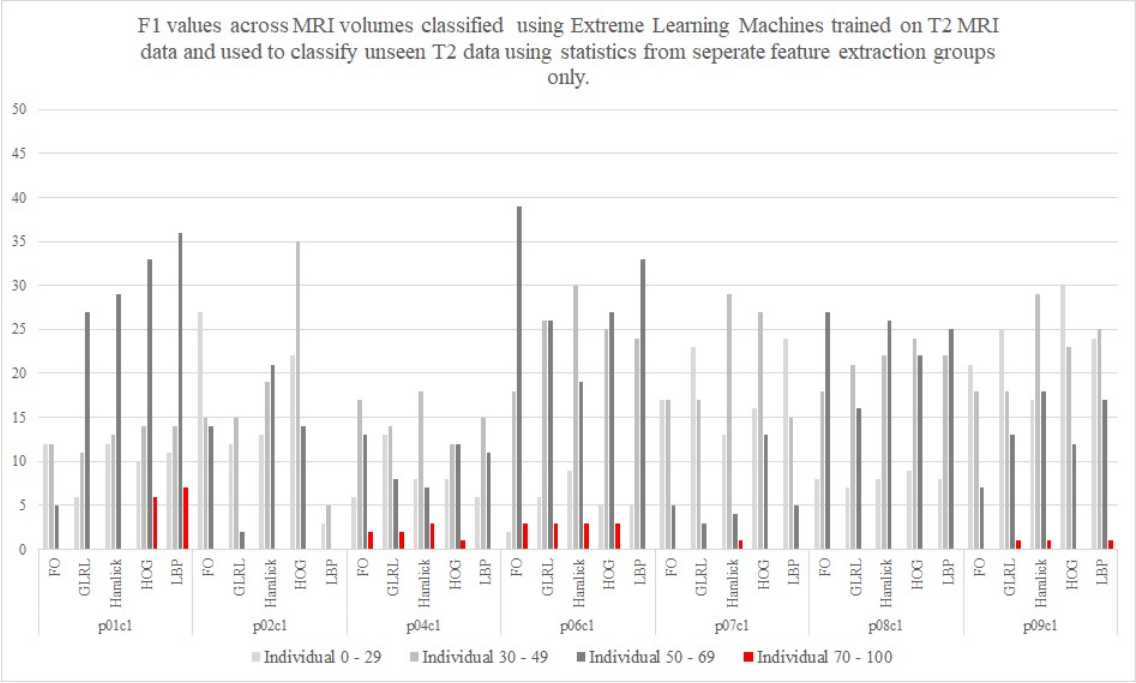


Figure 55: Bar charts displaying binned distributions of slice accuracies for Individual ELM experiments using individual feature extraction techniques. Red represents the number of slices within a volume scoring greater than 70.

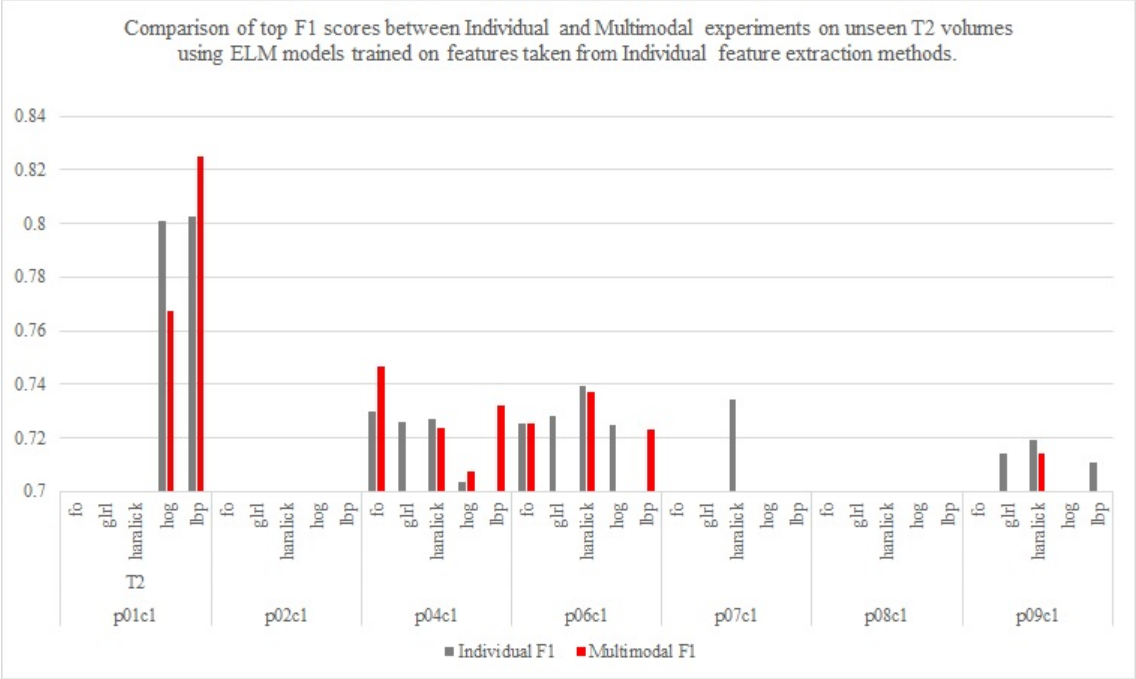


Figure 56: Bar charts displaying top slice scores for Individual and Multimodal ELM experiments using individual sets of features.

5.4.4 Multiple Sclerosis Lesion Segmentation Images

The following section presents a range of Binary Images predicted using both Individually trained and Multimodal Machine Learning Models.

Two patients are selected, p01c1 and p08c1. Patient p01c1 is included owing to the larger Lesion sizes, and overall success in classification, while p08c1 is included owing to the significantly lower Lesion Load to indicate potential failings in this research which are discussed in section 6.2.1.

Segmentation's for p01c1 slice number 113 using First Order Statistics are shown in fig. 57 for Machine Learning Models trained on T2 MRI data only, and fig. 58 for Machine Learning Models trained on Multimodal MRI data. The same setting but using grey level run length statistics are shown in figure appendix C fig. 89a for Individual models and appendix C fig. 89b for Multimodal models. Classification using Haralick Texture features on p01c1 slice 113 and Individual models are shown in appendix C fig. 90a and Multimodal models in appendix C fig. 90b. Histogram of Oriented Gradients are shown in appendix C fig. 91a for Individual and appendix C fig. 91b for Multimodal. Finally, Local Binary Pattern T2 segmentations of p01c1 slice 113 are shown in appendix C fig. 92a for Individual T2 models and appendix C fig. 92b for Multimodal machine learning models.

Segmentations for p08c1 slice number 73 using First Order Statistics are shown in appendix C fig. 93a for Machine Learning Models trained on T2 MRI data only, and appendix C fig. 93b for Machine Learning Models trained on Multimodal MRI data. The same setting but using grey level run length statistics are shown in figure appendix C fig. 94a for Individual models and appendix C fig. 94b for Multimodal models. Classification using Haralick Texture features on p08c1 slice 73 and Individual models are shown in appendix C fig. 95a and Multimodal models in appendix C fig. 95b. Histogram of Oriented Gradients are shown in appendix C fig. 96a for Individual and appendix C fig. 96b for Multimodal. Finally, Local Binary Pattern T2 segmentations of p08c1 slice 73 are shown in appendix C fig. 97a for Individual T2 models and appendix C fig. 97b for Multimodal machine learning models.

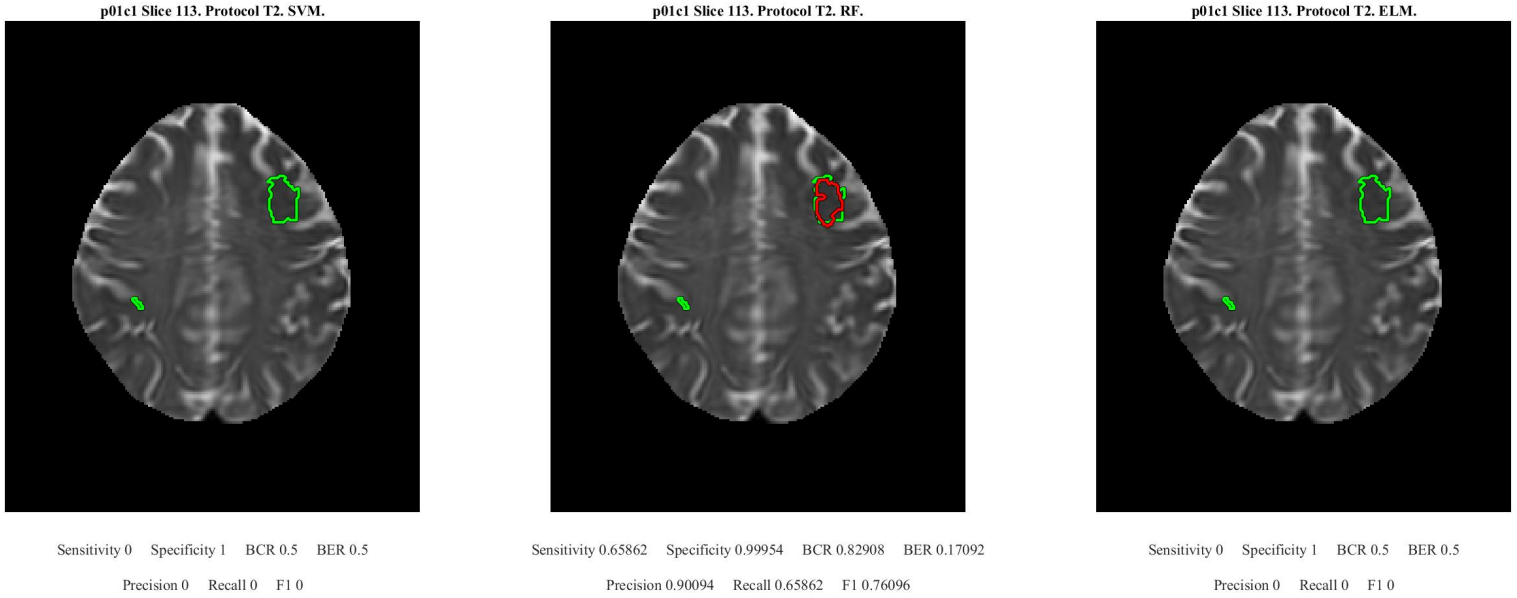


Figure 57: Individual sequence Segmented regions using all classifiers trained using only first-order statistics from T2 MRI sequence for unseen segmentation using dataset p01c1 T2 sequence on slice 113.

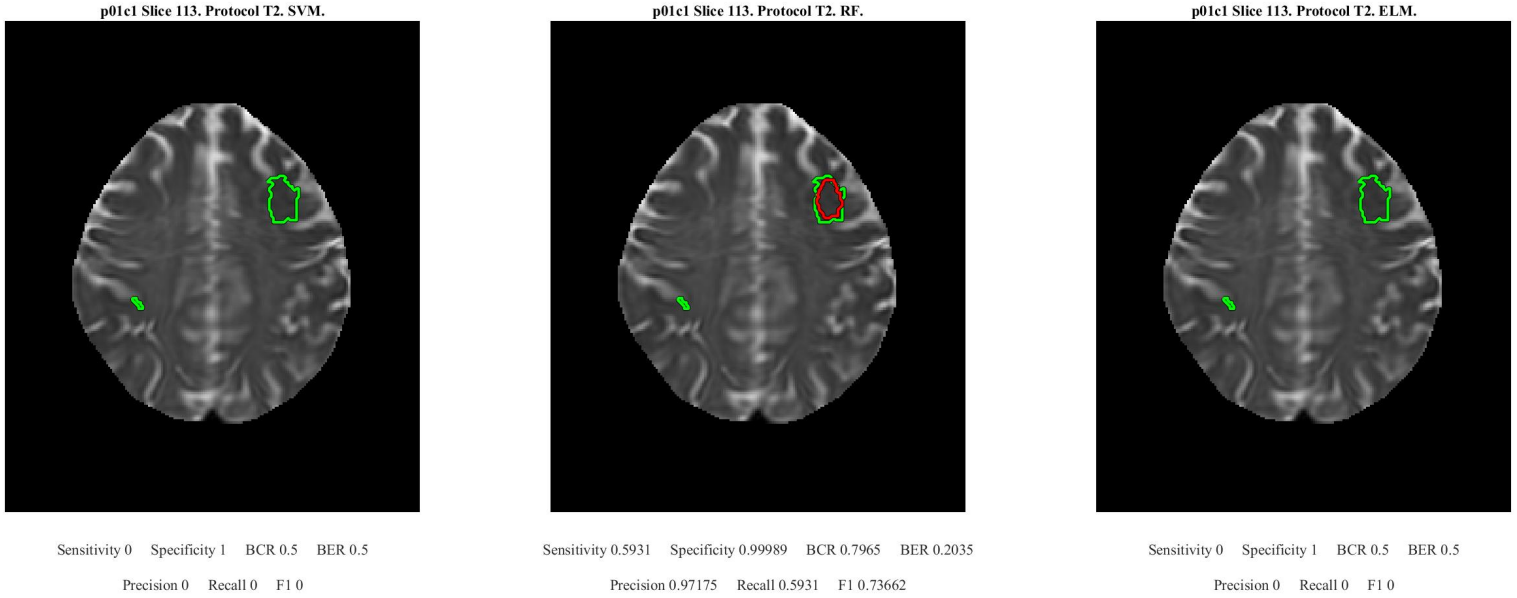


Figure 58: Multimodal sequence Segmented regions using all classifiers trained using only first-order statistics from Multimodal MRI sequences for unseen segmentation using dataset p01c1 T2 sequence on slice 113.

5.4.5 Cross-sectional Heatmap graphs

Individual feature results for patient dataset p01c1 using all MRI sequences and all Machine Learning models trained on Individual MRI sequence data are shown in fig. 59 for First Order Statistics, appendix C fig. 98 for Grey Level Run Lengths, appendix C fig. 100 for Haralick Texture features, appendix C fig. 102 for Histogram of Oriented Gradients and appendix C fig. 104 for Local Binary Patterns.

The results of the same experiment using machine learning models trained on Multimodal MRI data are shown in fig. 60 for First Order Statistics, appendix C fig. 99 for Grey Level Run Lengths, appendix C fig. 101 for Haralick Texture features, appendix C fig. 103 for Histogram of Oriented Gradients and appendix C fig. 105 for Local Binary Patterns.

All feature extraction methods exhibit increased classification accuracy towards upper regions of the main Lesion volume, yet struggle with classification towards the edges. Extreme Learning Machine and Support Vector Machine using T2 MRI data scores particularly low when using only First Order Statistics, however, the lowest performance is shown in Local Binary Patterns which achieve significantly lower results compared to all other feature extraction methods in dataset p01c1.

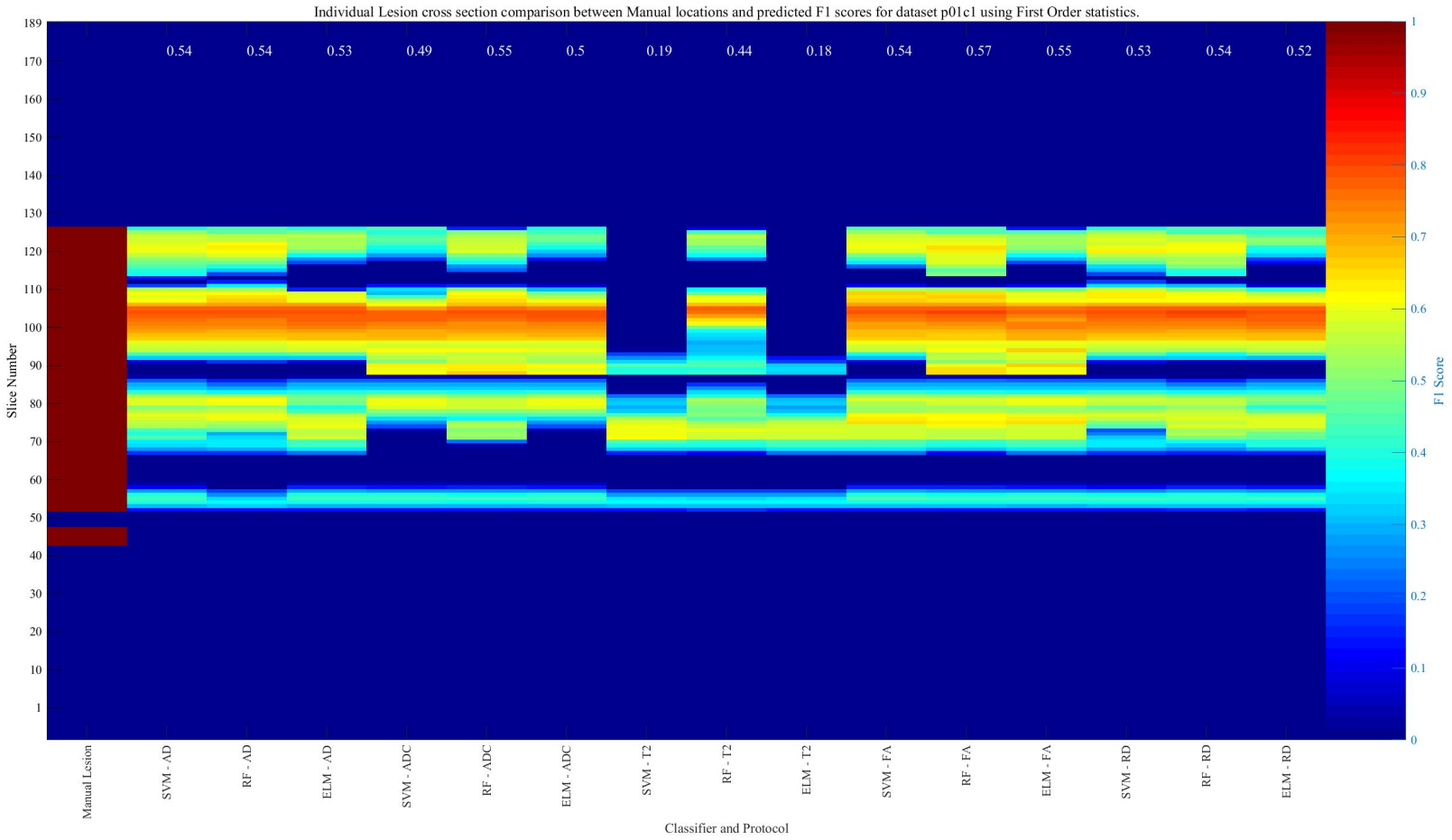


Figure 59: Cross sectional graph for p01c1 using Machine Learning models trained on Individual sequence First Order MRI data.

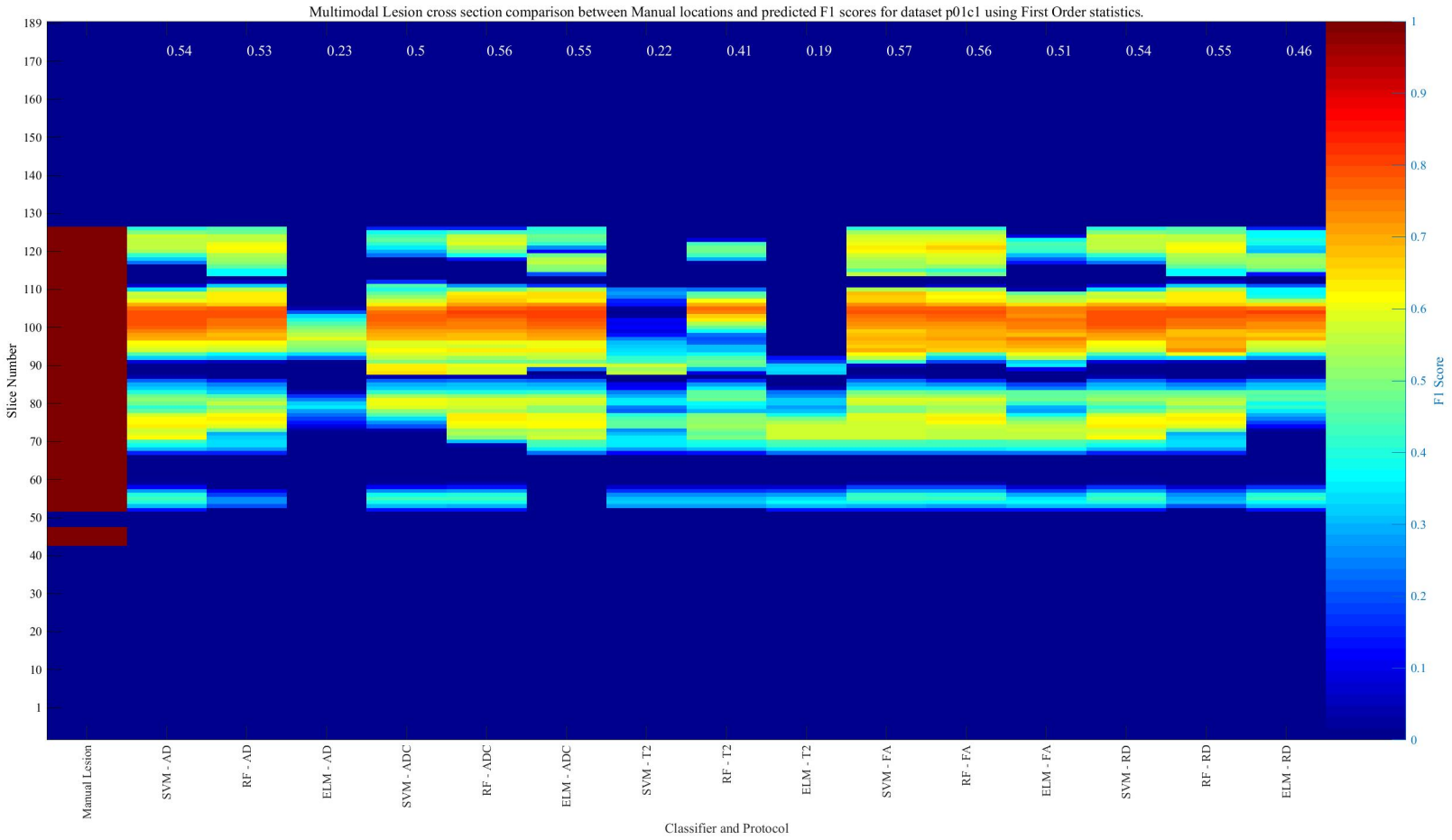


Figure 60: Cross sectional graph for p01c1 using Machine Learning models trained on Multimodal First Order MRI data.

The results for patient dataset p02c1 using Machine Learning models trained on Individual MRI sequence data are shown in fig. 61 for First Order Statistics, appendix C fig. 106 for Grey Level Run Lengths, appendix C fig. 108 for Haralick Texture features, appendix C fig. 110 for Histogram of Oriented Gradients and appendix C fig. 112 for Local Binary Patterns.

The results of the same experiment using models trained on Multimodal MRI data are shown in fig. 62 for First Order Statistics, appendix C fig. 107 for Grey Level Run Lengths, appendix C fig. 109 for Haralick Texture features, appendix C fig. 111 for Histogram of Oriented Gradients and appendix C fig. 113 for Local Binary Patterns.

Dataset p02c1 contains a significant number of missing slices across all feature extraction experiments with large numbers of slices missing across all Lesion volumes. In all figures, the highest slice scores and most consistent grouping is towards the lower slices within the coronal plane. The highest slice accuracies are shown using Multimodal Random Forest for T2 and FA sequences when classifying Histogram of Oriented Gradients appendix C fig. 111, as well as in Random Forest T2, ELM T2 and ELM FA classification using Local Binary Patterns features only appendix C fig. 113.

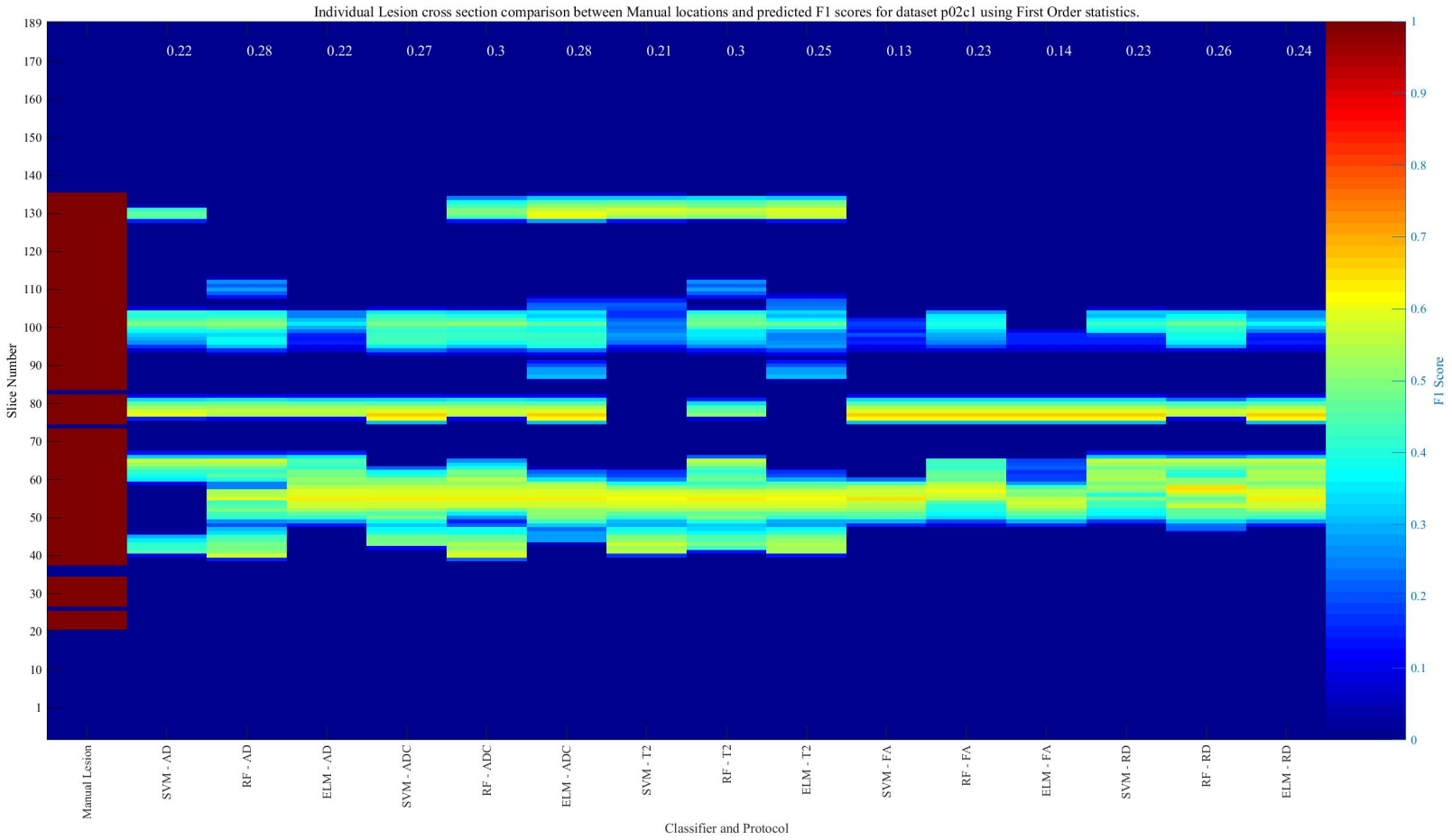


Figure 61: Cross sectional graph for p02c1 using Machine Learning models trained on Individual sequence First Order MRI data.

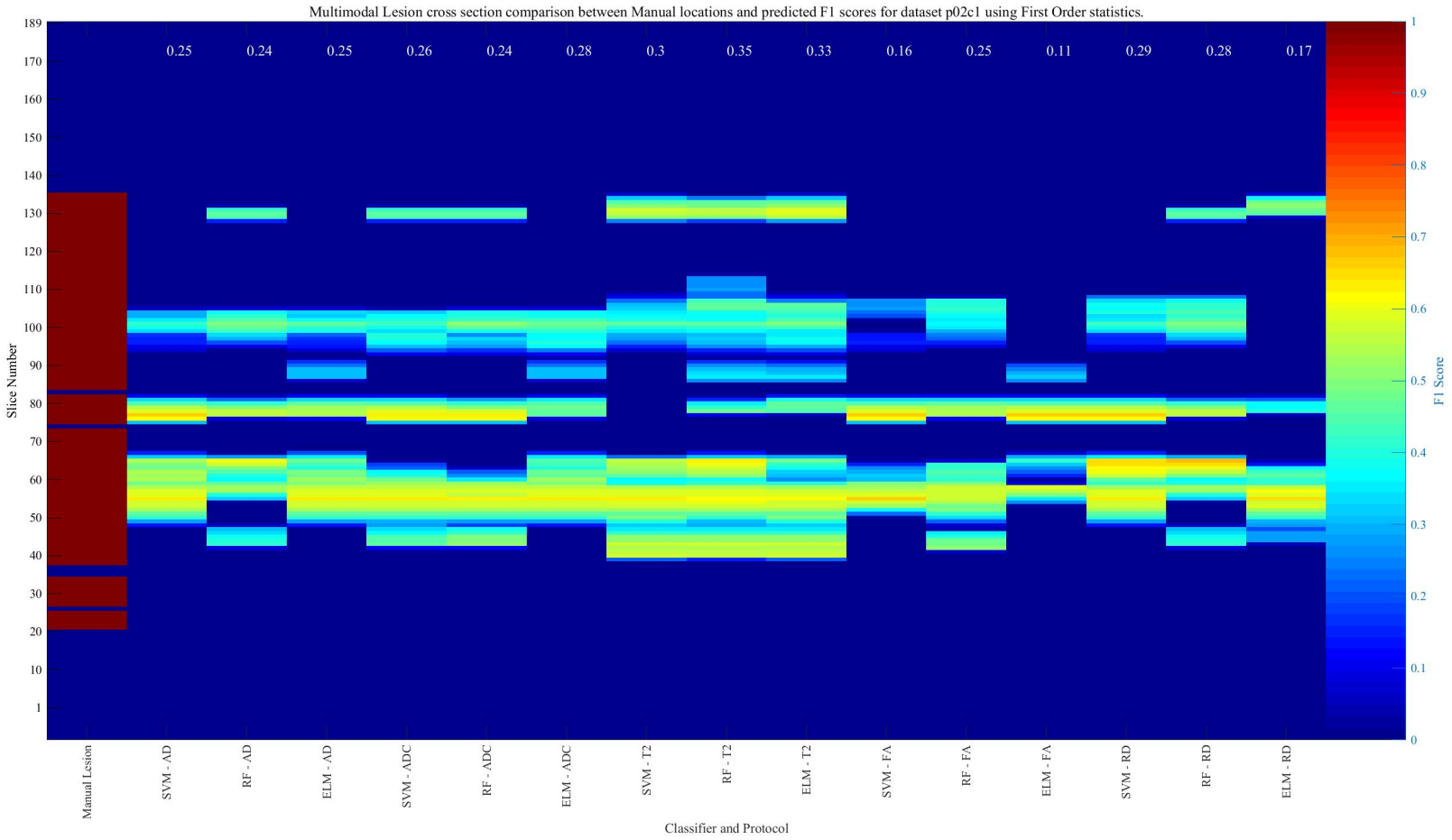


Figure 62: Cross sectional graph for p02c1 using Machine Learning models trained on Multimodal First Order MRI data.

Table 6: Results table for each Feature Extraction technique and Machine Learning Model using T2 sequence data.

Features	Data	Model	F1	Sensitivity	Specificity	BER	Precision	Recall
First Order	Individual	SVM	30.19 ± 15.53	62.38 ± 4.17	99.87 ± 0.04	18.87 ± 2.08	21.64 ± 13.71	62.38 ± 4.17
		RF	38.26 ± 10.18	64 ± 3.01	99.88 ± 0.04	18.06 ± 1.51	28.19 ± 10.03	64 ± 3.01
		ELM	34.72 ± 14.74	60.68 ± 4.49	99.88 ± 0.03	19.72 ± 2.24	26.57 ± 15.77	60.68 ± 4.49
	Multimodal	SVM	38.12 ± 13	60.48 ± 4.04	99.89 ± 0.03	19.82 ± 2.01	29.72 ± 14.65	60.48 ± 4.04
		RF	42.74 ± 8.96	61.44 ± 3.09	99.89 ± 0.03	19.33 ± 1.54	33.83 ± 11.12	61.44 ± 3.09
		ELM	39.26 ± 13.8	58.53 ± 3.75	99.89 ± 0.03	20.79 ± 1.87	31.65 ± 16.22	58.53 ± 3.75
GLRL	Individual	SVM	34.96 ± 11.61	60.09 ± 3.59	99.88 ± 0.04	20.02 ± 1.79	25.71 ± 11.3	60.09 ± 3.59
		RF	39.41 ± 8.73	60.46 ± 2.17	99.89 ± 0.04	19.83 ± 1.07	29.96 ± 9.28	60.46 ± 2.17
		ELM	36.39 ± 11.4	61.45 ± 2.69	99.88 ± 0.04	19.33 ± 1.34	26.92 ± 11.07	61.45 ± 2.69
	Multimodal	SVM	38.4 ± 13.73	59.23 ± 2.34	99.89 ± 0.03	20.44 ± 1.17	30.56 ± 15.94	59.23 ± 2.34
		RF	41.47 ± 12.09	59.72 ± 1.16	99.89 ± 0.03	20.19 ± 0.58	33.71 ± 14.83	59.72 ± 1.16
		ELM	31.68 ± 12.21	59.81 ± 4.94	99.88 ± 0.04	20.16 ± 2.46	22.52 ± 11.15	59.81 ± 4.94
Haralick	Individual	SVM	42.28 ± 8.42	62.26 ± 3.65	99.89 ± 0.04	18.92 ± 1.82	32.65 ± 9.07	62.26 ± 3.65
		RF	43.15 ± 8.99	62.16 ± 3.02	99.89 ± 0.04	18.97 ± 1.5	33.82 ± 10.03	62.16 ± 3.02
		ELM	39.37 ± 8.67	62.25 ± 3.35	99.88 ± 0.04	18.93 ± 1.67	29.39 ± 8.7	62.25 ± 3.35
	Multimodal	SVM	47.96 ± 8.2	60.41 ± 2.35	99.9 ± 0.03	19.84 ± 1.18	40.52 ± 10.36	60.41 ± 2.35
		RF	46.65 ± 8.45	61.15 ± 2.88	99.9 ± 0.03	19.47 ± 1.44	38.52 ± 10.48	61.15 ± 2.88
		ELM	48.02 ± 7.76	58.94 ± 2.76	99.91 ± 0.03	20.58 ± 1.38	41.11 ± 9.75	58.94 ± 2.76
HoG	Individual	SVM	44.2 ± 8.44	62.04 ± 2.21	99.89 ± 0.04	19.03 ± 1.1	35.05 ± 10.01	62.04 ± 2.21
		RF	44.26 ± 8.38	61.81 ± 2.65	99.89 ± 0.04	19.15 ± 1.32	35.14 ± 9.85	61.81 ± 2.65
		ELM	42.68 ± 8.56	60.9 ± 3.6	99.89 ± 0.04	19.6 ± 1.8	33.44 ± 9.7	60.9 ± 3.6
	Multimodal	SVM	31.91 ± 8.37	67.09 ± 2.66	99.87 ± 0.04	16.52 ± 1.33	21.36 ± 6.99	67.09 ± 2.66
		RF	48.05 ± 8.3	60.51 ± 2.43	99.9 ± 0.03	19.79 ± 1.21	40.54 ± 10.4	60.51 ± 2.43
		ELM	47.39 ± 7.17	60.35 ± 2.7	99.9 ± 0.03	19.88 ± 1.35	39.53 ± 8.85	60.35 ± 2.7
LBP	Individual	SVM	3.99 ± 4.32	65.1 ± 29.45	99.84 ± 0.07	10.38 ± 5.63	2.12 ± 2.4	65.1 ± 29.45
		RF	36.66 ± 7.5	64.7 ± 3.52	99.88 ± 0.05	17.71 ± 1.76	25.87 ± 6.98	64.7 ± 3.52
		ELM	38.11 ± 17.4	63.22 ± 3.46	99.89 ± 0.04	18.44 ± 1.72	30.43 ± 17.22	63.22 ± 3.46
	Multimodal	SVM	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		RF	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		ELM	47.27 ± 6.29	61.04 ± 4.74	99.9 ± 0.03	19.53 ± 2.37	38.79 ± 7.06	61.04 ± 4.74

6 Discussion

In this research three machine learning models were trained using a Multiple Sclerosis dataset provided by the University Hospital of Nancy, consisting of seven patient sets comprising five MRI sequences (Axial Diffusion, Radial Diffusion, Apparent Diffusion Coefficient, Fractional Anisotropy, T2). These datasets were pre-processed using skull stripping and background removal before feature extraction. The primary focus of this research was in the development of a Computer Vision and Classification methodology to understand the underlying Histopathology of MS Lesions contained within Multimodal MRI volumes.

Three initial experiments were carried out to achieve this aim. In particular, SVM, RF and ELM classifiers were trained on radiomic feature descriptors which were extracted using First Order Statistics, Grey Level Run Length statistics, Haralick texture features, Histogram of Oriented Gradients and Local Binary Patterns.

In the first experiment, LASSO and mRMR feature selection methods were applied to obtain a subset of features from the global feature set. These were then used to train Machine Learning models for Individual and Multimodal MRI classification. The second experiment evaluated the performance of feature groups through the creation of machine learning models trained on features from single extraction methods such as Haralick texture features. In doing so, the focus was to develop an understanding as to the classification performance of different features when applied, independently, to the task of MS Lesion detection in MRI volumes which have undergone a minimal pre-processing stage.

A significant portion of this research was based upon existing work in the field of Computer Vision and Machine Learning applications to Medical Imaging. During section 2, a Literature Review was conducted examining existing research in the application of similar techniques and to provide the foundation for the methodological underpinning of this work. The Methodology section 3 and Methods section 4 sections detailed the chosen approach and provided information on the chosen methods which used to complete this research. Finally, section 5 presented the results of both experiments carried out during this research.

The first section of the Discussion chapter presents a review of the Results section to understand whether the aim of this research outlined in section 1.5 has been accomplished. A critical evaluation in section 6.2 is carried out in reference to existing research in order to understand the validity of findings and success of this work. Finally, the conclusion section provides an overview of this research, including future research and recommendations in section 7.

6.1 Results Review

6.1.1 Training

During sample size calculation of models trained on Individual MRI sequences fig. 9, SVM and RF classifiers outperform ELM while maintaining a similar accuracy across both LASSO and mRMR feature selection methods. Extreme Learning machine shows significant instability when combined with mRMR feature selection compared to all other models, including the ELM classifier combined with LASSO feature selection. It is difficult to speculate upon the exact reasoning for this fluctuation in results when identifying a sample size owing to a lack of research in the applications mRMR and ELM applied to MS. However, examining existing work by (Zhu, Fan, He & Xu 2018) in which varying numbers of features and hidden neurons are applied, the sensitivity of ELM to parameter changes may be accountable for the instability of these results.

It should be noted that the sample size for Multimodal data, shown in fig. 19, was set at five multiplied by the selected Individual sample size of 4000, to account for all 5 MRI sequences. This sample size was selected owing to the steep increase in performance, followed by a reduction in classification accuracy afterwards, as shown in fig. 9.

LASSO tuning of Individual MRI sequence fig. 10 and Multimodal MRI sequence fig. 20 experiments found that a value of 70 provided the most consistent increase in accuracy in both Individual and Multimodal experiments. The ELM once more exhibits fluctuating accuracy, most noticeable in Multimodal experiments.

A coarse grid search algorithm was applied for model tuning to determine the optimal settings for each machine learning model. Individual LASSO fig. 11 and mRMR fig. 12 and Multimodal fig. 21 SVM tuning was carried out in line with (Hsu et al. 2003) who suggest an exponentially growing search space for C and γ performed within k-fold cross-validation. This research reinforces that finding, as evidenced by the convergence around a C value of 2 from all parameters. This lower value is desirable in research by (Roy et al. 2013) who advocate the setting of C to 1 for MS Lesion detection.

SVM converging around a γ value of 2 is significant in terms of the validity of this research. Despite the adopted minimal pre-processing stage, it is found that the ideal parameters for SVM are close to those proposed in similar research containing significant data pre-processing.

The final results from the training procedure, which were unexpected lie in the relationship between Individual and Multimodal training and both feature selection algorithms. Individual fig. 13 results show increased classification accuracy using mRMR feature selection across all five MRI sequences, while Multimodal fig. 22 results show LASSO with consistently higher accuracy. This trend is not related to sample size, which shows mostly equal accuracies between both feature selection methods. Instead, the classification of Multimodal data using LASSO feature selection and Random Forest may be considerably more successful after tuning the λ parameter compared to Individual classification post tuning.

From the training stage, the final results show that Random Forest achieved the highest accuracy in both LASSO and mRMR experiments with scores between 83-85%. Furthermore, of all models, the RF showed the least variance between MRI sequences and individual folds. Extreme Learning Machine scored the lowest with mean F1 scores below 60% despite good specificity, while SVM ranked between RF and ELM with scores between 73-75% as shown in table 4.

The ultimate purpose of the training section was to obtain a set of rankings from mRMR and LASSO algorithms which would be used in Individual and Multimodal experiments during model training and unseen dataset testing. As shown in fig. 27, the majority of features were selected from Local Binary Patterns in all cases with the remaining choices coming from First Order and Grey Level Run Length statistics. Haralick Texture Features were avoided entirely by LASSO and mRMR algorithms for every MRI sequence.

Speculation as to the exclusion of Haralick Texture Features is based on the spatial resolution of brain regions belonging to healthy and unhealthy classes. Owing to the lack of normalisation during pre-processing, information extracted from brain and Lesion regions using this method may not fit entirely into either class. As such, the feature selection algorithms will find no benefit in selecting them for inclusion in any feature vector and will instead select features which better represent a clear divide. A significant portion of research examined during the Literature Review chapter performed normalisation stages, such as (Tozer et al. 2009), (Loizou et al. 2015) and (Ghribi et al. 2018) who perform Haralick feature extraction after a rigorous pre-processing stage. In these experiments, the features perform well, indicating the necessity to normalise image volumes when using GLCM based features.

6.1.2 Experiment 1 – Feature selection Individual and Multimodal

Experiment 1 applied LASSO and mRMR feature selection methods to Individual and Multimodal MRI data for the creation of SVM, RF and ELM models trained on a subset of features taken from the global set. This research has found that both LASSO and mRMR feature selection methods are capable of identify a set of features which describe the underlying pathology of Multiple Sclerosis, as evidenced by the Individually trained SVM section 5.2.1, RF section 5.2.2, and ELM section 5.2.3, and Multimodal trained SVM section 5.3.1, RF section 5.3.2, and ELM section 5.3.3 results.

In the examination of the same MRI sequences there existed no correlation in classification performance between different patient datasets, and what instead is shown is scans within each patient dataset performing to a similar standard as all others for that patient. Furthermore, the trend across all results is that patient datasets containing considerably more Lesion volume are classified more accurately in all MRI sequences compared to those datasets containing small volume Lesions, as shown in table 2.

This is in line with the findings of (Saccà et al. 2018) who report their only misclassification was in a patient set consisting of the smallest Lesion load. Similar to their findings, the results of this research found that the only experiments to fail were those using SVM and ELM trained on

Individual MRI data attempting to classify p02c1, the second smallest patient within the available dataset. Unlike similar research which advocates the removal of smaller Lesions (Salem et al. 2018), the focus of this work was to identify all Lesions, especially those consisting of smaller volumes, owing to their importance in early disease detection (Fartaria et al. 2018).

The exact cause of misclassification is difficult to determine. Patient set p02c1 is the second smallest dataset based upon Lesion samples and percentage with 0.54% of the brain consisting of MS Lesion. Dataset p04c1 is the smallest, containing only 0.42% unhealthy samples within the brain volume.

The ELM achieved the lowest score of all machine learning models, although it should be noted that the variance of scores was significantly less than SVM and RF in LASSO fig. 32 and mRMR fig. 67b experiments. Research into the applicability of ELM to texture features describing MS Lesions is lacking, and this experiment has shown their potential resilience to outliers.

Visualising the success of each model is complex, owing to the sheer number of scans and potential Lesion examples. Because of this, the Individual section 5.2.4 and Multimodal section 5.3.4 segmentation images display a sample of the common trends which were present throughout all unseen patient datasets. It was found that all models successfully identified medium to large-sized Lesion regions but failed for much smaller volume Lesions, as shown in fig. 34 and fig. 69b.

Heatmap results of dataset p06c1 for LASSO fig. 36 and mRMR fig. 37 experiments show increased classification performance towards the upper section of Lesion regions, highlighting a potential increase in Lesion load towards this region. Classification has been successful throughout the entire Lesion region in the identification of slices containing unhealthy tissue samples. This could translate into clinical applications and operate as an indicator for medical experts to quickly understand where a Lesion volume may reside.

Looking at Multimodal results of p06c1 in fig. 47 and fig. 46, there is a slight trend in SVM and ELM performing to a higher standard then Individual SVM and ELM. Random Forest does not share this trend and instead Individual, and Multimodal results show similar accuracies. This trend only exists for this patient dataset, and from examining all other heatmap datasets, it is again clear that both LASSO and mRMR feature vectors are competitive in MS Lesion detection.

T2 classification using mRMR and LASSO feature selection techniques as shown in table 5, found that models trained on Multimodal MRI data outperformed Individual models in almost every case, with the exception of mRMR ELM. T2 is the only MRI sequence which exhibits this increase in performance. The importance of this finding is critical, owing to the primary use of T2 in clinical settings. When attempting to segment MS Lesion from a T2 scan, it is indicative with these results, that information from Multimodal MRI sequences is assisting in identifying Lesion regions more successfully than the same Machine Learning Models trained on Individual MRI data from T2 scans only.

The difference between slice accuracies and the overall sporadic nature of the results discussed in this section indicate significant differences amongst patients and MRI sequences. Deviation exists in accuracy between each patient dataset, as well as between the MRI sequences within these datasets. This does not mean that the results are poor or invalid, as evidenced by the successful identification of Multiple Sclerosis Lesions, it does, however, raise the question of the predictable performance and generalisation of future scans in relation to the proposed methodology when analysing new unseen MRI data.

In the evaluation of experiment one, it is clear that the proposed methodology is successful in understanding Multiple Sclerosis Lesion in medium to larger sized samples. A full analysis of the successes and limitations of this work is provided in section 6.2. However, the following section provides an interpretation of the results of experiment two, in which groups of features from each Feature Extraction method are used to segment MS Lesions.

6.1.3 Experiment 2 – Feature Groups Individual and Multimodal

Experiment 2 trained SVM, RF and ELM models on Individual and Multimodal MRI data taken from groups of features belonging to each of the implemented feature extraction techniques, outlined in section 4.4.5.

Examining the Dice score Boxplots for Individual fig. 48 and Multimodal fig. 83b results, the LBP descriptors performed poorly in comparison to all others with either complete failures or very low boxplot ranges. These findings suggest that LBP features should be avoided when attempting MS segmentation on unnormalized MRI data. This is contrary to the work of Abbasi & Tajeripour (2017) who suggest LBP features combined with HoG provides excellent description of targeted brain regions. It is noted that their work incorporated pre-processing for intensity normalisation and standardisation across MRI sequences before feature extraction.

In comparison, First Order and GLRL statistics classify to a lower standard than Haralick and HOG features which consistently achieve high Dice scores in Individual experiments. In Multimodal classification, the Haralick texture features achieve the highest upper quartile in all but one experiment (p06c1 First Order) indicating their ability to classify MS Lesion to a higher degree of accuracy in T2 data.

From the distribution graphs for Individual fig. 84b and Multimodal fig. 49, both experiments identified a high number of slices with accuracies between 50 and 69. Surprisingly, Individual SVM classified more slices to a score greater than 70% compared to Multimodal, which achieved most of these using Haralick texture features and First Order statistics. Furthermore, Haralick texture features consistently score more slices in the range of 50-60 compared to all other feature extraction techniques.

Examining the top slice values from feature groups experiments depicted in fig. 50, the maximum slice scores were classified within the largest patient dataset, p01c1, with HoG achieving 0.79% using Individual SVM, and Haralick achieving 0.77% using Multimodal SVM. It is noted that in

previous experiments using LASSO and mRMR that features from Haralick and HoG techniques were ignored entirely. This point is further discussed in section 6.2.

In comparison to SVM experiments, the RF classifiers Individual fig. 51 and Multimodal fig. 85b results, successfully identified MS Lesion within all datasets using all feature extraction techniques, including Local Binary Patterns. Experiments using feature groups cluster as patient datasets, however, unlike previous experiments it is possible to identify trends amongst feature extraction methods such as Haralick and HoG features consistently performing well, while First Order statistics and GLRL fluctuate in performance.

The number of top slices is considerably more in the case of Multimodal RF compared to Multimodal SVM, and consistency is seen between the two RF figures, highlighting the model's ability to classify unseen Individual and Multimodal data to a consistent standard across all feature extraction techniques. Furthermore, significantly more slices are classified in the range of 0.7 – 0.8%, as shown in fig. 53, and only two datasets fail to achieve top scores (p02c1 and p08c1). Interestingly, between each patient dataset and feature extraction method, the top slice score changes, indicating inconsistency in feature descriptors between patients when using the RF algorithm as SVM tended towards higher results using Haralick features.

In ELM boxplots for Individual fig. 54 and Multimodal fig. 87b feature group experiments, the results from Individual experiments group more consistently than Multimodal results, in which First Order and GLRL results are significantly lower than Haralick, HoG and LBP.

Examining the ELM distribution charts for Individual fig. 55 and Multimodal fig. 88b experiments, the number of top tier slices between Individual and Multimodal is considerably more when compared with SVM. Furthermore, Multimodal experiments appear to capture a greater number of slices with dice scores between 50-69 compared to Individual. ELM achieves the highest slice value of all experiments, using LBP features and Multimodal classification on patient p01c1, achieving a score of 0.825%, a huge contrast compared to SVM which failed to classify any multimodal LBP datasets.

Slice segmentations for individual feature groups show that the smallest Lesions are ignored, while medium to large sized areas are correctly identified. Figure fig. 96a highlights this, with HoG features being used in model training and unseen testing. Interestingly, the Multimodal model results depicted in fig. 96b highlights an attempt at smaller Lesion identification, once more demonstrating that Multimodal MRI data has the potential to describe smaller T2 regions which may not have been captured during 5x5 region window feature extraction.

While not achieving the highest accuracy, fig. 95a shows the results of experiments using Haralick texture features on T2, resulting in a good segmentation despite not achieving the highest accuracy. The Multimodal case in fig. 95b shows clear region identification of almost all Lesion areas across each machine learning model.

LBP features failed for SVM, and it may be the case that unnormalized MRI LBP features are heavily dependent on the choice of Machine Learning model as both RF and ELM successfully identified MS Lesions using these features. Contrary to speculation, the ELM classifies to a higher standard than the RF classifier in these examples.

SVM consistently achieved the lowest classification accuracy of MS Lesion, a finding in line with the research of (Speiser et al. 2019). The results of this work also reinforce the findings of (Sharma et al. 2016) who found that low amounts of training data, as is common in MS Lesion segmentation tasks, and larger feature vectors can significantly impact the classification performance of SVM.

Heatmap graphs for each feature selection method are shown in section 5.4.5 and each set of figures displays Individual and Multimodal results of every feature extraction method. With the exception of Local Binary Patterns, each experiment is successful in the two patients shown. In line with the findings of experiment one, all feature groups classify to a higher degree in regions of greater Lesion magnitude, as depicted in fig. 103.

Across all heatmaps, the results show increased slice accuracy for regions containing more Lesion samples. For patient p06c1, this is found to be towards the upper region of the Lesion volume. It was also found that throughout all experiments, there exists no outright winner between Individual and Multimodal experiments. Looking at the score table table 6, the highest machine learning model trained on Individual T2 data was RF with an average F1 score of 44.26±8.38 when trained with HoG features only. The highest model trained on Multimodal data and used to classify T2 was again RF and HoG features with a score of 48.05±8.3. The lowest score ignoring LBP features was Individual First Order experiments using SVM which achieved a volume accuracy of 30.19±15.53.

An unexpected and critical finding was the ELM classifier achieving two of the highest classification scores across the entire set of leave-out patients in Haralick texture features and Local Binary Patterns, with respective scores of 48.02±7.76 and 47.27±6.29. In both Individual and Multimodal experiments, the ELM managed to attain one of the highest volume accuracies where RF and SVM misclassified entirely in the Multimodal classification of LBP features.

The specificity of all experiments using individual feature groups did not fall below 0.99%, similar to experiment one using feature selection algorithms, and it was found that the feature descriptors themselves have proven successful in combination with the chosen machine learning algorithms to MS Lesion segmentation. There are, however, certain considerations which must be taken into account when evaluating the success of this research.

6.2 Critical Evaluation

A considerable number of Research Methodologies in MS Lesion Segmentation, as well as wider Medical Imaging applications, advocate varying pre-processing stages in order to improve the classification performance of proposed pipelines. The incorporation of unique and contradicting pre-processing steps results in less generative methods and an overall lack of standardisation in

clinical application and adoption of developed Medical Imaging systems. A critical step within this research was the negation of pre-processing involving intensity normalisation and noise removal in order to treat all MRI volumes as naturally as possible, in line with a clinical setting in which a medical expert would make a diagnosis.

The results of this research contradict the findings of (Sweeney et al. 2014) who found that the selection of a classification algorithm is not critical in comparison to the development of a set of meaningful feature descriptors for MS Lesion descriptions. The findings suggest that careful consideration should be given to Feature Descriptors, Feature Selection and Classification model choice when designing an MS Lesion segmentation methodology. In both experiments, it was found that Machine Learning model selection played a significant role in Classification Accuracy when examining unnormalized MRI data.

This is evidenced in the results of the comparative analysis between Feature Extraction techniques in table 5, where it was found that significant performance differences existed between the same Machine Learning models operating on different sets of features, as well as between each Machine Learning model operating on the same sets of Features. Furthermore, the choice of Feature Selection method also adds to the complexity of classification success as evidenced in table 5, where the identified set of 25 features has a significant impact on a Machine Learning Models ability to identify MS Lesion.

In all experiments carried out within this research, the specificity did not fall below 0.9983, highlighting the resilience of this methodology to False Positive classification through meaningful post-processing. All classification algorithms identified false Lesion regions which required morphological post-processing, carried out in line with the suggestions of (Roy et al. 2013). This research is not susceptible to the common problem of MS segmentation methodologies suffering from high False Positive rates owing to the similarities between Lesion and normal tissue and inability to meaningfully separate the two classes.

While the specificity of this work is high, the Sensitivity to MS Lesions falls between 0.58% and 0.67% across all experiments highlighting the necessity for an improved description of unhealthy tissue regions. A significant portion of reviewed literature proposes methodologies ranging in accuracy from 0.40% to 100% and a direct comparison is difficult given the huge variety of segmentation methodologies and limited available datasets for which to test the generalisation of developed systems (Danelakis et al. 2018).

Where this research succeeds is in the application of Multimodal Machine Learning model training using biomarkers extracted from multiple MRI sequences, uncommonly applied to MS Lesion identification. This is in line with the findings of (Davoudi et al. 2016) who also state that Multimodal biomarkers provide more information than Unimodal. The findings of this Thesis advocate consistently better classification performance in all but one experiment where Multimodal classification algorithms outperform the same algorithms trained on unimodal MRI data. From table 5 and table 6, it is apparent that when classifying unseen T2 datasets, information from all 5 MRI

sequences, combined for Multimodal experiments, provides more meaningful description than individual experiments using T2 data only.

A significant discovery from this research was the classification performance and consistency of the Extreme Learning Machine classifier, which outperformed the RF algorithm in several instances. Furthermore, the ELM correctly identified MS Lesion in Multimodal LBP feature group experiments where SVM and RF failed, as well as achieving the highest LBP score in the Individual MRI sequence version of this experiment. To the best of our knowledge, there is no existing literature on the application of ELM for texture classification of MS Lesions and these findings represent a significant possibility for the application of ELM to textural biomarker classification.

While each machine learning model consistently outperformed at least one other, the SVM failed to yield a winning score in any experiment. In line with existing research, it is speculated that the imbalanced nature of MS Lesion datasets and lack of training samples is disadvantageous to SVM classifiers. (Saccà et al. 2018) provide possible reasoning as to the reduced performance in SVM compared to RF, stating that their results of SVM and RF performance being comparable is most likely attributed to their pre-processing stage for denoising and artefact removal. As this research carries out no such stage, it is speculated that the gap in performance is attributed to SVM's disadvantage when presented with larger noisy datasets.

LASSO and mRMR feature selection algorithms were applied during the first set of Individual and Multimodal experiments for dimensionality reduction from 302 descriptors into a feature vector of 25 elements. The averaged results of all unseen patient datasets are shown in table 5. It was found that between both Feature Selection results, there existed no clear indication of the best method, and when analysing the results of feature selection experiments against the second feature groups experiments, table 6, in certain cases, the use of a single set of feature descriptors outperformed LASSO and mRMR's selection from the global feature set.

This finding is significant in terms of the hypothesised outcome from this research which followed the findings of (Yoo et al. 2018) who found the application of LASSO Feature Selection alongside RF classification resulted in classification accuracy increase. Our natural assumption was that both LASSO and mRMR algorithms would correctly identify a subset of descriptors which were capable of identifying MS Lesion from healthy brain tissue. From the final rankings, fig. 27, and results of both experiments, it is clear that the identified rankings from Individual MRI sequences for concatenation and Feature selection in Multimodal experiments yielded a set of descriptors, capable of describing MS Lesion, but unable to improve the performance over manual feature grouping.

In the available dataset, the highest resolution image was the clinical T2w, while the more experimental DWI techniques of, AD, ADC, FA and RD scans comprised significantly less information. The highest performing Feature Extraction technique using Multimodal MRI data was Haralick texture features captured using an averaged GLCM, as shown in table 6. These features were also the most consistent across the three machine learning models. A result which reinforces the

findings of (Mayerhoefer et al. 2009) who suggest that co-occurrence statistics are significantly more accurate than other texture descriptors when analysing groups of scans comprising varying resolutions.

Significant research has found that texture descriptors such as Haralick's Texture Features and Galloway's GLRL statistics have been widely applied to Multiple Sclerosis Lesion Segmentation tasks (Zhang et al. 2008, Michoux et al. 2015, Ghribi et al. 2018). While the application of LASSO and mRMR feature selection algorithms failed to identify any GLCM features, the second set of experiments comparing these methods did highlight their robustness to MRI sequences and different patient datasets.

6.2.1 Limitations

There are several limitations to this research. Firstly, the identified rankings from LASSO and mRMR algorithms across all 5 MRI sequences shown in table 6 failed to select any of the Haralick texture features which have proven successful in existing research in MS Lesion detection using T2w images (Ghribi et al. 2018, Samah et al. 2018). LASSO and mRMR also failed to select any HoG features which have been found to prove useful when combined with Local Binary Patterns and Random Forest classification for Brain Tumour detection (Abbasi & Tajeripour 2017), as well as showing good classification performance in experiment two results.

A critical issue with the experimental setup of this work involves the combination of identified features from Individual Machine Learning Training, for creation of a Multimodal feature subset for use in Multimodal training and feature selection. In doing so, a significant portion of MRI features are excluded during the second feature selection step where they may prove useful as Multimodal biomarkers.

While the results of the first experiment using Feature selection algorithms fell below hypothesised expectation, it is critical to note the success of the research owing to the overall classification performance and ability to segment MS Lesion regions. Furthermore, it is noted that the LASSO training procedure is correctly implemented owing to the similarities between the selections of both itself and mRMR, an algorithm which does not require training. It is therefore apparent that any failure is in the application of these algorithms to biomarkers extracted from unnormalized MRI data.

The reduced performance in comparison to existing state of the art research in MS Lesion segmentation is attributed to the lack of pre-processing and intensity normalisation of MRI volumes before feature extraction. Existing literature by (Samah et al. 2018) investigated the level of noise within MRI volumes and found GLCM and LBP features inherently unstable at higher levels of noise, a fundamental weakness for this research in which a primary focus was to leave brain regions untouched. This assumption is in line with research by (Brynolfsson et al. 2017), who found that the use of Haralick Texture features required standardisation in resolution and greyscale for meaningful description.

Another area of this research which has affected overall performance is the number of smaller sized Lesions contained within patient volumes. Smaller Lesions are not fully captured when using a 5x5 region window owing to the combination of boundaries between healthy Brain and unhealthy Lesion tissue regions. (García-Lorenzo et al. 2013) highlight the importance of taking into consideration the differing Lesion loads contained within a patient dataset for more accurate results analysis. During the analysis of unseen classification experiments, these smaller Lesions are misclassified owing to the blending of regions and unclear groupings between healthy and unhealthy classes.

The problem of Lesion load is negated entirely by (Salem et al. 2018) who remove small Lesions of less than three voxels from datasets before feature extraction in order to improve the classification performance of their methodology. This approach, however, is not desirable owing to the missed opportunity to assist in the diagnosis of early-stage Multiple Sclerosis and disease progression (Fartaria et al. 2018).

7 Conclusion

The focus of the research was to determine whether qualitative Multimodal feature descriptors extracted from a database of Multiple Sclerosis patients could correctly categorise Multiple Sclerosis Lesions from healthy brain tissue regions, irrespective of the MRI sequences selected for classification.

Two experiments were implemented, both of which have been outlined in section 4.4. The first of these applied LASSO and mRMR feature selection to Individual and Multimodal MRI data for the creation of two sets of Machine Learning models focused on Individual MRI sequence analysis and Multimodal sequence analysis. Multimodal analysis combined data from all MRI sequences in order to ascertain whether information from non-clinical sequences could be combined alongside standard T2 data in order to classify Multiple Sclerosis Lesions in unseen patient datasets.

The second set of experiments took groups of biomarkers from different feature extraction techniques and compared the performance of each of the methods outlined in section 4 in order to understand which feature extraction technique best describes MS Lesions found in unnormalized MRI volumes. Both of these experiments created Individual and Multimodal Machine Learning models for classification of unseen patient datasets.

The results indicate that biomarkers generated from non-clinical MRI sequences such as AD, ADC, FA and RD provide additional description alongside information extracted from T2 volumes, for use in Lesion identification and segmentation. Owing to the lack of pre-processing for standardisation and resolution differences between MRI volumes, this finding suggests that combinations of texture features extracted from these volumes can contain information useful across a broad range of scanning sequences, including clinical T2 images.

While the application of feature selection algorithms was successful, the performance increase was not as impactful as initially predicted and a number of useful descriptors, such as Haralick's

texture features and HoG descriptors, were left out of feature vectors generated by the two methods. Because of this, the initial experiment using feature selection algorithms played less of a role in answering the research question, and instead a more exhaustive search of the feature space would have been more appropriately suited.

Where this research exceeds expectations is in the classification performance of the Extreme Learning Machine, which has yet to be applied to the problem of texture classification of Multimodal MRI volumes. This work presents ELM's ability to compete with established RF and SVM classifiers, as well as outperforming them both in MS Lesion detection during a range of experiments.

This work also presents the ability to understand slice classifications in the context of the structural locations of MS Lesions in the coronal plane of an MRI image through Heatmap graphing. This combination of Heatmaps, Lesion structure and Classification accuracy goes a long way in providing a visually detailed and quick understanding of the algorithmic performance of a methodology.

7.1 Recommendations

To better understand the implications of these results, future studies could focus upon the analysis of pre-processing techniques as opposed to Machine Learning and Feature Extraction methodologies. The lack of adoption of developed Computer Systems in the Medical Community relates to poor generalisation between closed laboratory and clinical settings where a push for top results in academia can hinder the real-world impact of a developed system. Furthermore, the availability of datasets representing Multiple Sclerosis is considerably lacking, and a greater pool of labelled data is necessary across all Medical imaging systems, especially for diseases containing smaller samples outside of the domain of MS.

Depending upon the desired outcome from future research, the removal of smaller Lesions could significantly improve the classification success of a study. Within this research, the aim was in the evaluation of uncommonly applied MRI sequences to MS Lesion identification, as such, smaller Lesions were left untouched in order to understand whether the combination of all MRI sequences could identify biomarkers describing these critical areas.

The findings of this research indicate reduced performance improvement in the use of feature selection algorithms when operating on unnormalized MRI volumes. It is advised that a more robust feature selection stage is implemented, such as Sequential Forward Selection in which the best feature is initially identified, followed by the second until n features have been obtained. More exhaustive approaches such as these offer a detailed and more supervised search of a feature space compared to automatic and semi-automatic methods, at the cost of computational expense.

7.2 Future work

Following the outcome of this research, the transferability and generalisation of the proposed methodology must be tested against open-source datasets. This will provide a greater insight into exact future directions and changes which must be made in order to assist in the clinical

adoption of texture-based classification systems. Inline with state-of-the-art research, the sourced dataset could be used within a CNN methodology to establish how a fully automatic architecture understands the histopathology of MS Lesion represented in Multimodal MRI data, however, the number of available datasets once more limits the success of this direction.

During the early stages of this research, a full set of Wavelet transformed images were captured. These images were not used owing to the complexity of the proposed methodology and time taken to carry out feature extraction. These images have the potential to describe MS Lesions across all MRI sequences in a unique manner and could present a new set of biomarkers for use independently or alongside the existing database of features.

This research has successfully applied established textural descriptors alongside state-of-the-art classification and feature selection algorithms for the classification and segmentation of Multiple Sclerosis Lesion in Multimodal MRI data. The results highlight the successful application of the ELM classifier, which has yet to be applied to MS Lesion in this way, as well as challenging existing findings regarding the RF and SVM classifier and their reliability.

This work has sought to address a knowledge gap in MS Lesion detection in the classification of uncommonly applied and unnormalized MRI volumes of Axial Diffusion, Radial Diffusion, Apparent Diffusion Coefficient, Fractional Anisotropy, alongside the clinical T2w volume. It has been found that these MRI sequences provide combined information which assists in the diagnosis of MS Lesion when analysing an unseen volume from one of these sequences.

It is hoped that this research provides a useful guide into unnormalized MRI analysis utilising existing techniques in Computer Vision and Machine Learning, as well as operating as a platform for future development in MS Lesion detection for diagnosis, monitoring and prognosis.

References

Abbasi, S. & Tajeripour, F. (2017), 'Detection of brain tumor in 3d mri images using local binary patterns and histogram orientation gradient', *Neurocomputing* **219**, 526–535. Available from https://www.sciencedirect.com/science/article/pii/S0925231216310864 [accessed 15 February 2019].

Abdullah, B. A., Younis, A. A. & John, N. M. (2012), 'Multi-sectional views textural based svm for ms lesion segmentation in multi-channels mris', *The open biomedical engineering journal* **6**, 56. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3382289/ [accessed 08 May 2019].

Akbarpour, T., Shamsi, M., Daneshvar, S. & Pooreisa, M. (2017), 'Unsupervised multimodal magnetic resonance images segmentation and multiple sclerosis lesions extraction based on edge and texture features', *Applied Medical Informatics* **39**(1-2), 30–40. Available from https://ami.info.umfcluj.ro/index.php/AMI/article/view/619 [accessed 19 March 2019].

Ali, S. & Maher, A. (2016), Identifying multiple sclerosis lesions in mr images using image processing techniques, *in* '2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)', IEEE, pp. 1–5. Available from https://ieeexplore.ieee.org/abstract/document/7759913 [accessed 19 October 2017].

Alroughani, R., Akhtar, S., Ahmed, S., Behbehani, R. & Al-Hashel, J. (2016), 'Is time to reach edss 6.0 faster in patients with late-onset versus young-onset multiple sclerosis?', *PLOS ONE* **11**(11), 1–10. Available from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0165846 [accessed 15 December 2018].

Amin, J., Sharif, M., Raza, M. & Yasmin, M. (2018), 'Detection of brain tumor based on features fusion and machine learning', *Journal of Ambient Intelligence and Humanized Computing* . Available from https://link.springer.com/article/10.1007/s12652-018-1092-9 [accessed 08 May 2019].

Andronache, A., Rosazza, C., Sattin, D. P., Leonardi, M., D'Incerti, L. & Minati, L. (2013), 'Impact of functional mri data preprocessing pipeline on default-mode network detectability in patients with disorders of consciousness', *Frontiers in neuroinformatics* **7**, 16. Available from https://www.frontiersin.org/articles/10.3389/fninf.2013.00016/full [accessed 03 May 2019].

Arunkumar, N., Mohammed, M. A., Mostafa, S. A., Ibrahim, D. A., Rodrigues, J. J. & de Albuquerque, V. H. C. (2018), 'Fully automatic model-based segmentation and classification approach for mri brain tumor using artificial neural networks', *Concurrency and Computation: Practice and Experience* p. e4962. Available from https://onlinelibrary.wiley.com/doi/full/10.1002/cpe.4962 [accessed 06 May 2019].

Aslani, S., Dayan, M., Storelli, L., Filippi, M., Murino, V., Rocca, M. A. & Sona, D. (2019), 'Multi-branch convolutional neural network for multiple sclerosis lesion segmentation', *NeuroImage* **196**, 1–15. Available from https://www.sciencedirect.com/science/article/pii/S105381191930268X [accessed 22 April 2019].

Bell, Daniel, J. & Jones, J. (2019), 'Mri'. Available from https://radiopaedia.org/articles/mri-2?lang=gb [accessed 12 May 2019].

Bercovich, E. & Javitt, M. C. (2018), 'Medical imaging: From roentgen to the digital revolution, and beyond', *Rambam Maimonides medical journal* **9**(4). Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6186003/ [accessed 15 December 2018].

Bharathi, P. & Subashini, P. (2014), 'Optimal feature subset selection using differential evolution and extreme learning machine', *Int. J. Sci. Res* **3**(7), 1898–1905. Available from http://www.academia.edu/download/37003864/IJSR_Published_Paper.pdf [accessed 06 May 2019].

Boser, B. E., Guyon, I. M. & Vapnik, V. N. (1992), A training algorithm for optimal margin classifiers, *in* 'Proceedings of the fifth annual workshop on Computational learning theory', ACM, pp. 144–152. Available from https://dl.acm.org/citation.cfm?id=130401 [accessed 08 May 2019].

Breiman, L. (2001), 'Random forests', *Machine Learning* **45**(1), 5–32. Available from https://link.springer.com/article/10.1023/A:1010933404324 [accessed 08 May 2019].

Brynolfsson, P., Nilsson, D., Torheim, T., Asklund, T., Karlsson, C. T., Trygg, J., Nyholm, T. & Garpebring, A. (2017), 'Haralick texture features from apparent diffusion coefficient (adc) mri images depend on imaging and pre-processing parameters', *Scientific reports* **7**(1), 4041. Available from https://www.nature.com/articles/s41598-017-04151-4 [accessed 08 January 2019].

Cai, J., Luo, J., Wang, S. & Yang, S. (2018), 'Feature selection in machine learning: A new perspective', *Neurocomputing* **300**, 70 – 79. Available from https://www.sciencedirect.com/science/article/pii/S0925231218302911 [accessed 07 May 2019].

Carrigan, N., Dysch, L. & Salkovskis, P. M. (2018), 'The impact of health anxiety in multiple sclerosis: a replication and treatment case series', *Behavioural and cognitive psychotherapy* **46**(2), 148–167. Available from https://www.ncbi.nlm.nih.gov/pubmed/28988546 [accessed 12 May 2019].

Chawla, S., Kister, I., Sinnecker, T., Wuerfel, J., Brisset, J.-C., Paul, F. & Ge, Y. (2018), 'Longitudinal study of multiple sclerosis lesions using ultra-high field (7t) multiparametric mr imaging', *PloS one* **13**(9), e0202918. Available from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0202918 [accessed 30 April 2019].

Cortes, C. & Vapnik, V. (1995), 'Support-vector networks', *Machine Learning* **20**(3), 273–297. Available from https://link.springer.com/article/10.1023/A:1022627411411 [accessed 08 May 2019].

Dalal, N. & Triggs, B. (2005), Histograms of oriented gradients for human detection, *in* '2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)', Vol. 1, pp. 886–893 vol. 1. Available from https://ieeexplore.ieee.org/document/1467360 [accessed 30 April 2019].

Damadian, R. (1971), 'Tumor detection by nuclear magnetic resonance', *Science* **171**(3976), 1151–1153. Available from http://science.sciencemag.org/content/171/3976/1151.short [accessed 12 May 2019].

Damadian, R., Goldsmith, M. & Minkoff, L. (1977), 'Nmr in cancer: Xvi. fonar image of the uve human body'. Available from http://www.uprightmriimaging.com/pdf/doc_17.pdf [accessed 12 May 2019].

Danelakis, A., Theoharis, T. & Verganelakis, D. A. (2018), 'Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging', *Computerized Medical Imaging and Graphics* **70**, 83–100. Available from https://www.sciencedirect.com/science/article/abs/pii/S0895611183032274 [accessed 22 July 2019].

Davoudi, Y., Foroughipour, M., Torabi, R., Layegh, P., Matin, N. & Shoeibi, A. (2016), 'Diffusion weighted imaging in acute attacks of multiple sclerosis', *Iranian Journal of Radiology* **13**(2). Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5035938/ [accessed 13 May 2019].

Dworkin, J. D., Linn, K. A., Oguz, I., Fleishman, G. M., Bakshi, R., Nair, G., Calabresi, P., Henry, R., Oh, J., Papinutto, N. et al. (2018), 'An automated statistical technique for counting distinct multiple sclerosis lesions', *American Journal of Neuroradiology* **39**(4), 626–633. Available from http://www.ajnr.org/content/39/4/626.abstract [accessed 03 May 2019].

Fartaria, M. J., Todea, A., Kober, T., O'brien, K., Krueger, G., Meuli, R., Granziera, C., Roche, A. & Cuadra, M. B. (2018), 'Partial volume-aware assessment of multiple sclerosis lesions', *NeuroImage: Clinical* **18**, 245–253. Available from https://www.ncbi.nlm.nih.gov/pubmed/29868448 [accessed 30 April 2019].

Filippi, M., Brück, W., Chard, D., Fazekas, F., Geurts, J. J. G., Enzinger, C., Hametner, S., Kuhlmann, T., Preziosa, P., Àlex Rovira, Schmierer, K., Stadelmann, C. & Rocca, M. A. (2019), 'Association between pathological and mri findings in multiple sclerosis', *The Lancet Neurology* **18**(2), 198 – 210. Available from http://www.sciencedirect.com/science/article/pii/S1474442218304514 [accessed 12 May 2019].

Fonti, V. & Belitser, E. (2017), 'Feature selection using lasso', *VU Amsterdam Research Paper in Business Analytics* . Available from https://www.researchgate.net/profile/David_Booth14/post/How_to_model_high_dimensional_data_pn/attachment/5b6bfda43843b04aed7940a4/AS%3A657635544141832%401533803940729/download/werkstuk-fonti_tcm235-836234.pdf [accessed 07 May 2019].

Freeman, W. T. & Roth, M. (1995), Orientation histograms for hand gesture recognition, *in* 'International workshop on automatic face and gesture recognition', Vol. 12, pp. 296–301. Available from http://aimm02.cse.ttu.edu.tw/class_2009_2/CV/OpenCV/References/Orientation%20histograms%20for%20hand%20gesture.pdf [accessed 30 April 2019].

Galassi, F., Commowick, O., Vallee, E. & Barillot, C. (2019), Voxel-wise comparison with a-contrario analysis for automated segmentation of multiple sclerosis lesions from multimodal mri, *in* A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes & T. van Walsum, eds, 'Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries', Springer International Publishing, Cham, pp. 180–188. Available from https://link.springer.com/chapter/10.1007/978-3-030-11723-8_18#citeas [accessed 03 May 2019].

Galloway, M. M. (1975), 'Texture analysis using gray level run lengths', *Computer Graphics and Image Processing* **4**(2), 172 – 179. Available from http://www.sciencedirect.com/science/article/pii/S0146664X75800086 [accessed 24 November 2017].

García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L. & Collins, D. L. (2013), 'Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging', *Medical image analysis* **17**(1), 1–18. Available from https://www.sciencedirect.com/science/article/abs/pii/S1361841512001338 [accessed 28 July 2019].

Ghribi, O., Sellami, L., Slima, M. B., Mhiri, C., Dammak, M. & Hamida, A. B. (2018), 'Multiple sclerosis exploration based on automatic mri modalities segmentation approach with advanced volumetric evaluations for essential feature extraction', *Biomedical Signal Processing and Control* **40**, 473–487. Available from https://www.sciencedirect.com/science/article/abs/pii/S1746809417301404 [accessed 15 February 2019].

Giacalone, M., Rasti, P., Debs, N., Frindel, C., Cho, T.-H., Grenier, E. & Rousseau, D. (2018), 'Local spatio-temporal encoding of raw perfusion mri for the prediction of final lesion in stroke', *Medical Image Analysis* **50**, 117–126. Available from http://www.sciencedirect.com/science/article/pii/S1361841518306807 [accessed 06 May 2019].

Gillies, R. J., Kinahan, P. E. & Hricak, H. (2015), 'Radiomics: images are more than pictures,

they are data', *Radiology* **278**(2), 563–577. Available from https://link.springer.com/article/10.1007/s00415-017-8683-9 [accessed 14 May 2019].

Glen, S. (2016), 'Jaccard index / similarity coefficient'. Available from https://www.statisticshowto.datasciencecentral.com/jaccard-index/ [accessed 15 May 2019].

Goldenberg, M. M. (2012), 'Multiple sclerosis review', *Pharmacy and Therapeutics* **37**(3), 175. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351877/ [accessed 23 December 2018].

Gonzalez-Arias, C., Viafara, C., Coronado, J. & Martinez, F. (2019), 'Automatic classification of severe and mild wear in worn surface images using histograms of oriented gradients as descriptor', *Wear* **426-427**, 1702 – 1711. Available from https://www.sciencedirect.com/science/article/pii/S0043164818315023 [accessed 06 May 2019].

Gonzalez, R. C. & Woods, R. E. (2008), *Digital Image Processing, 3rd edition*, Uttar Pradesh: Pearson India Education Services.

Gumaei, A., Hassan, M. M., Hassan, M. R., Alelaiwi, A. & Fortino, G. (2019), 'A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification', *IEEE Access* **7**, 36266–36273. Available from https://ieeexplore.ieee.org/document/8664160 [accessed 10 May 2019].

Gupta, N., Bhatele, P. & Khanna, P. (2018), 'Identification of gliomas from brain mri through adaptive segmentation and run length of centralized patterns', *Journal of Computational Science* **25**, 213–220. Available from https://www.sciencedirect.com/science/article/pii/S1877750317302077 [accessed 06 May 2019].

Hacking, C. & Bashir, U. (2019), 'T2 relaxation'. Available from https://radiopaedia.org/articles/t2-relaxation?lang=gb [accessed 15 September 2019].

Hancer, E., Xue, B. & Zhang, M. (2018), 'Differential evolution for filter feature selection based on information theory and feature ranking', *Knowledge-Based Systems* **140**, 103 – 119. Available from https://www.sciencedirect.com/science/article/pii/S0950705117304987 [accessed 07 May 2019].

Hanchuan Peng, Fuhui Long & Ding, C. (2005), 'Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(8), 1226–1238. Available from https://ieeexplore.ieee.org/document/1453511 [accessed 07 May 2019].

Haq, A. U., Li, J. P., Memon, M. H., Nazir, S. & Sun, R. (2018), 'A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms', *Mobile*

Information Systems **2018**. Available from https://www.hindawi.com/journals/misy/2018/3860146/abs/ [accessed 07 May 2019].

Haralick, R. M., Shanmugam, K. & Dinstein, I. (1973), 'Textural features for image classification', *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3**(6), 610–621. Available from https://ieeexplore.ieee.org/document/4309314 [accessed 05 May 2019].

Hartung, H.-P., Graf, J., Aktas, O., Mares, J. & Barnett, M. H. (2019), 'Diagnosis of multiple sclerosis: revisions of the mcdonald criteria 2017–continuity and change', *Current opinion in neurology* **32**(3), 327–337. Available from https://www.ncbi.nlm.nih.gov/pubmed/30985371 [accessed 12 May 2019].

Hastie, T., Tibshirani, R. & Friedman, J. (2009), 'The elements of statistical learning: data mining, inference, and prediction, springer series in statistics'. Page 658.

Hsu, C.-W., Chang, C.-C., Lin, C.-J. et al. (2003), 'A practical guide to support vector classification'. [accessed 10 July 2019].

Huang, G.-B., Zhu, Q.-Y. & Siew, C.-K. (2006), 'Extreme learning machine: Theory and applications', *Neurocomputing* **70**(1), 489 – 501. Available from https://www.sciencedirect.com/science/article/pii/S0925231206000385 [accessed 08 May 2019].

Huang, G., Zhou, H., Ding, X. & Zhang, R. (2012), 'Extreme learning machine for regression and multiclass classification', *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **42**(2), 513–529. Available from https://ieeexplore.ieee.org/abstract/document/6035797 [accessed 08 May 2019].

Isoglu, S., Koca, E. I. & Duru, D. G. (2017), Comparative multiple sclerosis lesion segmentation in magnetic resonance images, *in* '2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)', IEEE, pp. 1–4. Available from https://ieeexplore.ieee.org/abstract/document/7956784 [accessed 14 October 2017].

Jadhav, S., He, H. & Jenkins, K. (2018), 'Information gain directed genetic algorithm wrapper feature selection for credit rating', *Applied Soft Computing* **69**, 541 – 553. Available from https://www.sciencedirect.com/science/article/pii/S1568494618302242 [accessed 07 May 2019].

Jafarpour, S., Sedghi, Z. & Amirani, M. C. (2012), 'A robust brain mri classification with glcm features', *International Journal of Computer Applications* **37**(12), 1–5. Available from https://pdfs.semanticscholar.org/19c7/f566a4e32e7969a223ae268f6a2726ce4b46.pdf [accessed 08 May 2019].

Jalal, A., Quaid, M. A. K. & Sidduqi, M. A. (2019), A triaxial acceleration-based human motion detection for ambient smart home system, *in* '2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)', pp. 353–358. Available from https://ieeexplore.ieee.org/abstract/document/8667183 [accessed 08 May 2019].

Jiang, L., Kong, G. & Li, C. (2019), 'Wrapper framework for test-cost-sensitive feature selection', *IEEE Transactions on Systems, Man, and Cybernetics: Systems* pp. 1–10. Available from https://ieeexplore.ieee.org/abstract/document/8674538 [accessed 07 May 2019].

Kalina, J. & Schlenker, A. (2015), 'A robust supervised variable selection for noisy high-dimensional data', *BioMed research international* **2015**. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4468284/ [accessed 07 May 2019].

Kang, C., Huo, Y., Xin, L., Tian, B. & Yu, B. (2019), 'Feature selection and tumor classification for microarray data using relaxed lasso and generalized multi-class support vector machine', *Journal of Theoretical Biology* **463**, 77 – 91. Available from http://www.sciencedirect.com/science/article/pii/S0022519318306003 [accessed 07 May 2019].

Karimaghaloo, Z., Rivaz, H., Arnold, D. L., Collins, D. L. & Arbel, T. (2015), 'Temporal hierarchical adaptive texture crf for automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain mri', *IEEE Transactions on Medical Imaging* **34**(6), 1227–1241. Available from https://ieeexplore.ieee.org/abstract/document/6987348 [accessed 07 May 2019].

Karimaghaloo, Z., Shah, M., Francis, S. J., Arnold, D. L., Collins, D. L. & Arbel, T. (2012), 'Automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain mri using conditional random fields', *IEEE Transactions on Medical Imaging* **31**(6), 1181–1194. Available from https://ieeexplore.ieee.org/abstract/document/6145687 [accessed 17 April 2019].

Kim, Y., Cho, H.-h., Kim, S. T., Park, H., Nam, D. & Kong, D.-S. (2018), 'Radiomics features to distinguish glioblastoma from primary central nervous system lymphoma on multi-parametric mri', *Neuroradiology* **60**(12), 1297–1305. Available from https://link.springer.com/article/10.1007/s00234-018-2091-4 [accessed 07 May 2019].

Köhler, C., Wahl, H., Ziemssen, T., Linn, J. & Kitzler, H. H. (2019), 'Exploring individual multiple sclerosis lesion volume change over time: development of an algorithm for the analyses of longitudinal quantitative mri measures', *NeuroImage: Clinical* **21**, 101623. Available from https://www.sciencedirect.com/science/article/pii/S2213158218303711 [accessed 03 May 2019].

Kuhlmann, T., Lingfeld, G., Bitsch, A., Schuchardt, J. & Brück, W. (2002), 'Acute axonal damage in multiple sclerosis is most extensive in early disease stages and decreases over time', *Brain* **125**(10), 2202–2212. Available from https://academic.oup.com/brain/article/125/10/2202/300498 [accessed 23 December 2018].

Kwak, G.-H. & Park, N.-W. (2019), 'Impact of texture information on crop classification with machine learning and uav images', *Applied Sciences* **9**(4), 643. Available from https://www.mdpi.com/2076-3417/9/4/643 [accessed 16 April 2019].

Laing, C. M., Phillips, L. H., Cooper, C. L., Hosie, J. A. & Summers, F. (2015), 'Anger, quality of life and mood in multiple sclerosis', *Journal of Multiple Sclerosis* . Available from http://aura.abdn.ac.uk/bitstream/handle/2164/9399/anger_quality_of_life_and_mood_in_multiple_sclerosis_2376_0389.1000127.pdf?sequence=1 [accessed 23 December 2018].

Lalkhen, A. G. & McCluskey, A. (2008), 'Clinical tests: sensitivity and specificity', *Continuing Education in Anaesthesia Critical Care & Pain* **8**(6), 221–223. Available from https://academic.oup.com/bjaed/article/8/6/221/406440 [accessed 14 May 2019].

Lam, C., Yu, C., Huang, L. & Rubin, D. (2018), 'Retinal lesion detection with deep learning using image patches', *Investigative ophthalmology & visual science* **59**(1), 590–596. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5788045/ [accessed 03 May 2019].

Lassmann, H., Brück, W. & Lucchinetti, C. F. (2007), 'The immunopathology of multiple sclerosis: an overview', *Brain pathology* **17**(2), 210–218. Available from https://onlinelibrary.wiley.com/doi/full/10.1111/j.1750-3639.2007.00064.x [accessed 12 May 2019].

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J. & Liu, H. (2017), 'Feature selection: A data perspective', *ACM Comput. Surv.* **50**(6), 94:1–94:45. Available from https://dl.acm.org/citation.cfm?id=3136625 [accessed 07 May 2019].

Lladó, X., Ganiler, O., Oliver, A., Martí, R., Freixenet, J., Valls, L., Vilanova, J. C., Ramió-Torrentà, L. & Rovira, À. (2012), 'Automated detection of multiple sclerosis lesions in serial brain mri', *Neuroradiology* **54**(8), 787–807. Available from https://link.springer.com/article/10.1007/s00234-011-0992-6 [accessed 17 April 2019].

Loizou, C., Petroudi, S., Seimenis, I., Pantziaris, M. & Pattichis, C. (2015), 'Quantitative texture analysis of brain white matter lesions derived from t2-weighted mr images in ms patients with clinically isolated syndrome', *Journal of Neuroradiology* **42**(2), 99–114. Available from https://www.sciencedirect.com/science/article/pii/S0150986114001771 [accessed 08 January 2019].

Lötsch, J., Schiffmann, S., Schmitz, K., Brunkhorst, R., Lerch, F., Ferreiros, N., Wicker, S., Tegeder, I., Geisslinger, G. & Ultsch, A. (2018), 'Machine-learning based lipid mediator serum concentration patterns allow identification of multiple sclerosis patients with high accuracy', *Scientific reports* **8**. Available from https://www.nature.com/articles/s41598-018-33077-8.pdf?origin=ppub [accessed 08 May 2019].

Lu, M. (2019), 'Embedded feature selection accounting for unknown data heterogeneity', *Expert Systems with Applications* **119**, 350 – 361. Available from https://www.sciencedirect.com/science/article/pii/S0957417418307267 [accessed 07 May 2019].

Lubetzki, C. & Stankoff, B. (2014), Chapter 4 - demyelination in multiple sclerosis, *in* D. S. Goodin, ed., 'Multiple Sclerosis and Related Disorders', Vol. 122 of *Handbook of Clinical Neurology*, Elsevier, pp. 89 – 99.

Masdeu, J. & Belen, P. (2019), 'Fractional anisotropy'. Available from `https://www.sciencedirect.com/topics/neuroscience/fractional-anisotropy` [accessed 20 September 2019].

Mathworks (2019*a*), 'Support vector machines for binary classification'. Available from `https://uk.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html` [accessed 15 May 2019].

Mathworks (2019*b*), 'Treebagger'. Available from `https://uk.mathworks.com/help/stats/treebagger.html` [accessed 15 May 2019].

Mayerhoefer, M. E., Szomolanyi, P., Jirak, D., Materka, A. & Trattnig, S. (2009), 'Effects of mri acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study', *Medical physics* **36**(4), 1236–1243. Available from `https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3081408` [accessed 16 April 2019].

McConnell, R. (1986), 'Method of and apparatus for pattern recognition'. Available from `https://www.osti.gov/biblio/6007283` [accessed 30 April 2019].

Michoux, N., Guillet, A., Rommel, D., Mazzamuto, G., Sindic, C. & Duprez, T. (2015), 'Texture analysis of t2-weighted mr images to assess acute inflammation in brain ms lesions', *PloS one* **10**(12), e0145497. Available from `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0145497` [accessed 16 April 2019].

Mobahi, H., Rao, S. R., Yang, A. Y., Sastry, S. S. & Ma, Y. (2011), 'Segmentation of natural images by texture and boundary compression', *International journal of computer vision* **95**(1), 86–98. Available from `https://link.springer.com/article/10.1007/s11263-011-0444-0` [accessed 05 May 2019].

Monzel, R. (2007), 'haralicktexturefeatures'. Available from `https://www.mathworks.com/matlabcentral/fileexchange/58769-haralicktexturefeatures` [accessed 16 October 2019].

Morozova, O., Levina, O., Uusküla, A. & Heimer, R. (2015), 'Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in russia', *BMC Medical Research Methodology* **15**(1), 71. Available from `https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-015-0066-2` [accessed 07 May 2019].

Mustafi, S. M., Harezlak, J., Kodiweera, C., Randolph, J. S., Ford, J. C., Wishart, H. A. & Wu, Y.-C. (2019), 'Detecting white matter alterations in multiple sclerosis using advanced diffusion

magnetic resonance imaging', *Neural regeneration research* **14**(1), 114. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6262996/ [accessed 13 May 2019].

Nabizadeh, N. & Kubat, M. (2015), 'Brain tumors detection and segmentation in mr images: Gabor wavelet vs. statistical features', *Computers & Electrical Engineering* **45**, 286 – 301. Available from https://www.sciencedirect.com/science/article/pii/S0045790615000324 [accessed 06 May 2019].

Nair, T., Precup, D., Arnold, D. L. & Arbel, T. (2018), 'Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation', *CoRR* **abs/1808.01200**. Available from http://arxiv.org/abs/1808.01200 [accessed 03 May 2019].

NLoM (2013), 'The hand of mrs. wilhelm roentgen: the first x-ray image, 1895'. Available from https://www.nlm.nih.gov/dreamanatomy/da_g_Z-1.html [accessed 12 May 2019].

Nogueira, M. A., Abreu, P. H., Martins, P., Machado, P., Duarte, H. & Santos, J. (2017), 'Image descriptors in radiology images: a systematic review', *Artificial Intelligence Review* **47**(4), 531–559. Available from https://link.springer.com/article/10.1007/s10462-016-9492-8 [accessed 28 July 2019].

NTU (2013), 'Treebagger'. Available from http://www.ntu.edu.sg/home/egbhuang/ [accessed 15 May 2019].

Ojala, T., Pietikainen, M. & Maenpaa, T. (2002), 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7), 971–987. Available from https://ieeexplore.ieee.org/document/1017623 [accessed 30 April 2019].

Ojala, T., Pietikäinen, M. & Harwood, D. (1996), 'A comparative study of texture measures with classification based on featured distributions', *Pattern Recognition* **29**(1), 51 – 59. Available from https://www.sciencedirect.com/science/article/pii/0031320395000674 [accessed 30 April 2019].

Öztürk, S. & Akdemir, B. (2018), 'Application of feature extraction and classification methods for histopathological image using glcm, lbp, lbglcm, glrlm and sfta', *Proc. Comput. Sci* **132**, 40–46. Available from https://www.sciencedirect.com/science/article/pii/S1877050918307890 [accessed 08 May 2019].

Parmar, C., Grossmann, P., Bussink, J., Lambin, P. & Aerts, H. J. (2015), 'Machine learning methods for quantitative radiomic biomarkers', *Scientific reports* **5**, 13087. Available from https://www.nature.com/articles/srep13087 [accessed 03 May 2019].

Peng, X., Lin, P., Zhang, T. & Wang, J. (2013), 'Extreme learning machine-based classification of adhd using brain structural mri data', *PLOS ONE* **8**(11), 1–12. Available from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0079476 [accessed 08 May 2019].

Pietikäinen, M. & Zhao, G. (2015), Two decades of local binary patterns: A survey, *in* 'Advances in independent component analysis and learning machines', Elsevier, pp. 175–210. Available from http://www.sciencedirect.com/science/article/pii/B9780128028063000099 [accessed 06 May 2019].

Qian, Z., Li, Y., Wang, Y., Li, L., Li, R., Wang, K., Li, S., Tang, K., Zhang, C., Fan, X., Chen, B. & Li, W. (2019), 'Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers', *Cancer Letters* **451**, 128 – 135. Available from http://www.sciencedirect.com/science/article/pii/S0304383519301545 [accessed 05 May 2019].

Rampun, A., Wang, L., Malcolm, P. & Zwiggelaar, R. (2016), 'A quantitative study of texture features across different window sizes in prostate t2-weighted mri', *Procedia Computer Science* **90**, 74–79. Available from https://www.sciencedirect.com/science/article/pii/S1877050916312066 [accessed 05 May 2019].

Reischauer, C., Patzwahl, R., Koh, D.-M., Froehlich, J. M. & Gutzeit, A. (2018), 'Texture analysis of apparent diffusion coefficient maps for treatment response assessment in prostate cancer bone metastases—a pilot study', *European journal of radiology* **101**, 184–190. Available from https://www.sciencedirect.com/science/article/pii/S0720048X1830069X [accessed 05 May 2019].

Rodriguez-Galiano, V., Luque-Espinar, J., Chica-Olmo, M. & Mendes, M. (2018), 'Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods', *Science of The Total Environment* **624**, 661 – 672. Available from https://www.sciencedirect.com/science/article/pii/S0048969717335751 [accessed 07 May 2019].

Roffo, G. (2016), 'Feature selection library (matlab toolbox)', *arXiv preprint arXiv:1607.01327* . Available from https://arxiv.org/pdf/1607.01327.pdf [accessed 14 May 2019].

Roffo, G. & Melzi, S. (2016), 'Feature selection via eigenvector centrality', *CoRR* pp. 88–97. Available from https://arxiv.org/pdf/1706.05933.pdf [accessed 14 May 2019].

Roffo, G., Melzi, S., Castellani, U. & Vinciarelli, A. (2017), Infinite latent feature selection: A probabilistic latent graph-based ranking approach, *in* 'Proceedings of the IEEE International Conference on Computer Vision', pp. 1398–1406. Available from https://arxiv.org/pdf/1707.07538.pdf [accessed 14 May 2019].

Roura, E., Oliver, A., Cabezas, M., Valverde, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À. & Lladó, X. (2015), 'A toolbox for multiple sclerosis lesion segmentation', *Neuroradiology* **57**(10), 1031–1043. Available from https://link.springer.com/article/10.1007/s00234-015-1552-2 [accessed 14 October 2017].

Roy, P. K., Bhuiyan, A. & Ramamohanarao, K. (2013), Automated segmentation of multiple sclerosis lesion in intensity enhanced flair mri using texture features and support vector machine, *in* '2013 IEEE International Conference on Image Processing', IEEE, pp. 4277–4281. Available from https://ieeexplore.ieee.org/abstract/document/6738881 [accessed 15 February 2019].

Rudick, R. A., Cohen, J. A., Weinstock-Guttman, B., Kinkel, R. P. & Ransohoff, R. M. (1997), 'Management of multiple sclerosis', *New England Journal of Medicine* **337**(22), 1604–1611. Available from https://www.nejm.org/doi/full/10.1056/NEJM199711273372207 [accessed 23 December 2018].

Saccà, V., Sarica, A., Novellino, F., Barone, S., Tallarico, T., Filippelli, E., Granata, A., Chiriaco, C., Bossio, R. B., Valentino, P. et al. (2018), 'Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state fmri connectivity data', *Brain imaging and behavior* pp. 1–12. Available from https://link.springer.com/article/10.1007/s11682-018-9926-9 [accessed 30 March 2019].

Salem, M., Cabezas, M., Valverde, S., Pareto, D., Oliver, A., Salvi, J., Àlex Rovira & Lladó, X. (2018), 'A supervised framework with intensity subtraction and deformation field features for the detection of new t2-w lesions in multiple sclerosis', *NeuroImage: Clinical* **17**, 607 – 615. Available from https://www.sciencedirect.com/science/article/pii/S2213158217302954 [accessed 18 July 2019].
URL: *http://www.sciencedirect.com/science/article/pii/S2213158217302954*

Samah, Y., Yassine, B. S. & Naceur, A. M. (2018), Multiple sclerosis lesions detection from noisy magnetic resonance brain images tissue, *in* '2018 15th International Multi-Conference on Systems, Signals & Devices (SSD)', IEEE, pp. 240–245. Available from https://ieeexplore.ieee.org/abstract/document/8570679 [accessed 13 April 2019].

Schmidt, P., Pongratz, V., Küster, P., Meier, D., Wuerfel, J., Lukas, C., Bellenberg, B., Zipp, F., Groppa, S., Sämann, P. G., Weber, F., Gaser, C., Franke, T., Bussas, M., Kirschke, J., Zimmer, C., Hemmer, B. & Mühlau, M. (2019), 'Automated segmentation of changes in flair-hyperintense white matter lesions in multiple sclerosis on serial magnetic resonance imaging', *NeuroImage: Clinical* **23**, 101849. Available from https://www.sciencedirect.com/science/article/pii/S2213158219301998#bb0035 [accessed 13 May 2019].

Schwartz, W. R., de Siqueira, F. R. & Pedrini, H. (2012), 'Evaluation of feature descriptors for texture classification', *Journal of Electronic Imaging* **21**(2), 023016. Available from https://www.spiedigitallibrary.org/journals/Journal-of-Electronic-Imaging/volume-21/issue-2/023016/Evaluation-of-feature-descriptors-for-texture-classification/10.1117/1.JEI.21.2.023016.short?SSO=1 [accessed 05 May 2019].

Schwenkenbecher, P., Wurster, U., Konen, F. F., Gingele, S., Sühs, K.-W., Wattjes, M. P., Stangel, M. & Skripuletz, T. (2019), 'Impact of the mcdonald criteria 2017 on early diagnosis of

relapsing-remitting multiple sclerosis', *Frontiers in Neurology* **10**. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6428717/ [accessed 12 May 2019].

Seeram, E. (2015), *Computed tomography: physical principles, clinical applications, and quality control*, Elsevier Health Sciences.

Sharma, S., Laule, C., Moore, G. W., Li, D. K. & Zhang, Y. (2019), 'Correlating new directional measures of myelin and axonal integrity in t2-weighted mri with quantitative histology in multiple sclerosis', *Journal of Neuroscience Methods* **311**, 369 – 376. Available from https://www.sciencedirect.com/science/article/pii/S0165027018302875 [accessed 04 May 2019].

Sharma, V., Baruah, D., Chutia, D., Raju, P. & Bhattacharya, D. K. (2016), An assessment of support vector machine kernel parameters using remotely sensed satellite data, *in* '2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)', pp. 1567–1570. Available from https://ieeexplore.ieee.org/document/7808096 [accessed 08 May 2019].

Sheng, M., Dong, Z. & Xie, Y. (2018), 'Identification of tumor-educated platelet biomarkers of non-small-cell lung cancer', *OncoTargets and therapy* **11**, 8143. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6241732/ [accessed 07 May 2019].

Shoshtari, A. S., Lin, S., McKeown, M. J. & Wang, Z. J. (2016), A multimodal data fusion approach efficiently predicts disease duration in multiple sclerosis, *in* '2016 International Joint Conference on Neural Networks (IJCNN)', pp. 4518–4524. Available from https://ieeexplore.ieee.org/document/7727791 [accessed 03 May 2019].

Silva, M., Milanese, G., Seletti, V., Ariani, A. & Sverzellati, N. (2018), 'Pulmonary quantitative ct imaging in focal and diffuse disease: current research and clinical applications', *The British journal of radiology* **91**(xxxx), 20170644. Available from https://www.birpublications.org/doi/full/10.1259/bjr.20170644 [accessed 12 May 2019].

Smirniotopoulos, J., Smith, A., Rushing, E., Murphy, F. & Horkayne-Szakaly, I. (2009), 'Radiologic-pathologic correlation of white matter disease', *The Neuroradiology Journal* **22**, 26–32. Available from https://journals.sagepub.com/doi/pdf/10.1177/19714009090220S107?casa_token=sBV-MQuSr1oAAAAA:7R1qnnJsy5sn1sL3EKGYQo599Cikw0OxIHrxLZJ8KcSSqznoW1wOVvycVceeauWwttPnNiD5Y_5wqg [accessed 15 December 2018].

Song, Y., Zhang, S., He, B., Sha, Q., Shen, Y., Yan, T., Nian, R. & Lendasse, A. (2018), 'Gaussian derivative models and ensemble extreme learning machine for texture image classification', *Neurocomputing* **277**, 53 – 64. Available from https://www.sciencedirect.com/science/article/pii/S0925231217314017#! [accessed 10 May 2019].

Soto, Y. & Nadrljanski, M. (2019), 'Computed tomography'. Available from https://radiopaedia.org/articles/computed-tomography?lang=gb [accessed 12 May 2019].

Speiser, J. L., Wolf, B. J., Chung, D., Karvellas, C. J., Koch, D. G. & Durkalski, V. L. (2019), 'Bimm forest: A random forest method for modeling clustered and longitudinal binary outcomes', *Chemometrics and Intelligent Laboratory Systems* **185**, 122 – 134. Available from https://www.sciencedirect.com/science/article/pii/S0169743918304362 [accessed 08 May 2019].

Sweeney, E. M., Vogelstein, J. T., Cuzzocreo, J. L., Calabresi, P. A., Reich, D. S., Crainiceanu, C. M. & Shinohara, R. T. (2014), 'A comparison of supervised machine learning algorithms and feature vectors for ms lesion segmentation using multimodal structural mri', *PloS one* **9**(4), e95753. Available from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0095753 [accessed 19 March 2019].

Sørensen, L. & Nielsen, M. (2018), 'Ensemble support vector machine classification of dementia using structural mri and mini-mental state examination', *Journal of Neuroscience Methods* **302**, 66 – 74. Available from https://www.sciencedirect.com/science/article/pii/S0165027018300177 [accessed 08 May 2019].

Thaler, C., Faizy, T. D., Sedlacik, J., Bester, M., Stellmann, J.-P., Heesen, C., Fiehler, J. & Siemonsen, S. (2018), 'The use of multiparametric quantitative magnetic resonance imaging for evaluating visually assigned lesion groups in patients with multiple sclerosis', *Journal of Neurology* **265**(1), 127–133. Available from https://pubs.rsna.org/doi/full/10.1148/radiol.2015151169 [accessed 13 May 2019].

Thanh Noi, P. & Kappas, M. (2018), 'Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery', *Sensors* **18**(1), 18. Available from https://www.mdpi.com/1424-8220/18/1/18 [accessed 08 May 2019].

Theocharakis, P., Glotsos, D., Kalatzis, I., Kostopoulos, S., Georgiadis, P., Sifaki, K., Tsakouridou, K., Malamas, M., Delibasis, G., Cavouras, D. et al. (2009), 'Pattern recognition system for the discrimination of multiple sclerosis from cerebral microangiopathy lesions based on texture analysis of magnetic resonance images', *Magnetic resonance imaging* **27**(3), 417–422. Available from https://www.sciencedirect.com/science/article/abs/pii/S0730725X08002592 [accessed 17 April 2019].

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288. Available from http://www.jstor.org/stable/2346178 [accessed 07 May 2019].

Tin Kam Ho (1995), Random decision forests, *in* 'Proceedings of 3rd International Conference on Document Analysis and Recognition', Vol. 1, pp. 278–282 vol.1. Available from https://ieeexplore.ieee.org/document/598994 [accessed 08 May 2019].

Tozer, D., Marongiu, G., Swanton, J., Thompson, A. & Miller, D. (2009), 'Texture analysis of magnetization transfer maps from patients with clinically isolated syndrome and multiple sclerosis', *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* **30**(3), 506–513. Available from https://onlinelibrary.wiley.com/doi/full/10.1002/jmri.21885 [accessed 17 April 2019].

Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., Oliver, A. & Lladó, X. (2017), 'Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach', *NeuroImage* **155**, 159–168. Available from https://www.sciencedirect.com/science/article/pii/S1053811917303270 [accessed 10 January 2019].

VenkateswarLal, P., Nitta, G. R. & Prasad, A. (2019), 'Ensemble of texture and shape descriptors using support vector machine classification for face recognition', *Journal of Ambient Intelligence and Humanized Computing* . Available from https://link.springer.com/article/10.1007/s12652-019-01192-7 [accessed 06 May 2019].

Verma, A. & Bashir, U. (2019), 'Diffusion weighted imaging'. Available from https://radiopaedia.org/articles/diffusion-weighted-imaging-1?lang=gb [accessed 17 September 2019].

Verma, R. K., Wiest, R., Locher, C., Heldner, M. R., Schucht, P., Raabe, A., Gralla, J., Kamm, C. P., Slotboom, J. & Kellner-Weldon, F. (2017), 'Differentiating enhancing multiple sclerosis lesions, glioblastoma, and lymphoma with dynamic texture parameters analysis (dtpa): A feasibility study', *Medical physics* **44**(8), 4000–4008. Available from https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.12356 [accessed 17 April 2019].

Vial, A., Stirling, D., Field, M., Ros, M., Ritz, C., Carolan, M., Holloway, L. & Miller, A. A. (2018), 'The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review', *Translational Cancer Research* **7**(3). Available from http://tcr.amegroups.com/article/view/21823 [accessed 03 May 2019].

Wattjes, M., Steenwijk, M. & Stangel, M. (2015), 'Mri in the diagnosis and monitoring of multiple sclerosis: an update', *Clinical neuroradiology* **25**(2), 157–165. Available from https://link.springer.com/article/10.1007/s00062-015-0430-y [accessed 13 May 2019].

Webel, J., Gola, J., Britz, D. & Mücklich, F. (2018), 'A new analysis approach based on haralick texture features for the characterization of microstructure on the example of low-alloy steels', *Materials Characterization* **144**, 584–596. Available from

https://www.sciencedirect.com/science/article/pii/S1044580318313871 [accessed 16 April 2019].

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Colliot, O., Stankoff, B. & Ayache, N. (2018), Flair mr image synthesis by using 3d fully convolutional networks for multiple sclerosis, *in* 'ISMRM-ESMRMB 2018-Joint Annual Meeting', pp. 1–6. Available from https://hal.inria.fr/hal-01723070/ [accessed 08 May 2019].

Wei, X. (2007), 'Gray level run length matrix toolbox v1.0'. Available from https://www.mathworks.com/matlabcentral/fileexchange/17482-gray-level-run-length-matrix-toolbox [accessed 16 October 2019].

Wilson, James, A. & Smirniotopoulos, James, G. (2019), 'Brain imaging in multiple sclerosis'. Available from https://emedicine.medscape.com/article/342254-overview [accessed 13 May 2019].

Winklewski, P. J., Sabisz, A., Naumczyk, P., Jodzio, K., Szurowska, E. & Szarmach, A. (2018), 'Understanding the physiopathology behind axial and radial diffusivity changes—what do we know?', *Frontiers in neurology* **9**, 92. Available from https://www.frontiersin.org/articles/10.3389/fneur.2018.00092/full [accessed 20 September 2019].

Xia, J., Ghamisi, P., Yokoya, N. & Iwasaki, A. (2018), 'Random forest ensembles and extended multiextinction profiles for hyperspectral image classification', *IEEE Transactions on Geoscience and Remote Sensing* **56**(1), 202–216. Available from https://ieeexplore.ieee.org/abstract/document/8046025 [accessed 08 May 2019].

Yang, Y., Yan, L.-F., Zhang, X., Nan, H.-Y., Hu, Y.-C., Han, Y., Zhang, J., Liu, Z.-C., Sun, Y.-Z., Tian, Q., Yu, Y., Sun, Q., Wang, S.-Y., Zhang, X., Wang, W. & Cui, G.-B. (2019), 'Optimizing texture retrieving model for multimodal mr image-based support vector machine for classifying glioma', *Journal of Magnetic Resonance Imaging* **49**(5), 1263–1274. Available from https://onlinelibrary.wiley.com/doi/full/10.1002/jmri.26524 [accessed 08 May 2019].

Yoo, Y., Tang, L. Y., Brosch, T., Li, D. K., Kolind, S., Vavasour, I., Rauscher, A., MacKay, A. L., Traboulsee, A. & Tam, R. C. (2018), 'Deep learning of joint myelin and t1w mri features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls', *NeuroImage: Clinical* **17**, 169–178. Available from https://www.sciencedirect.com/science/article/pii/S2213158217302553 [accessed 19 March 2019].

Yu, H., Yang, X., Zheng, S. & Sun, C. (2019), 'Active learning from imbalanced data: A solution of online weighted extreme learning machine', *IEEE Transactions on Neural Networks and Learning Systems* **30**(4), 1088–1103. Available from https://ieeexplore.ieee.org/document/8443399 [accessed 10 May 2019].

Zhang, J., Tong, L., Wang, L. & Li, N. (2008), 'Texture analysis of multiple sclerosis: a comparative study', *Magnetic resonance imaging* **26**(8), 1160–1166. Available from https://www.sciencedirect.com/science/article/abs/pii/S0730725X08000660 [accessed 17 April 2019].

Zhang, X., Wei, C. & Wan, X. (2018), Mammogram classification method based on gmm and glcm-pso-pnn, *in* '2018 3rd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2018)', Atlantis Press. Available from https://www.atlantis-press.com/proceedings/amcce-18/25895678 [accessed 10 May 2019].

Zhang, Y.-D., Zhang, Y., Phillips, P., Dong, Z. & Wang, S. (2017), 'Synthetic minority oversampling technique and fractal dimension for identifying multiple sclerosis', *Fractals* **25**(04), 1740010. Available from https://www.worldscientific.com/doi/abs/10.1142/S0218348X17400102 [accessed 15 April 2019].

Zhao, J., Meng, Z., Wei, L., Sun, C., Zou, Q. & Su, R. (2019), 'Supervised brain tumor segmentation based on gradient and context-sensitive features', *Frontiers in Neuroscience* **13**, 144. Available from https://www.frontiersin.org/article/10.3389/fnins.2019.00144 [accessed 07 May 2019].

Zhong, Y., Qi, S., Kang, Y., Feng, W. & Haacke, E. M. (2012), Automatic skull stripping in brain mri based on local moment of inertia structure tensor, *in* '2012 IEEE International Conference on Information and Automation', IEEE, pp. 437–440. Available from https://ieeexplore.ieee.org/abstract/document/6246845 [accessed 01 May 2019].

Zhou, M., Scott, J., Chaudhury, B., Hall, L., Goldgof, D., Yeom, K. W., Iv, M., Ou, Y., Kalpathy-Cramer, J., Napel, S. et al. (2018), 'Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches', *American Journal of Neuroradiology* **39**(2), 208–216. Available from https://www.ncbi.nlm.nih.gov/pubmed/28982791 [accessed 03 May 2019].

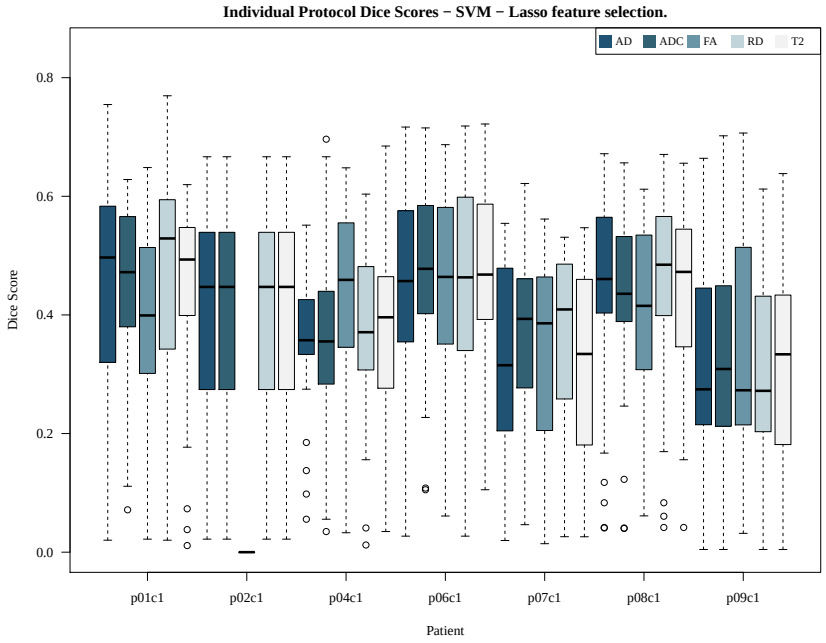
Zhu, M., Xia, J., Jin, X., Yan, M., Cai, G., Yan, J. & Ning, G. (2018), 'Class weights random forest algorithm for processing class imbalanced medical data', *IEEE Access* **6**, 4641–4652. Available from https://ieeexplore.ieee.org/document/8246503 [accessed 08 May 2019].

Zhu, Q.-X., Fan, Y., He, Y.-L. & Xu, Y. (2018), Effective cancer classification based on gene expression data using multidimensional mutual information and elm, *in* '2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)', IEEE, pp. 954–958. Available from https://www.researchgate.net/publication/329495412_Effective_Cancer_Classification_based_on_Gene_Expression_Data_using_Multidimensional_Mutual_Information_and_ELM [accessed 18 July 2019].

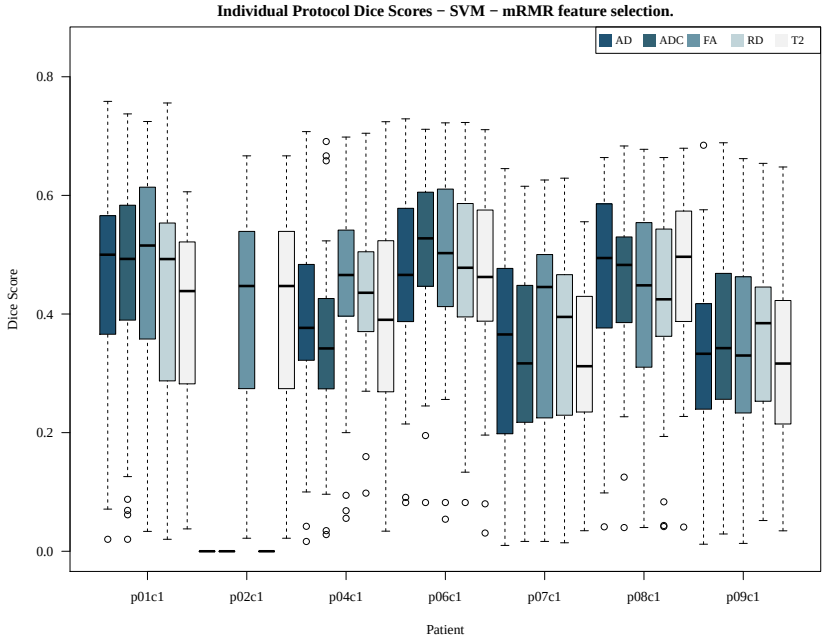
Zurita, M., Montalba, C., Labbé, T., Cruz, J. P., da Rocha, J. D., Tejos, C., Ciampi, E., Cárcamo, C., Sitaram, R. & Uribe, S. (2018), 'Characterization of relapsing-remitting multiple

sclerosis patients using support vector machine classifications of functional and diffusion mri data', *NeuroImage: Clinical* **20**, 724 – 730. Available from `http://www.sciencedirect.com/science/article/pii/S221315821830278X` [accessed 08 May 2019].

A Results from Unseen Dataset Classification - Individual

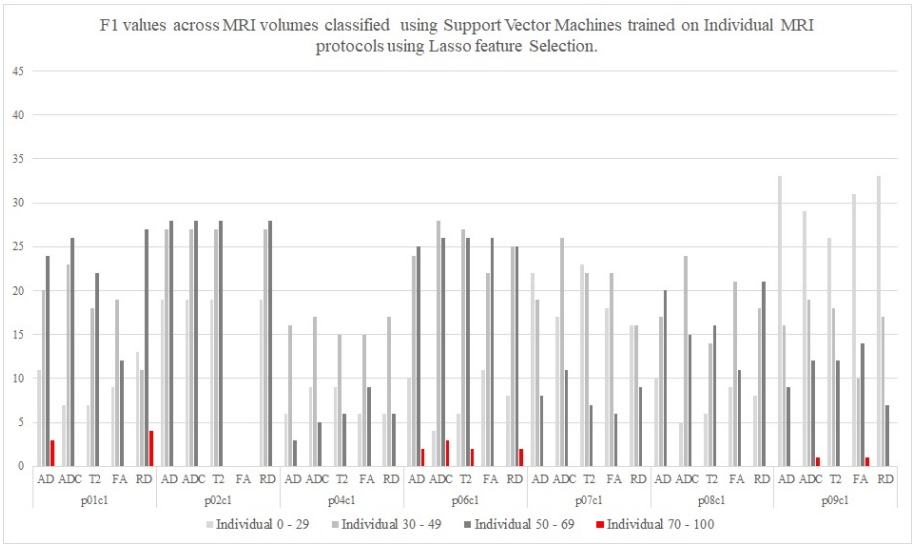


(a)

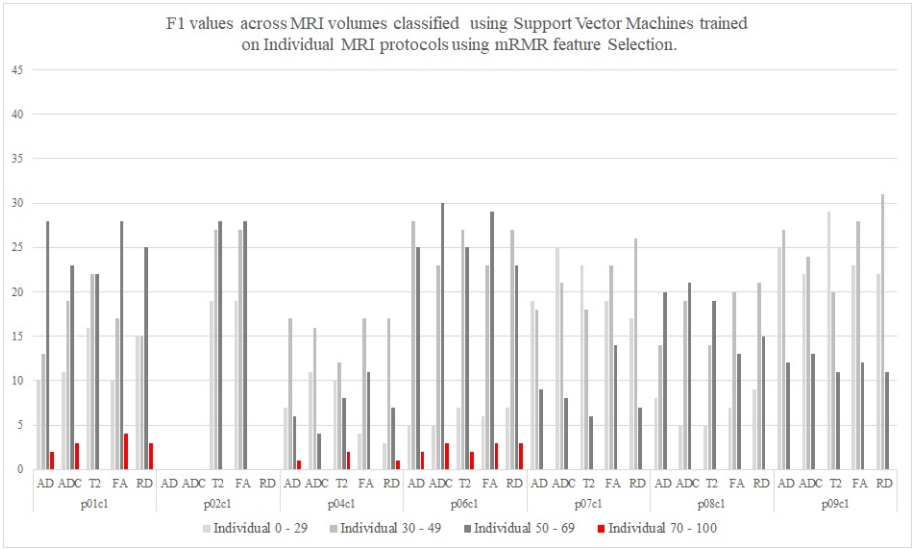


(b)

Figure 63: (a) Dice score performance for entire MRI volumes using Support Vector Machine Models trained on Individual MRI sequences and LASSO Feature Selection. (b) As for (a) but using mRMR feature Selection.

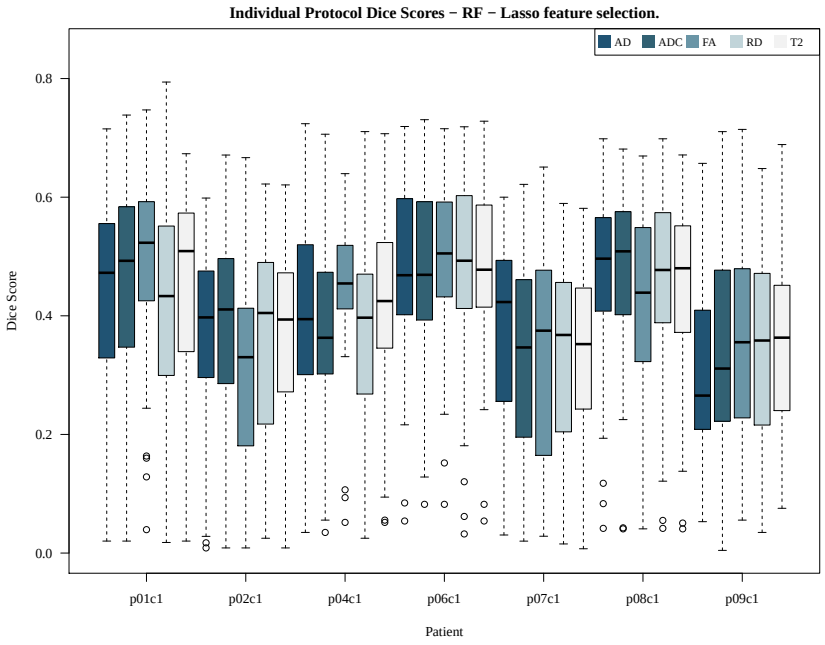


(a)

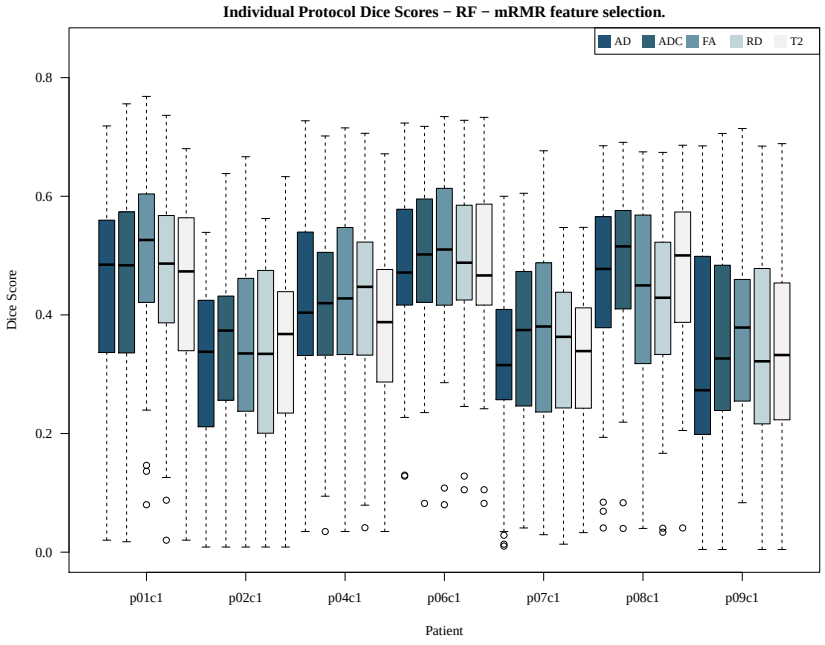


(b)

Figure 64: (a) Bar charts displaying binned distributions of slice accuracies for Individual SVM experiments using (a) LASSO feature selection and (b) mRMR feature selection methods. Red represents the number of slices within a volume scoring greater than 70.

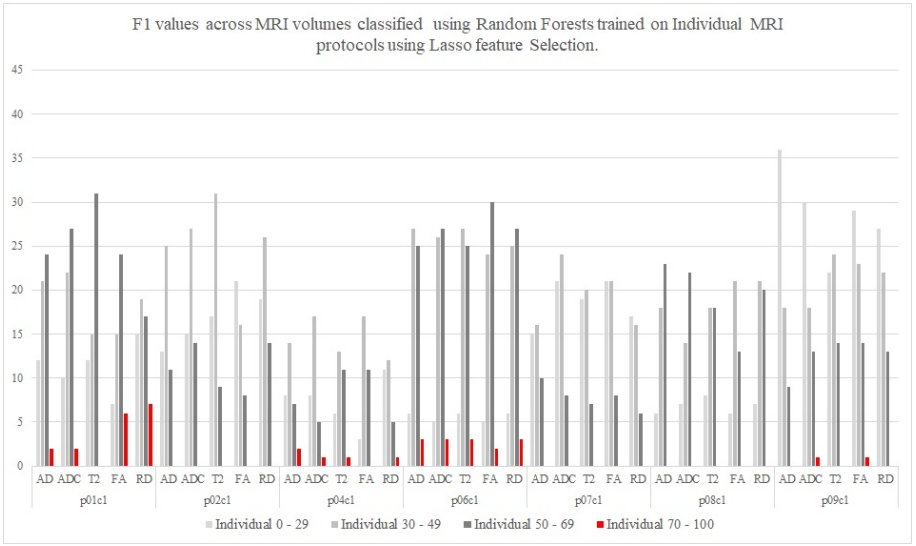


(a)

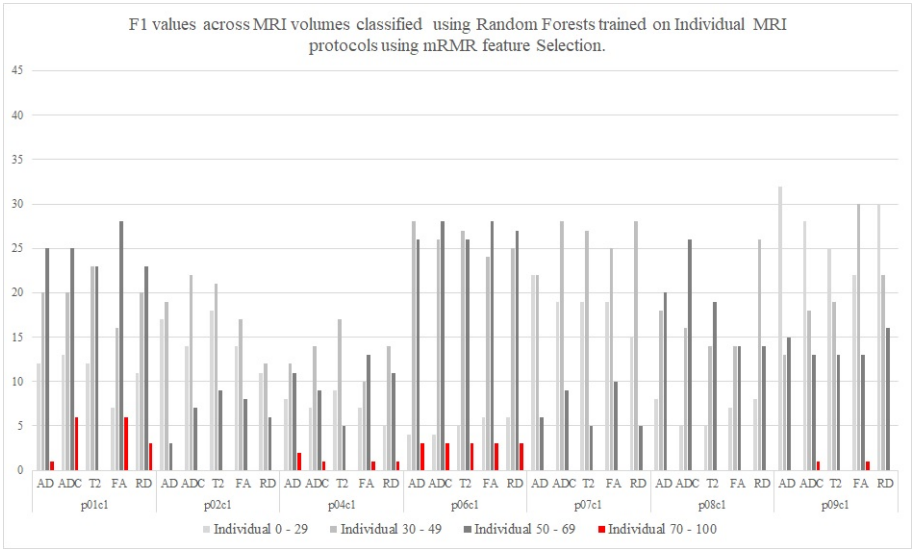


(b)

Figure 65: (a) Dice score performance for entire MRI volumes using Random Forest Models trained on Individual MRI sequences and LASSO Feature Selection. (b) As for (a) but using mRMR feature Selection.

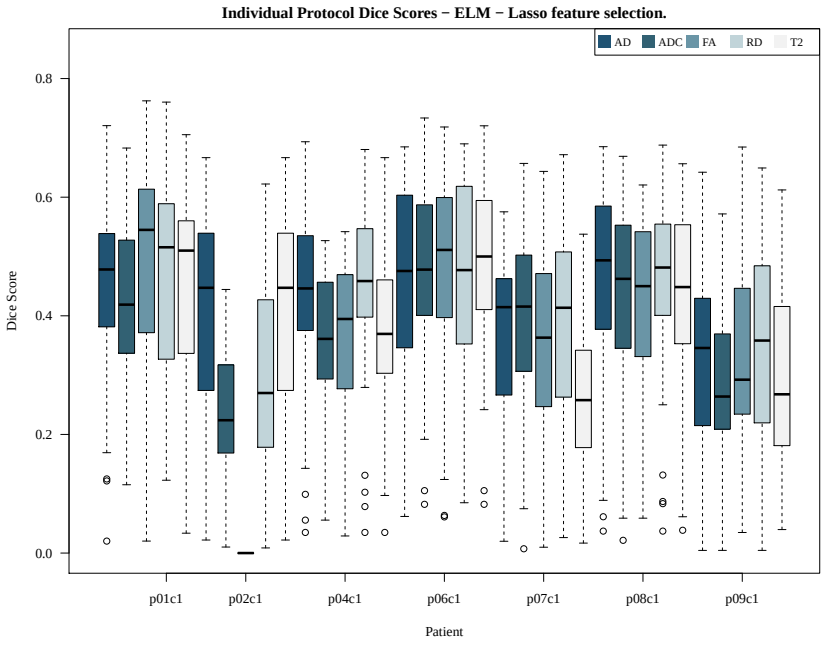


(a)

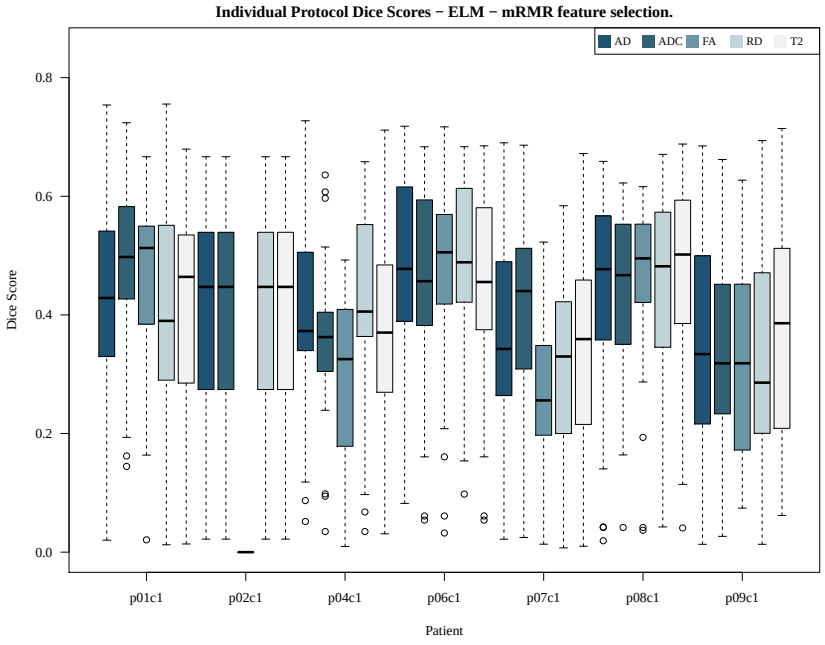


(b)

Figure 66: (a) Bar charts displaying binned distributions of slice accuracies for Individual RF experiments using (a) LASSO feature selection and (b) mRMR feature selection methods. Red represents the number of slices within a volume scoring greater than 70.

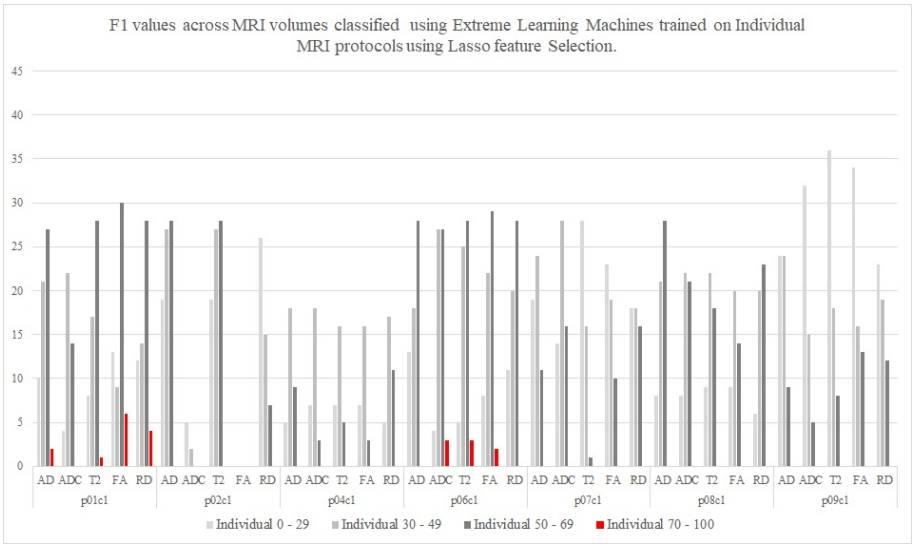


(a)

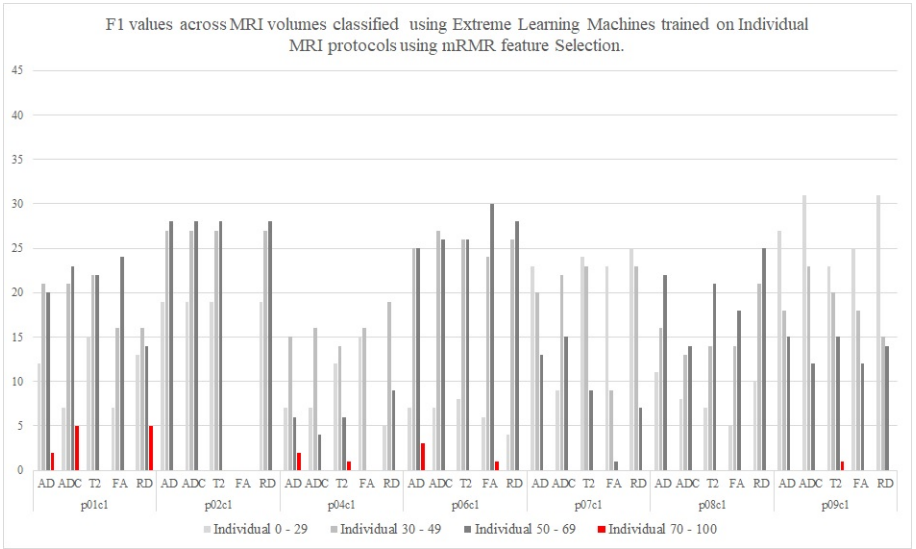


(b)

Figure 67: (a) Dice score performance for entire MRI volumes using Extreme Learning Machine Models trained on Individual MRI sequences and LASSO feature Selection.(b) As for (a) but using mRMR feature Selection.

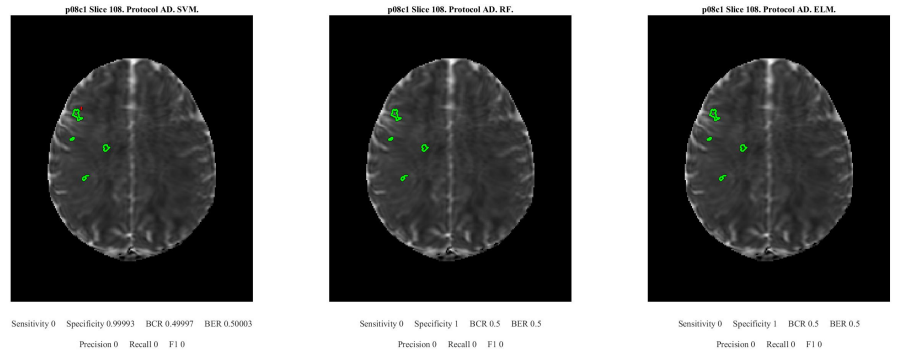


(a)

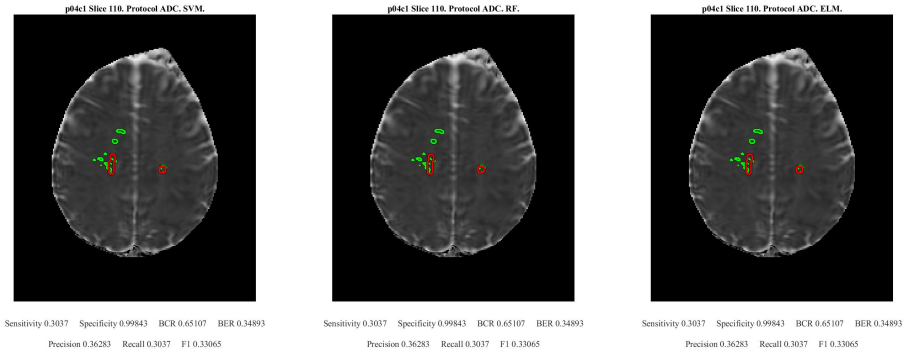


(b)

Figure 68: (a) Bar charts displaying binned distributions of slice accuracies for Individual ELM experiments using (a) LASSO feature selection and (b) mRMR feature selection methods. Red represents the number of slices within a volume scoring greater than 70.

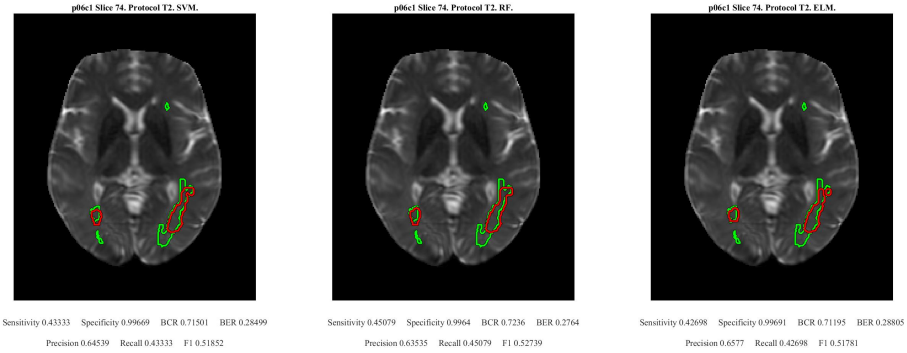


(a)

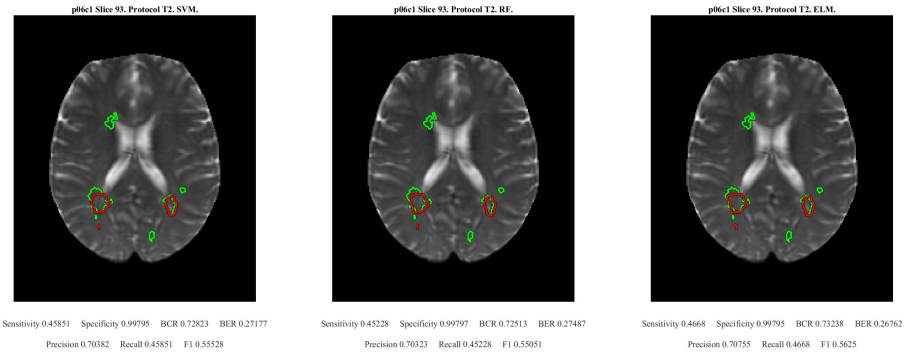


(b)

Figure 69: (a) Segmentations using all machine learning classifiers trained on AD MRI sequence and features selected using LASSO feature selection for MS identification of dataset p08c1 AD sequence on slice 109. (b) Models trained on ADC data only used to predict p04c1 slice 110. Both using mRMR feature selection.



(a)



(b)

Figure 70: (a) Segmentations using all machine learning classifiers trained on T2 MRI sequence and features selected using mRMR feature selection for MS identification of dataset p06c1 T2 sequence on slice 74. (b) Models trained on T2 data only used to predict p06c1 slice 93 with LASSO feature selection.

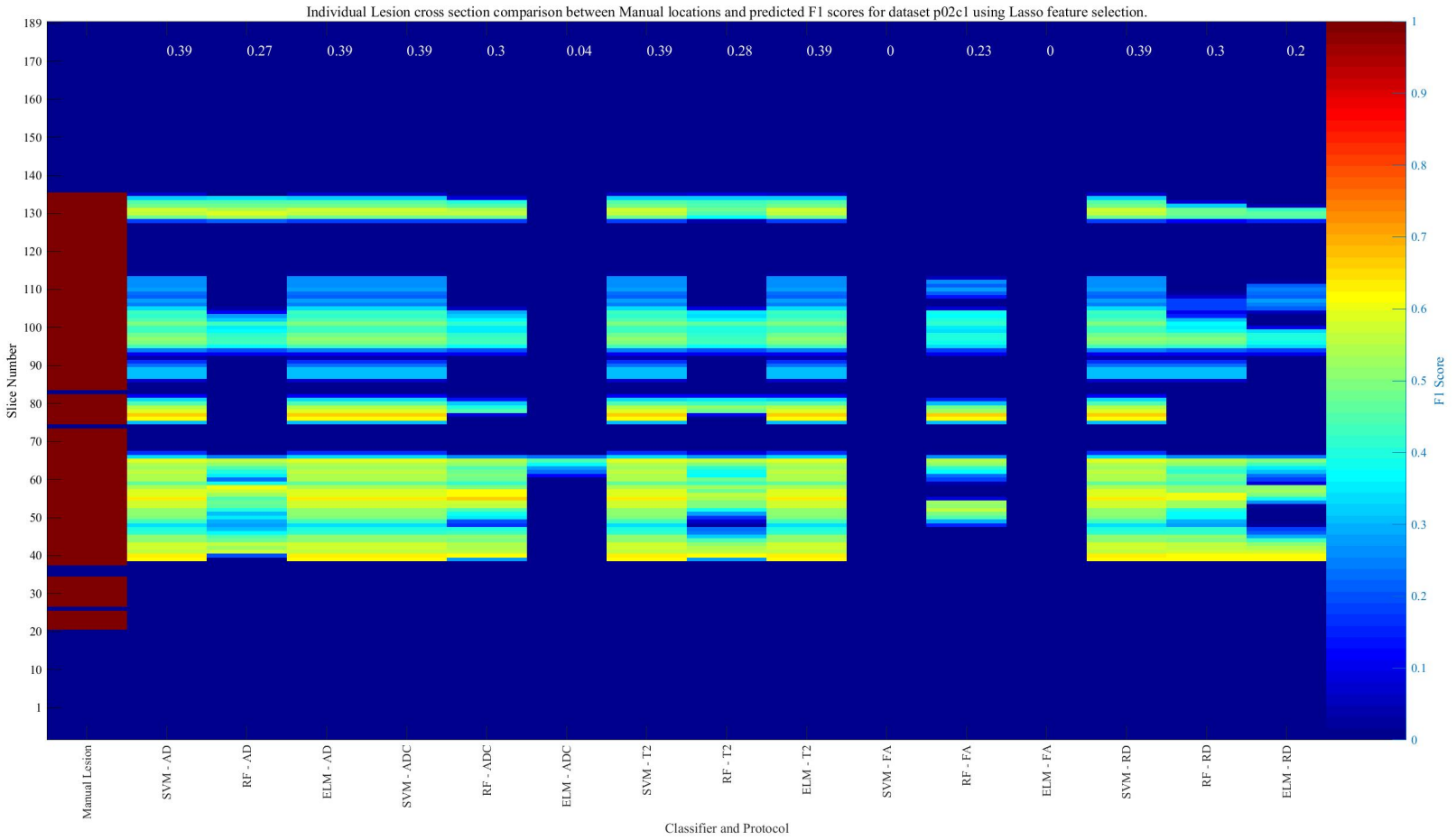


Figure 71: Cross sectional graph for p02c1 using Machine Learning models trained on Individual MRI data and LASSO feature selection.

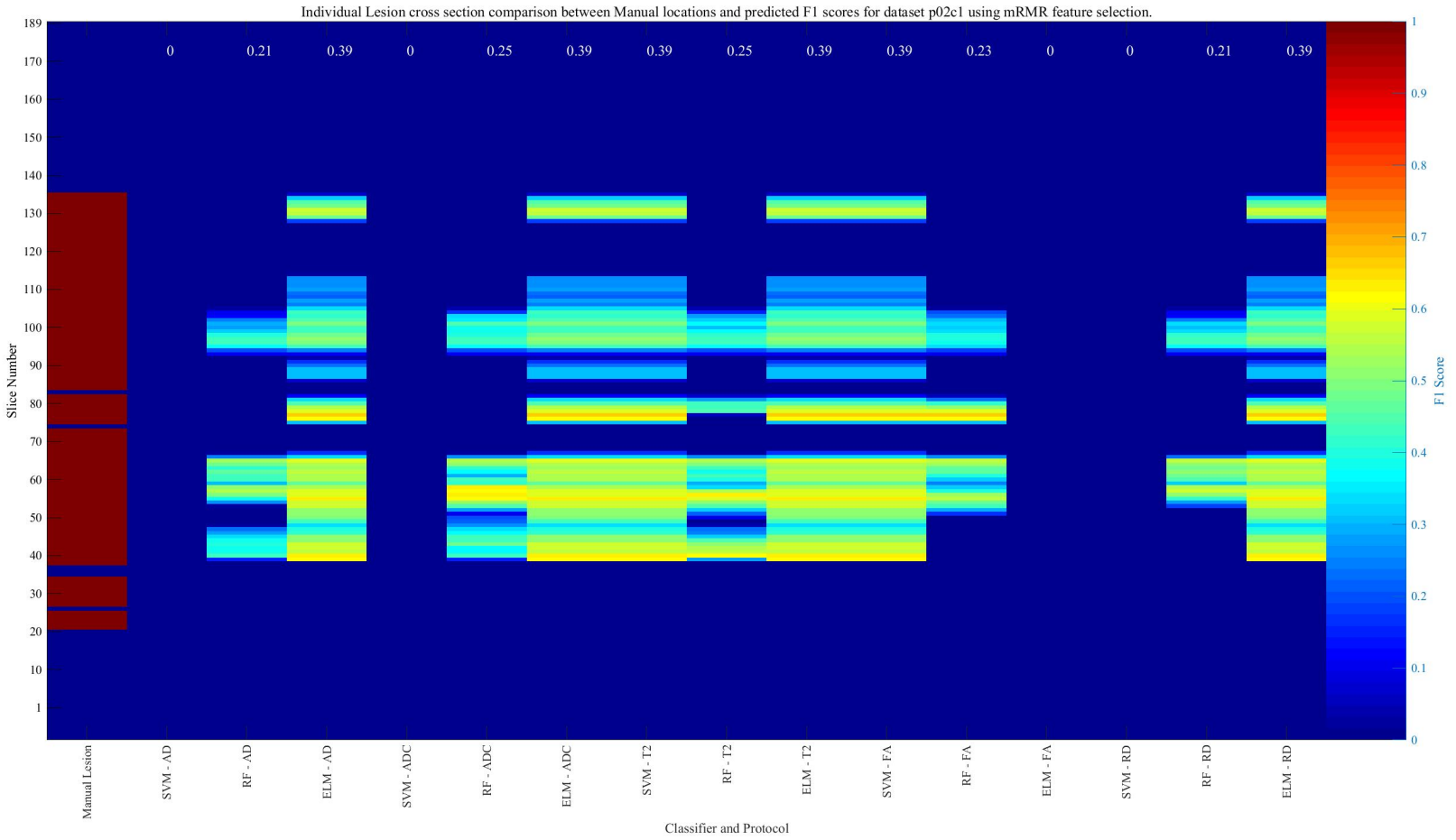
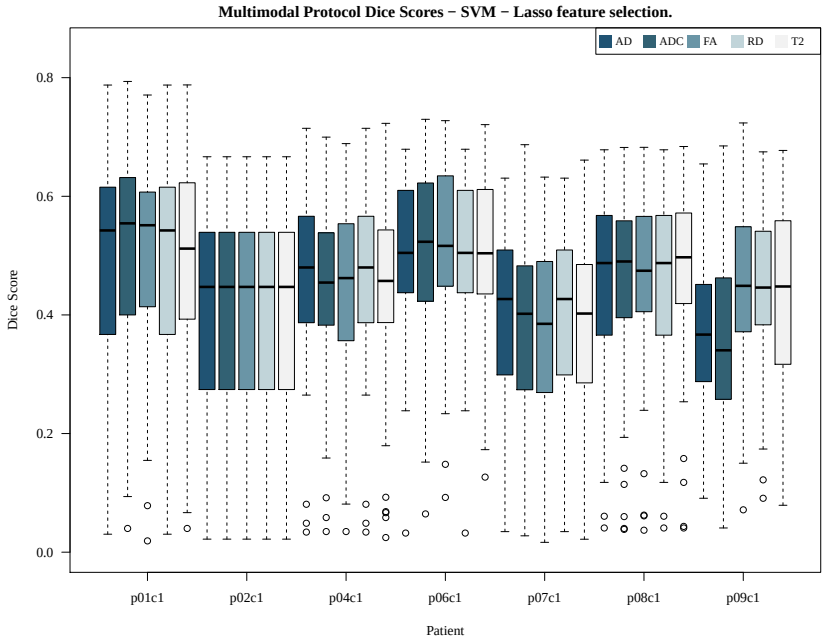
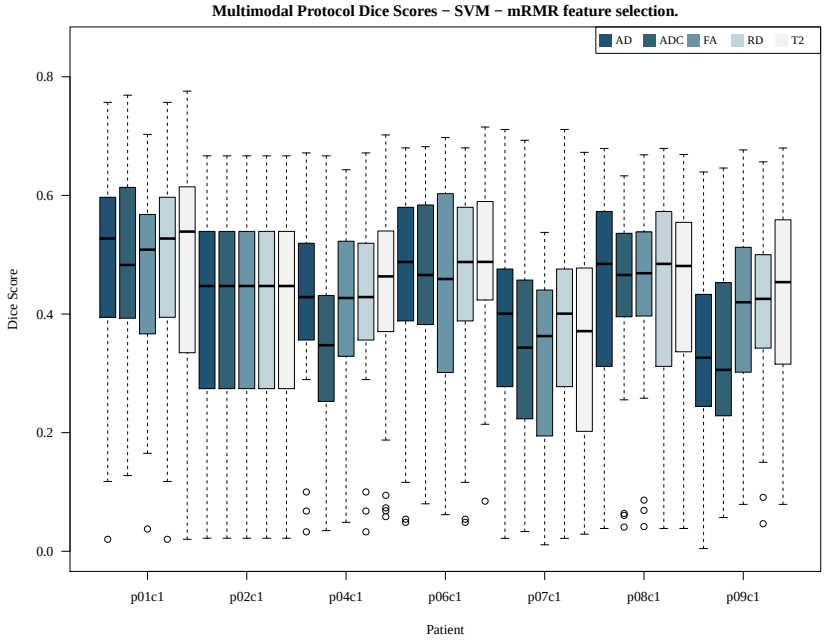


Figure 72: Cross sectional graph for p02c1 using Machine Learning models trained on Individual MRI data and mRMR feature selection.

B Results from Unseen Dataset Classification - Multimodal

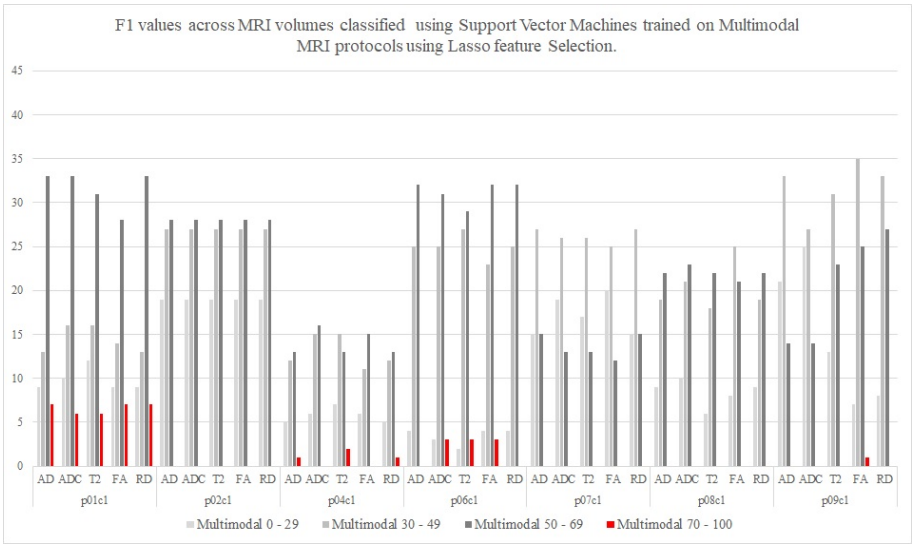


(a)

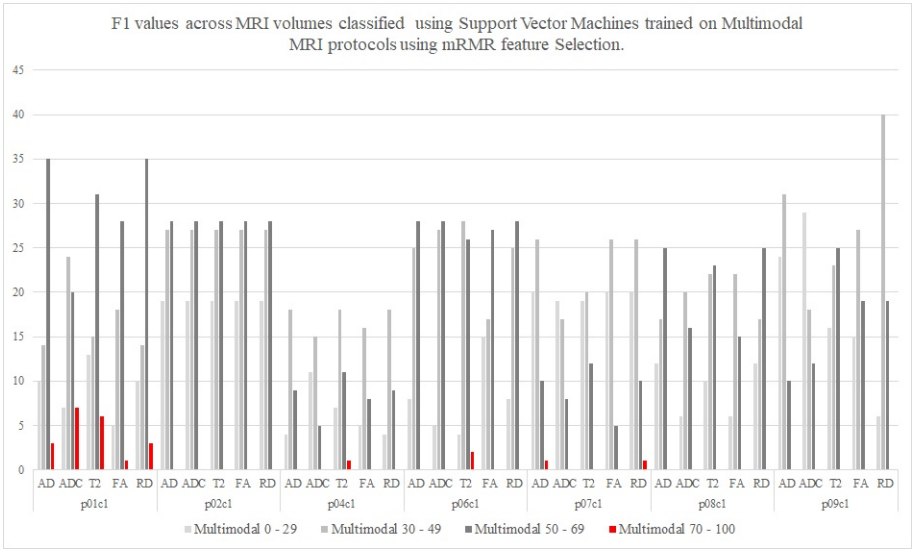


(b)

Figure 73: (a) Dice score performance for entire MRI volumes using Support Vector Machine Models trained on Multimodal MRI data and LASSO Feature Selection. (b) As for (a) but using mRMR feature Selection.

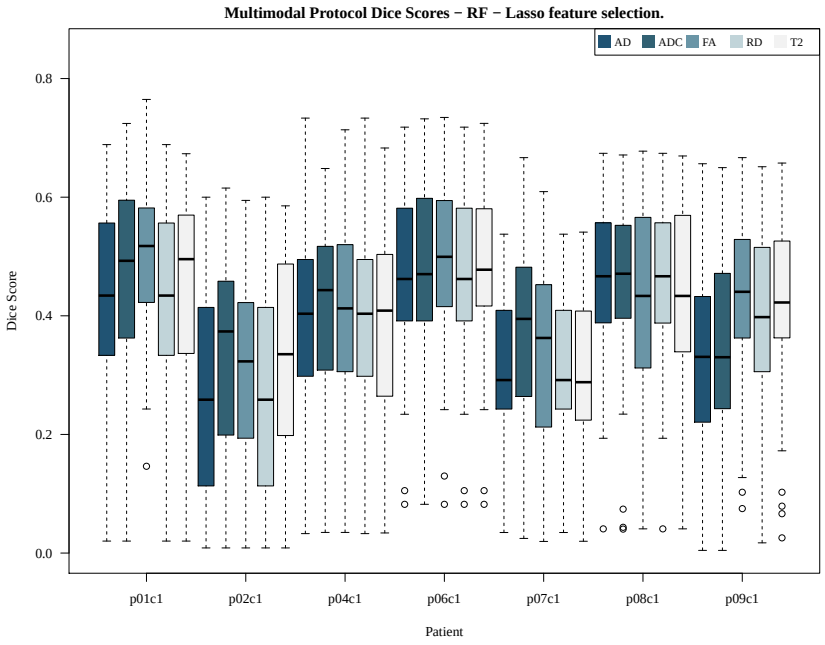


(a)

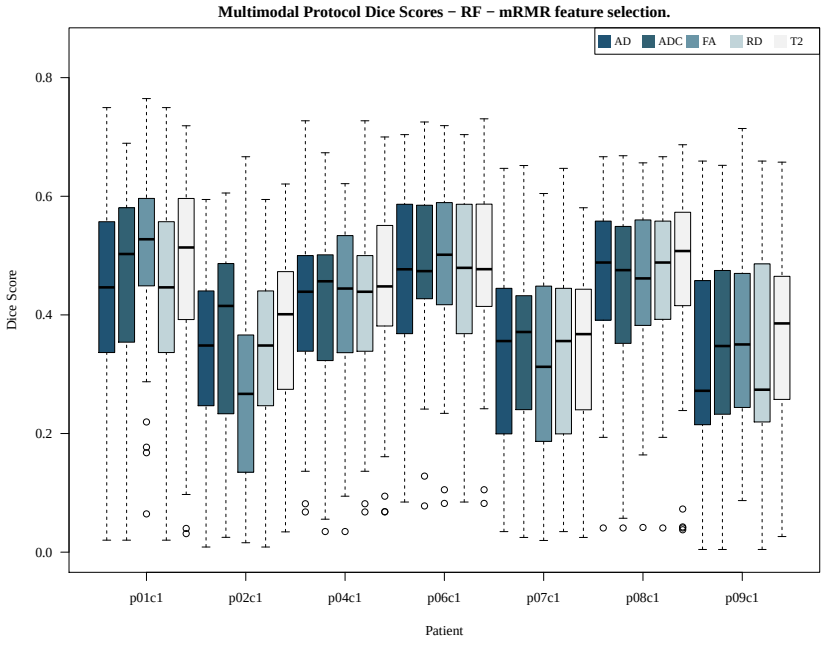


(b)

Figure 74: (a) Bar charts displaying binned distributions of slice accuracies for Multimodal SVM experiments using (a) LASSO feature selection and (b) mRMR feature selection methods. Red represents the number of slices within a volume scoring greater than 70.

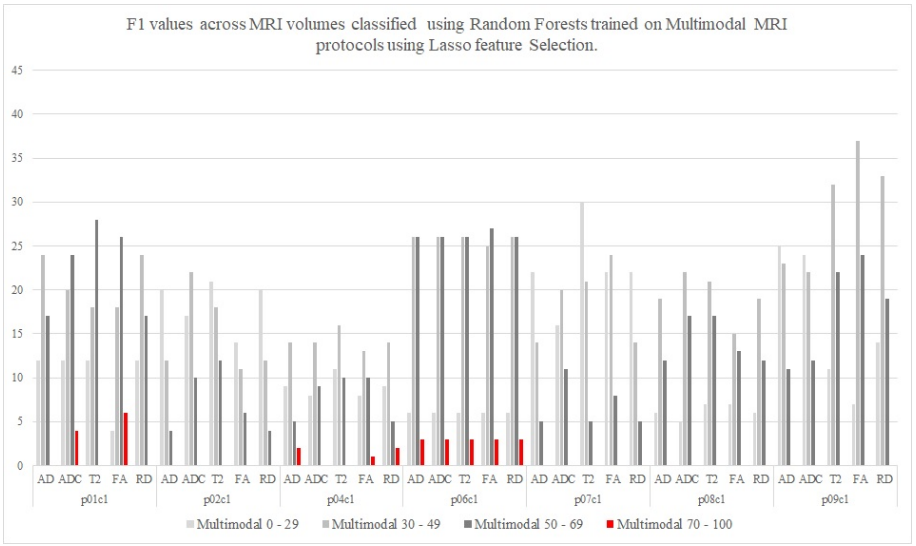


(a)

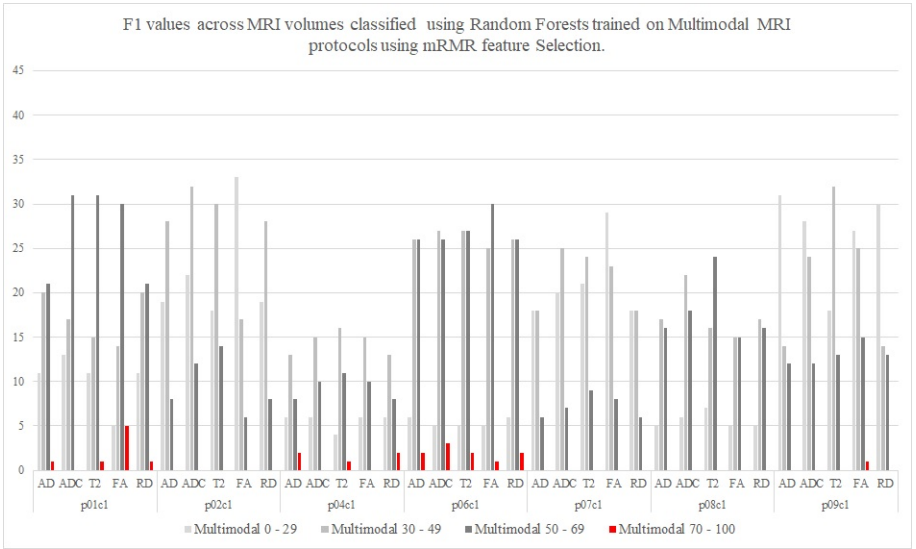


(b)

Figure 75: (a) Dice score performance for entire MRI volumes using Random Forest Models trained on Multimodal MRI data and LASSO Feature Selection. (b) As for (a) but using mRMR feature Selection.

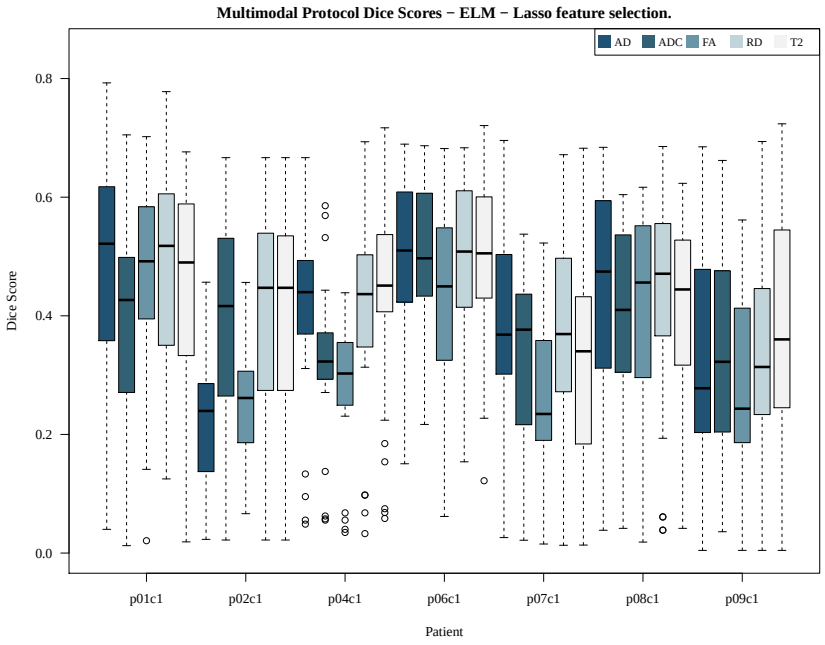


(a)

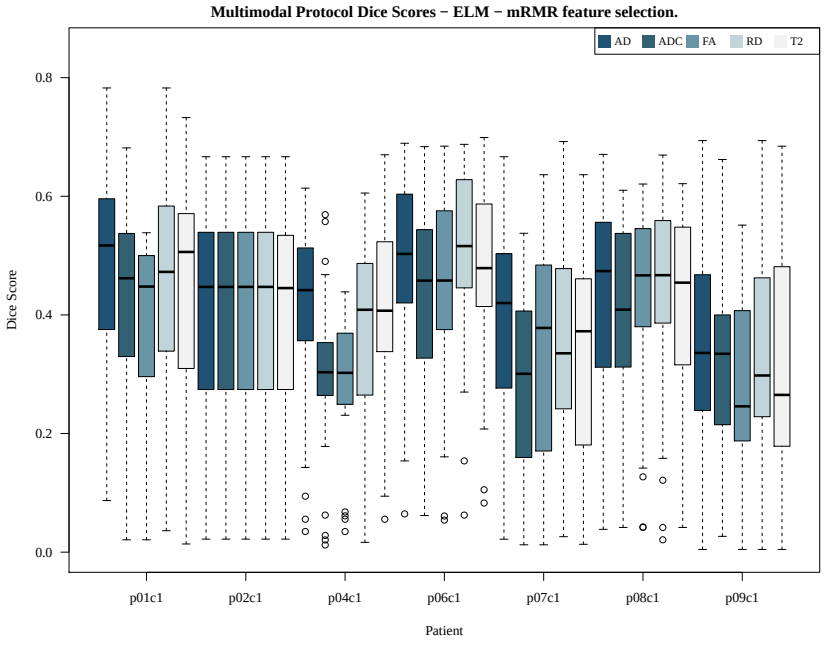


(b)

Figure 76: (a) Bar charts displaying binned distributions of slice accuracy for Multimodal RF experiments using (a) LASSO feature selection and (b) mRMR feature selection methods. Red represents the number of slices within a volume scoring greater than 70.

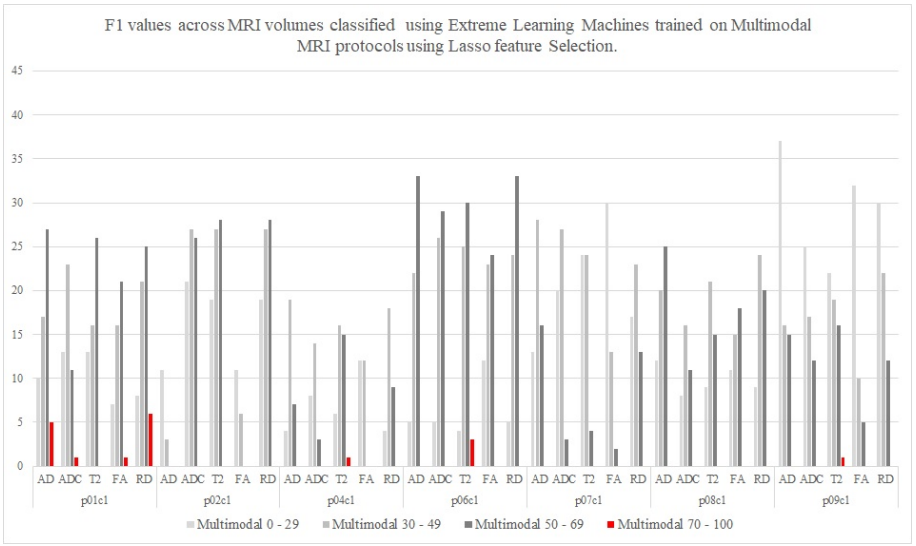


(a)

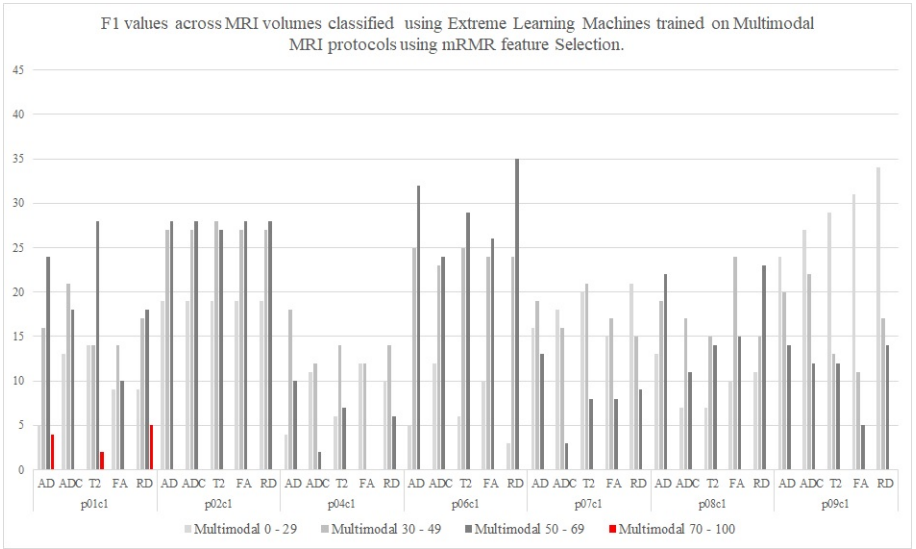


(b)

Figure 77: (a) Dice score performance for entire MRI volumes using Extreme Learning Machine Models trained on Multimodal MRI data and LASSO feature Selection.(b) As for (a) but using mRMR feature Selection.

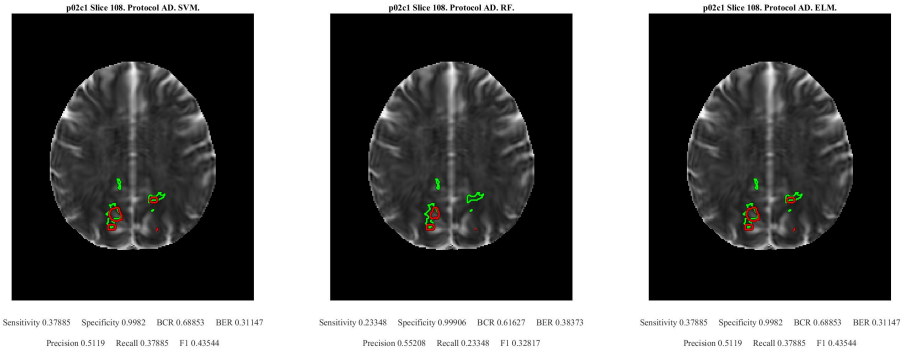


(a)

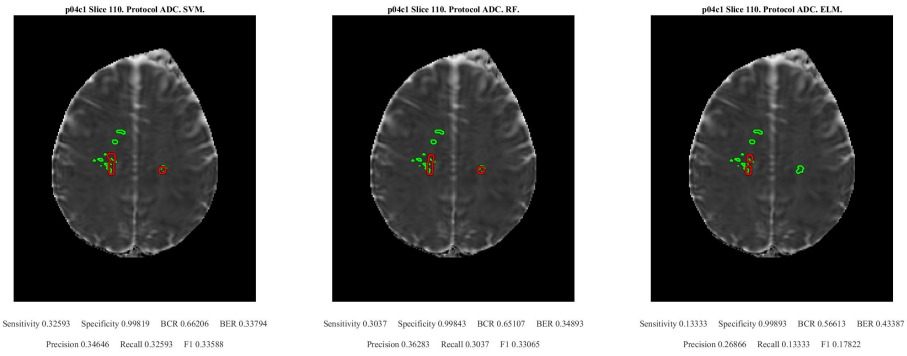


(b)

Figure 78: (a) Bar charts displaying binned distributions of slice accuracies for Multimodal ELM experiments using (a) LASSO feature selection and (b) mRMR feature selection methods. Red represents the number of slices within a volume scoring greater than 70.

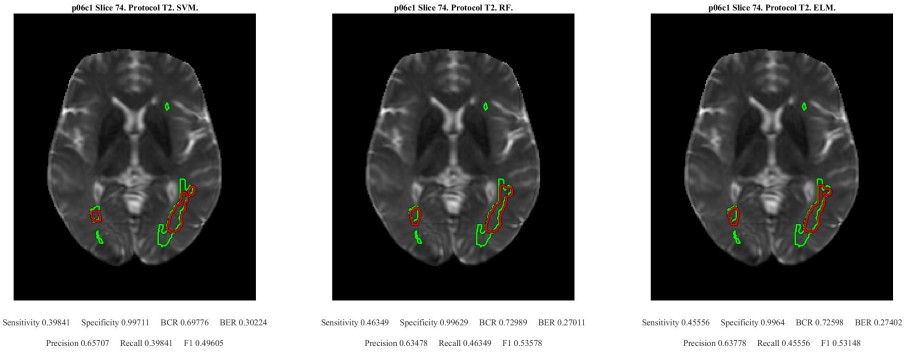


(a)

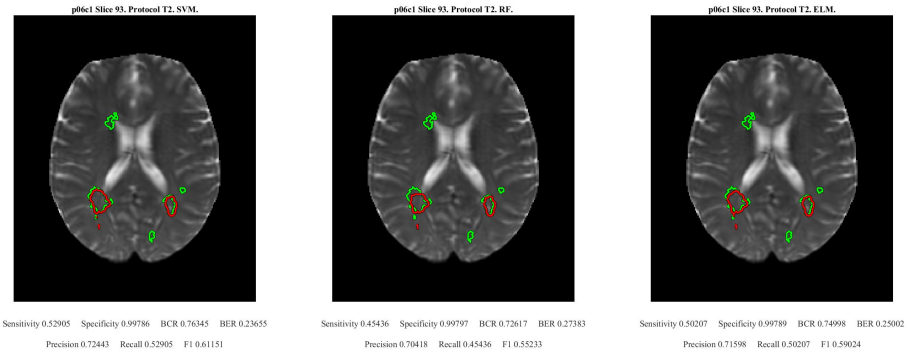


(b)

Figure 79: (a) Segmentations using all machine learning classifiers trained on Multimodal MRI data and features selected using mRMR feature selection for MS identification of dataset p02c1 AD sequence on slice 108. (b) as for (a) but identifying dataset p04c1 slice 110 with mRMR feature selection.



(a)



(b)

Figure 80: (a) Segmentations using all machine learning classifiers trained on Multimodal MRI data and features selected using mRMR feature selection for MS identification of dataset p06c1 T2 sequence on slice 74. (b) as for (a) but identifying dataset p06c1 slice 93 with LASSO feature selection.

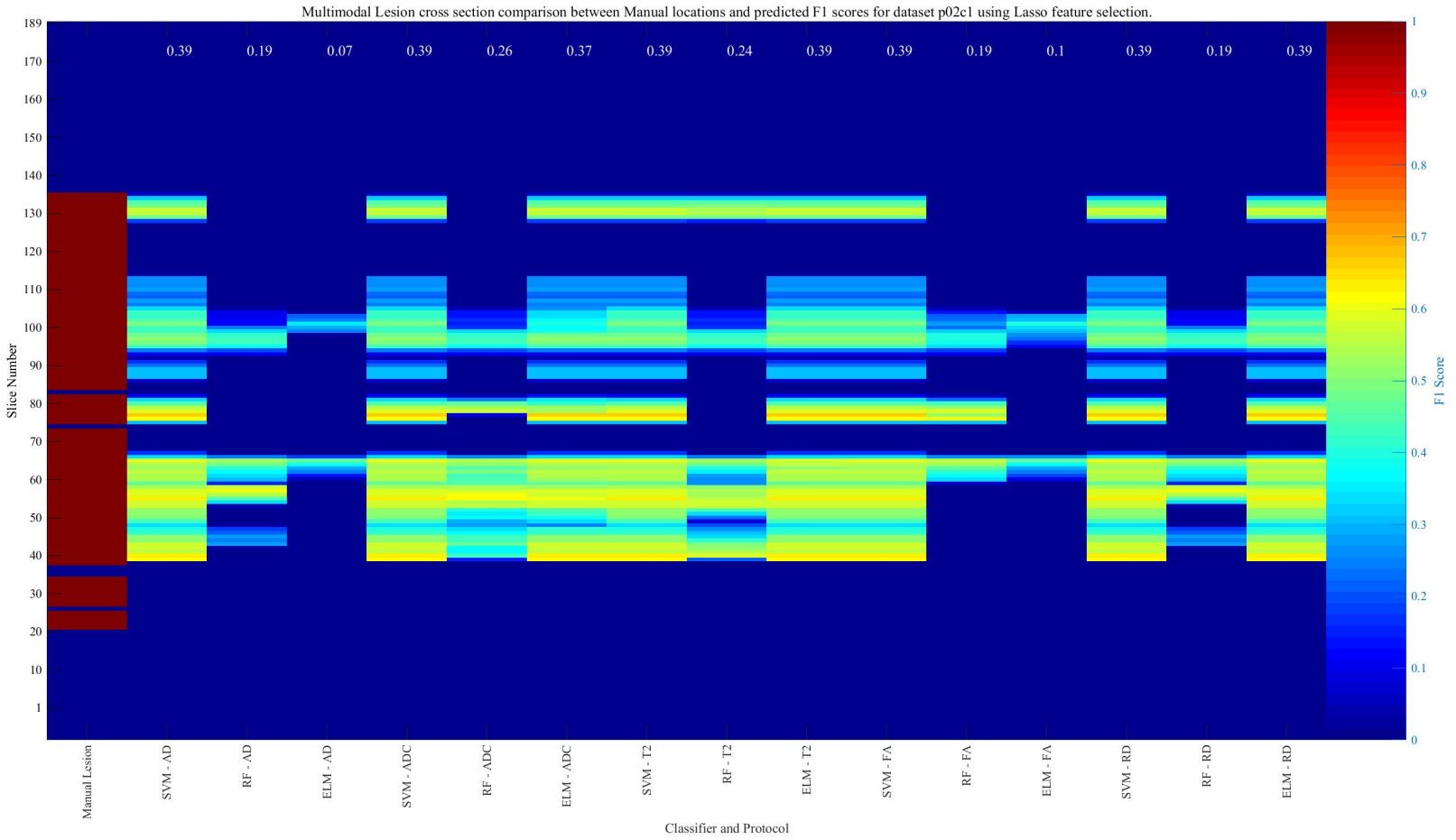


Figure 81: Cross sectional graph for p02c1 using Machine Learning models trained on Multimodal MRI data and LASSO feature selection.

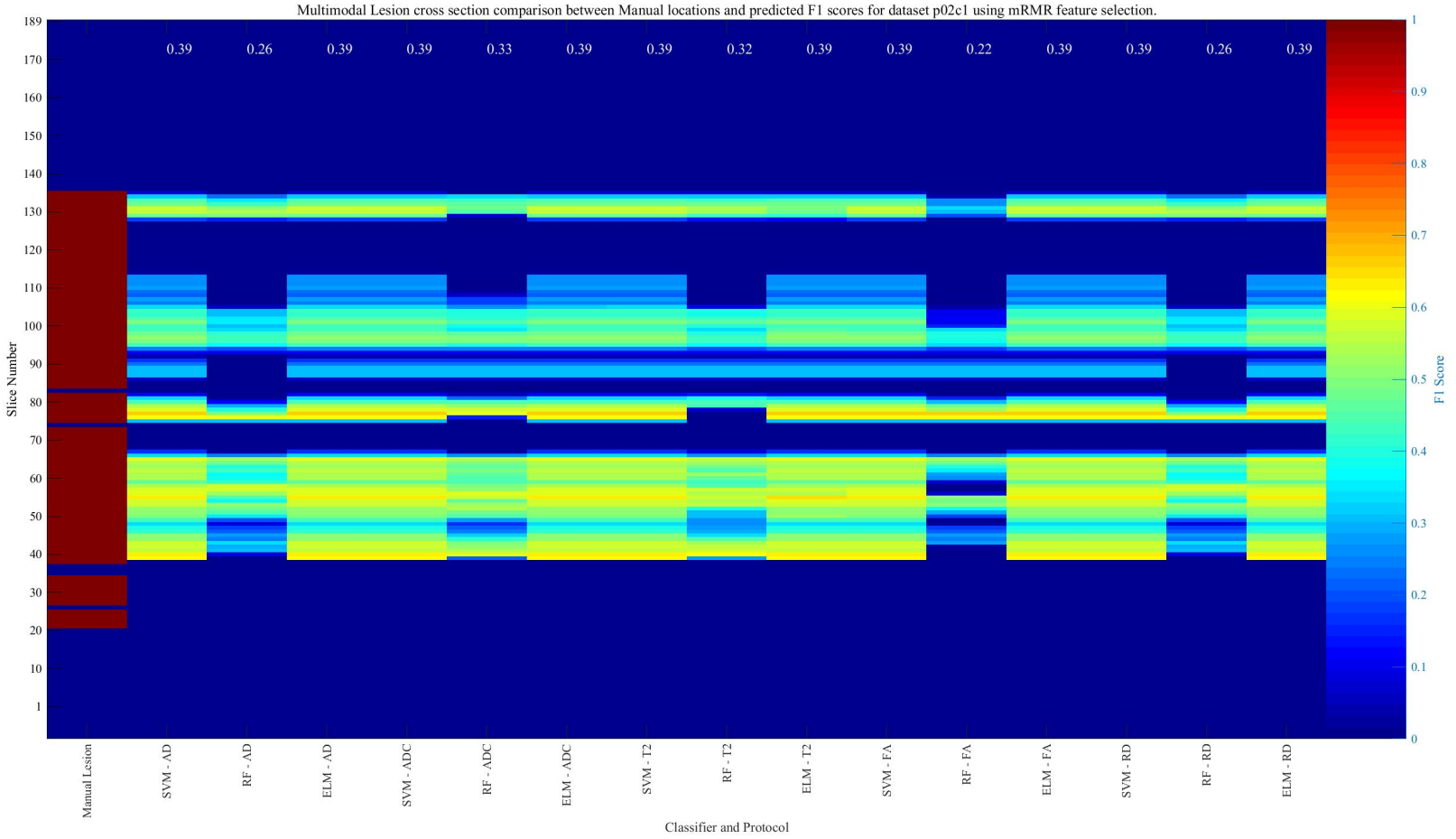
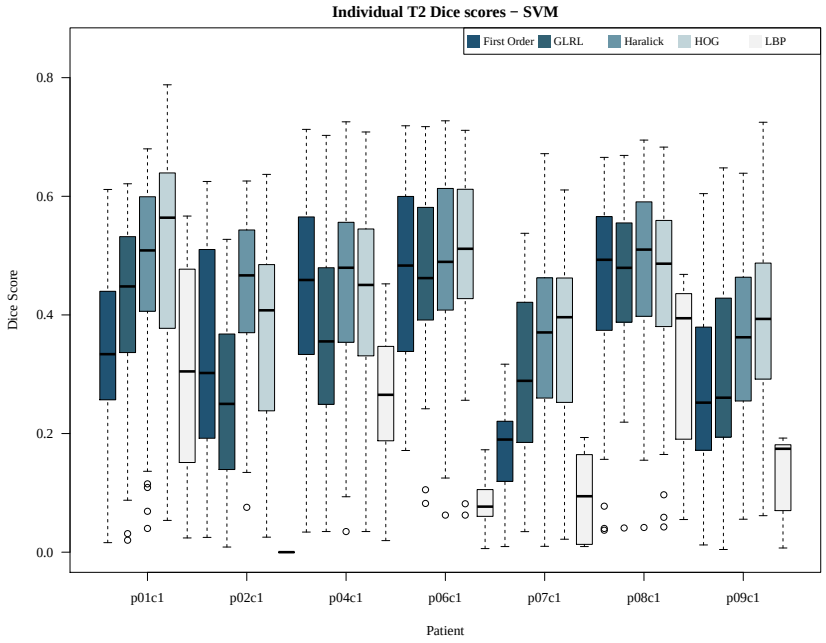
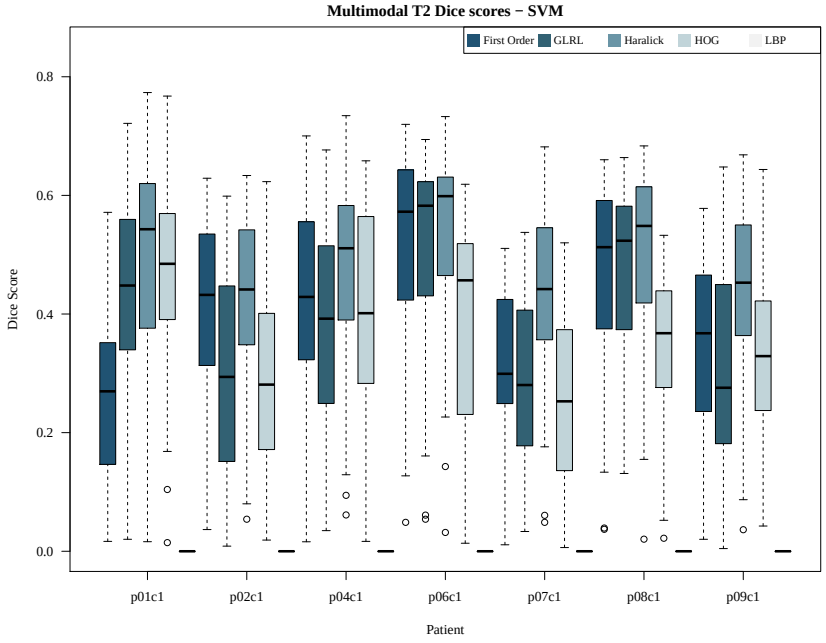


Figure 82: Cross sectional graph for p02c1 using Machine Learning models trained on Multimodal MRI data and mRMR feature selection.

C Results from Unseen Dataset Classification - Feature Groups

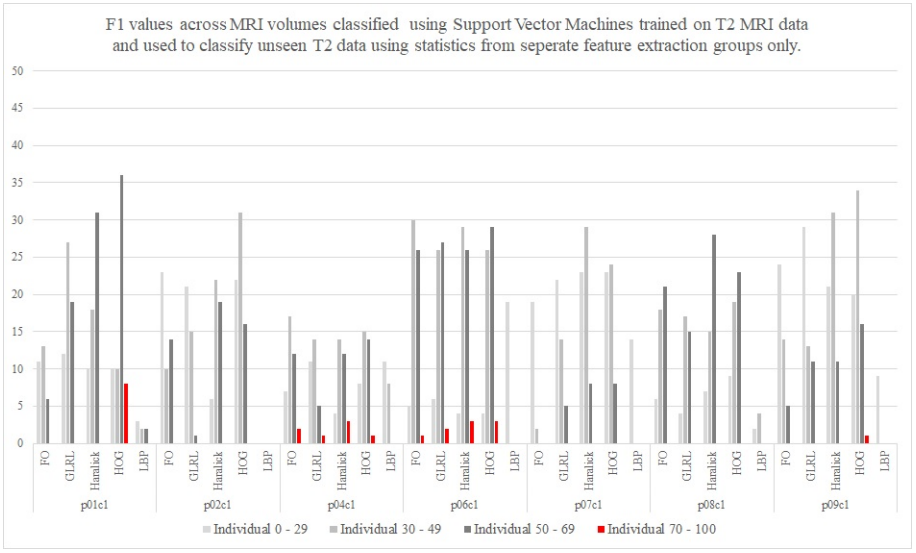


(a)

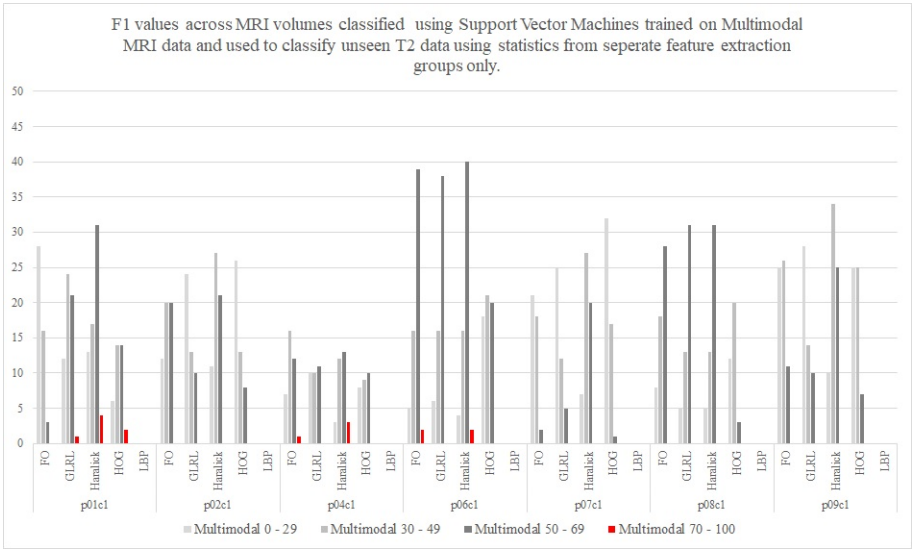


(b)

Figure 83: (a) Individual sequence Dice Scores using Support Vector Machine Classifier applied on Individual feature sets using leaveout patient datasets T2 MRI sequence. (b) As for (a) but using Multimodal MRI data in classifier training.

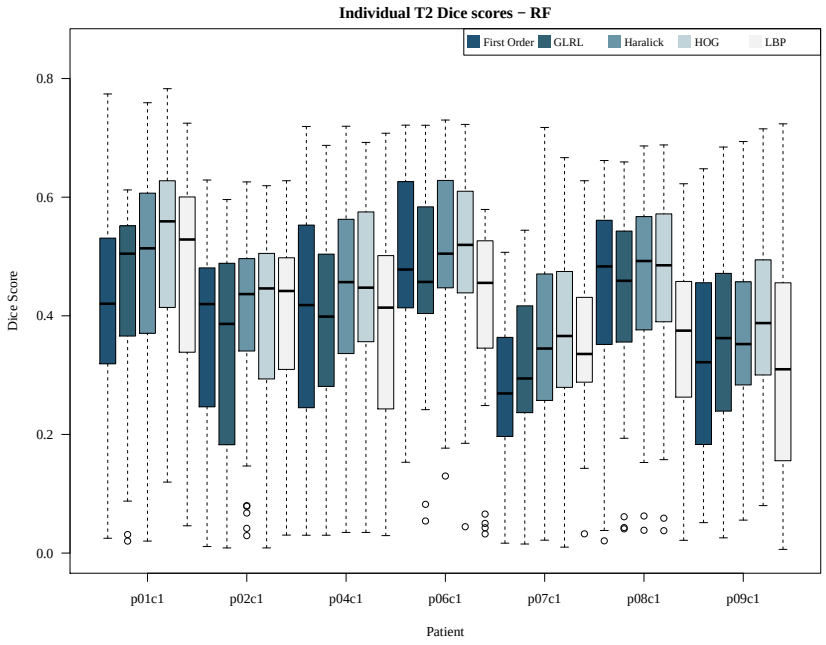


(a)

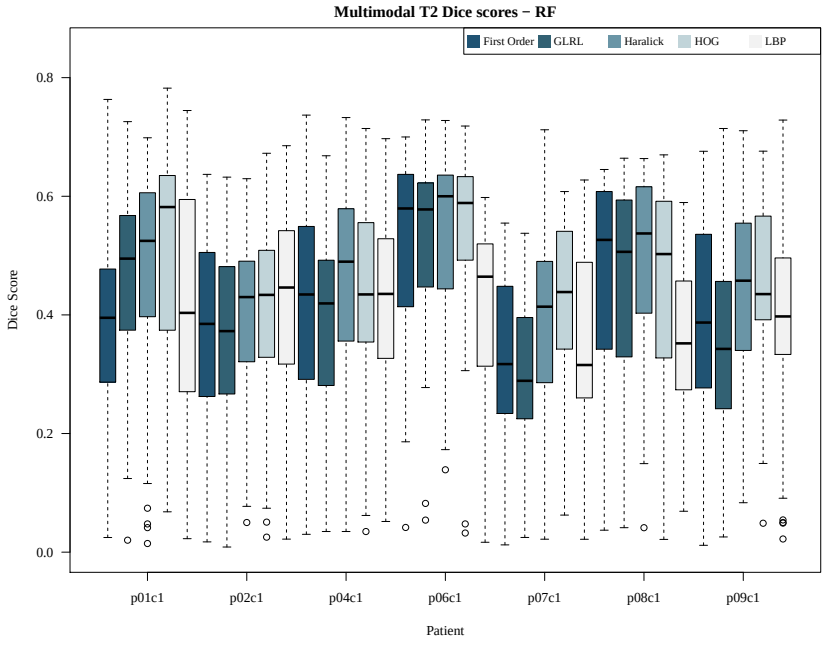


(b)

Figure 84: (a) Bar charts displaying binned distributions of slice accuracies for Individual SVM experiments using individual feature extraction techniques (b) as for a but showing results of SVM trained using Multimodal MRI data. Red represents the number of slices within a volume scoring greater than 70.

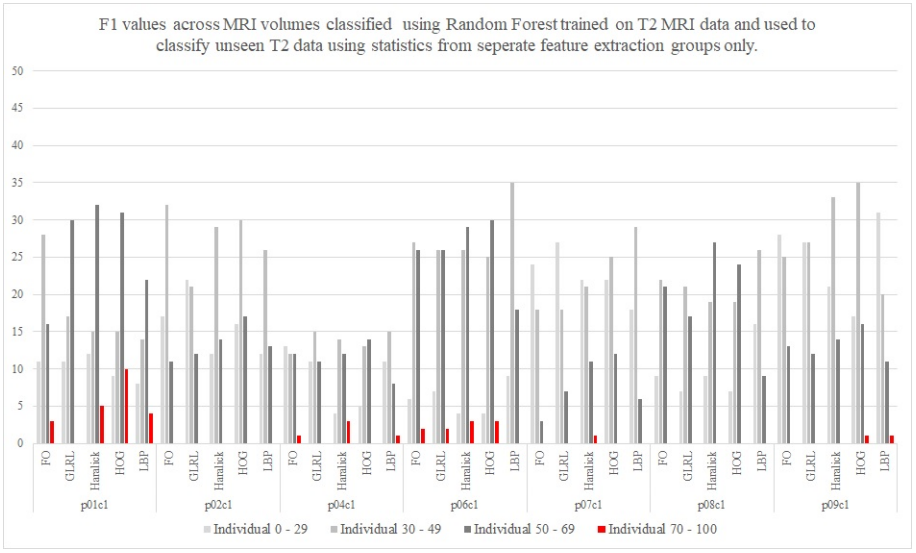


(a)

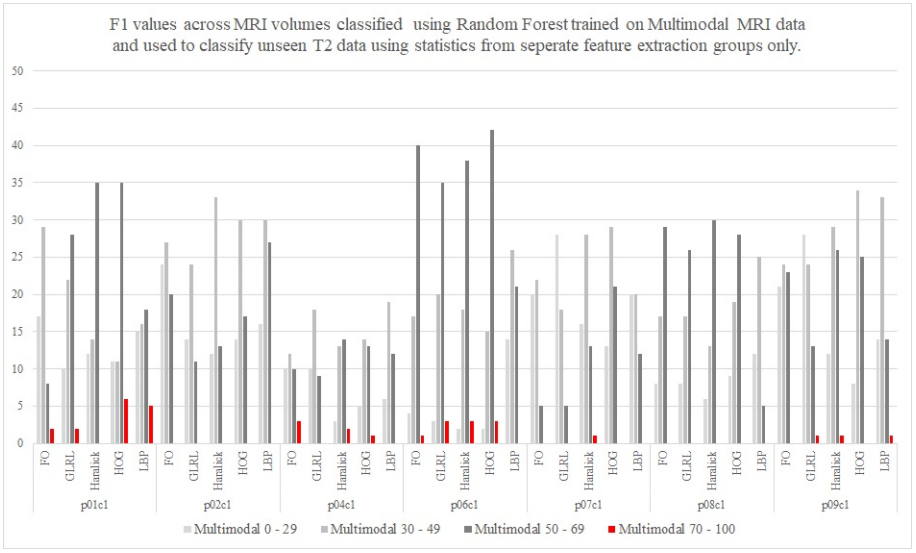


(b)

Figure 85: (a) Individual sequence Dice Scores using Random Forest Classifier applied on Individual feature sets using leaveout patient datasets T2 MRI sequence. (b) As for (a) but using Multimodal MRI data in classifier training.

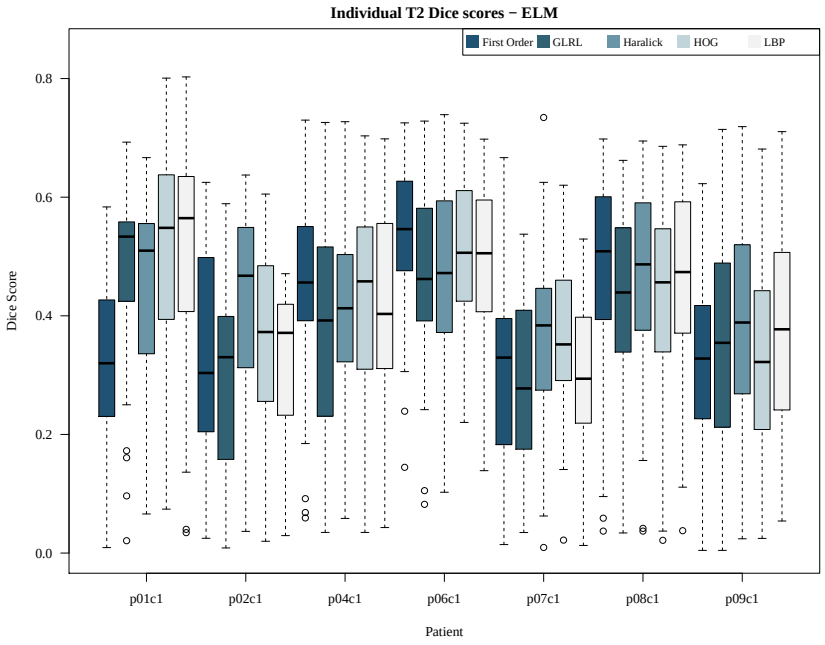


(a)

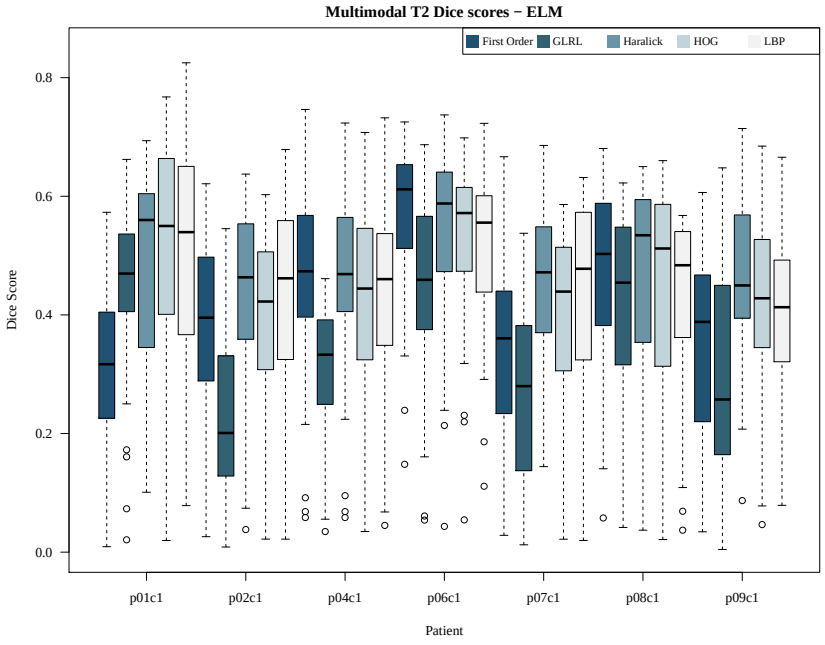


(b)

Figure 86: (a) Bar charts displaying binned distributions of slice accuracies for Individual RF experiments using individual feature extraction techniques (b) as for a but showing results of RF trained using Multimodal MRI data. Red represents the number of slices within a volume scoring greater than 70.

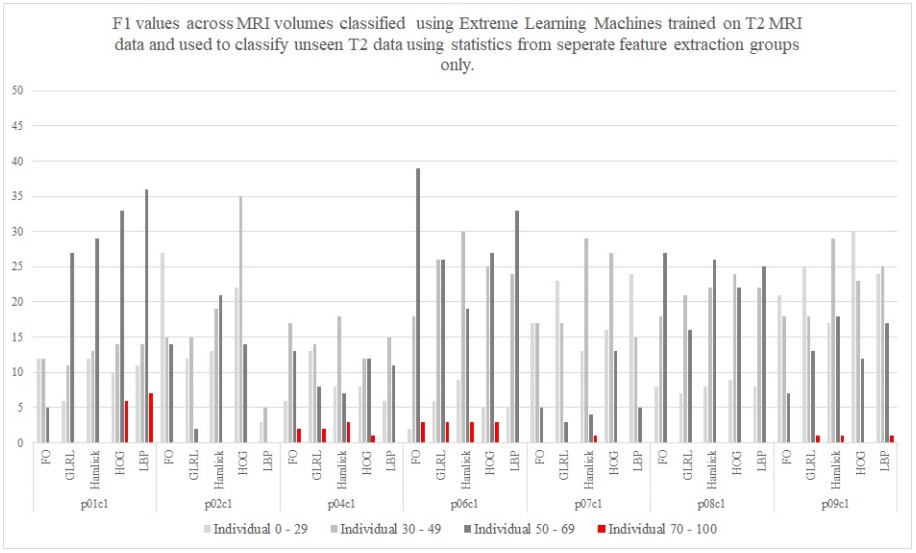


(a)

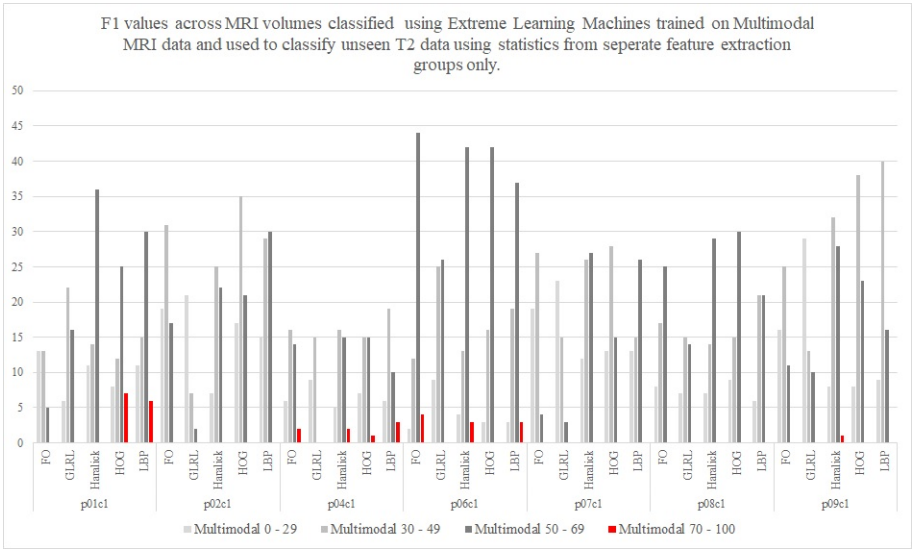


(b)

Figure 87: (a) Individual sequence Dice Scores using Extreme Learning Machine Classifier applied on Individual feature sets using leaveout patient datasets T2 MRI sequence. (b) As for (a) but using Multimodal MRI data in classifier training.

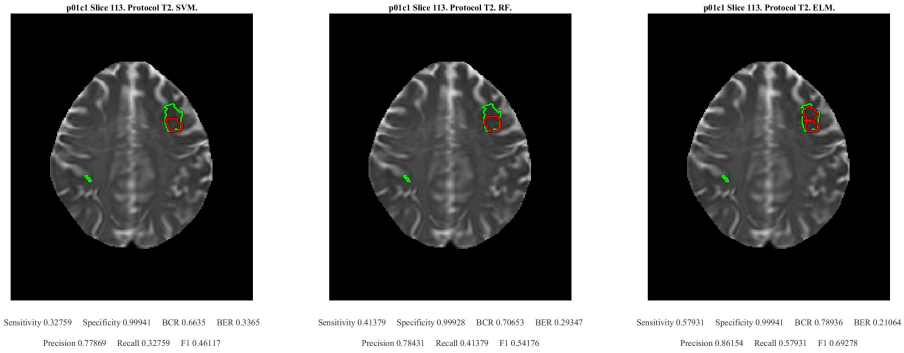


(a)

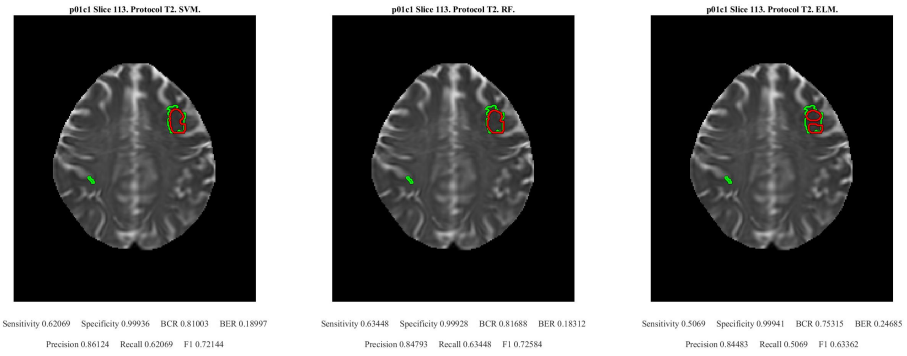


(b)

Figure 88: (a) Bar charts displaying binned distributions of slice accuracies for Individual ELM experiments using individual feature extraction techniques (b) as for a but showing results of ELM trained using Multimodal MRI data. Red represents the number of slices within a volume scoring greater than 70.

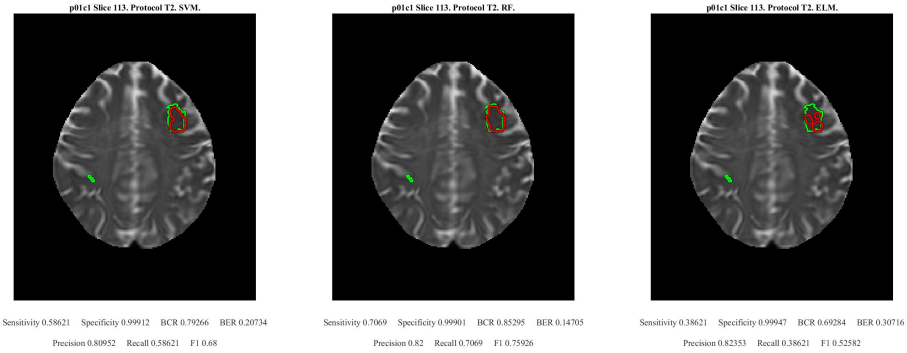


(a)

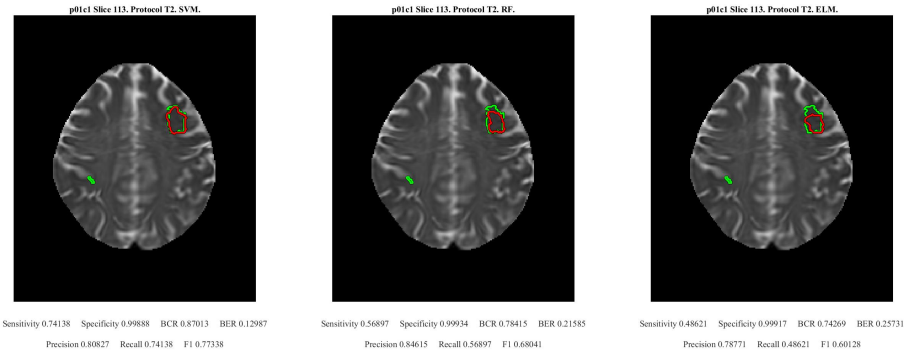


(b)

Figure 89: (a) Individual sequence Segmented regions using all classifiers trained using only Gray Level Run Length statistics from T2 MRI sequence for unseen segmentation using dataset p01c1 T2 sequence on slice 113. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.

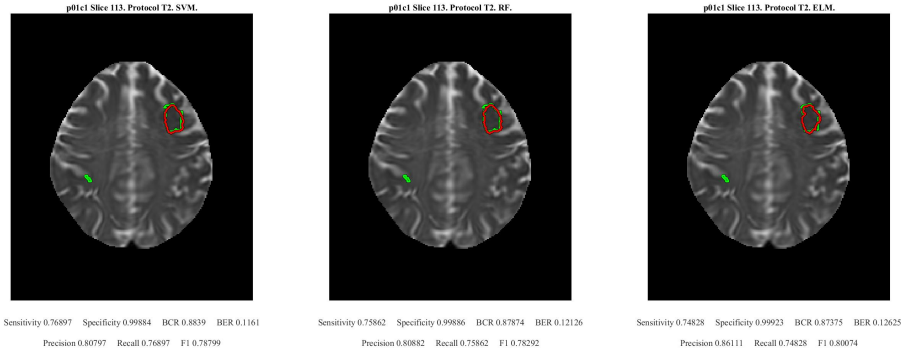


(a)

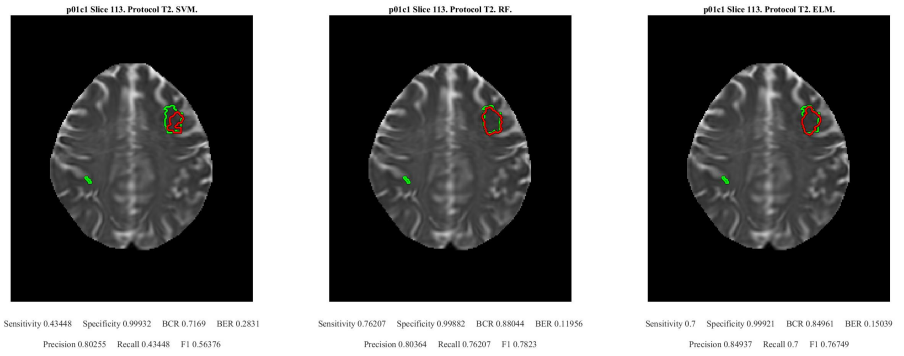


(b)

Figure 90: (a) Individual sequence Segmented regions using all classifiers trained using only Haralick Texture statistics from T2 MRI sequence for unseen segmentation using dataset p01c1 T2 sequence on slice 113. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.

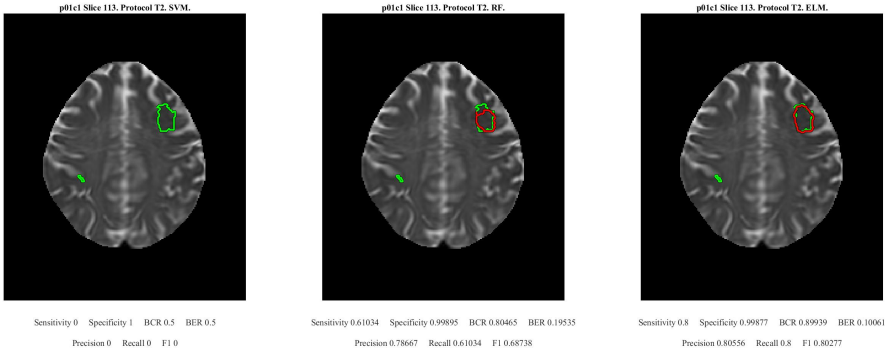


(a)

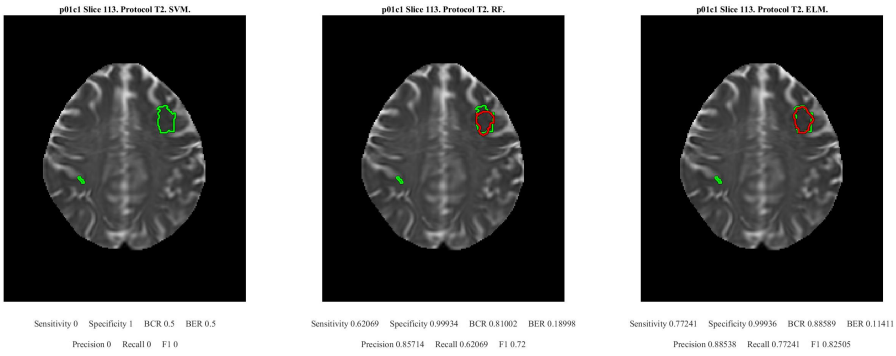


(b)

Figure 91: (a) Individual sequence Segmented regions using all classifiers trained on Histogram of Oriented Gradient statistics from T2 MRI sequence for unseen segmentation using dataset p01c1 T2 sequence on slice 113. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.

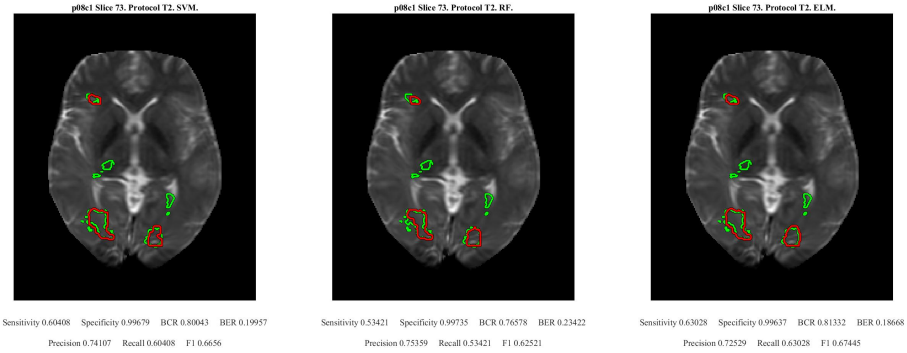


(a)

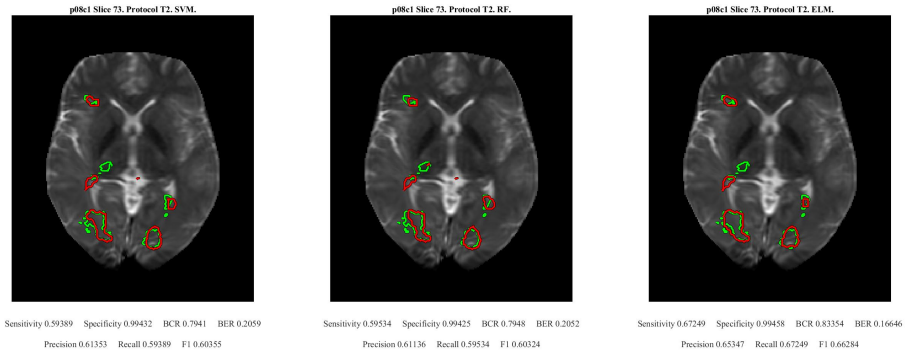


(b)

Figure 92: (a) Individual sequence Segmented regions using all classifiers trained using only Local Binary Pattern statistics from T2 MRI sequence for unseen segmentation using dataset p01c1 T2 sequence on slice 113. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.

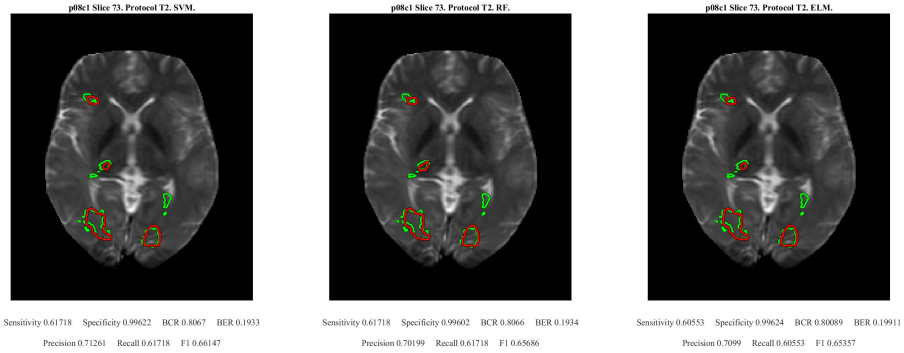


(a)

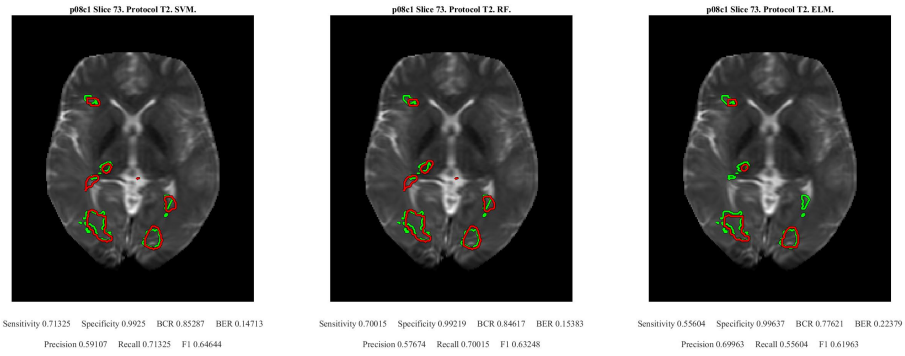


(b)

Figure 93: (a) Individual sequence Segmented regions using all classifiers trained using only first-order statistics from T2 MRI sequence for unseen segmentation using dataset p08c1 T2 sequence on slice 73. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.

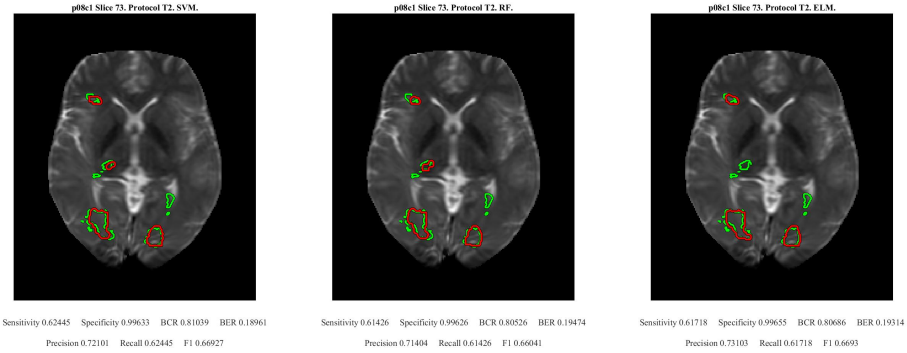


(a)

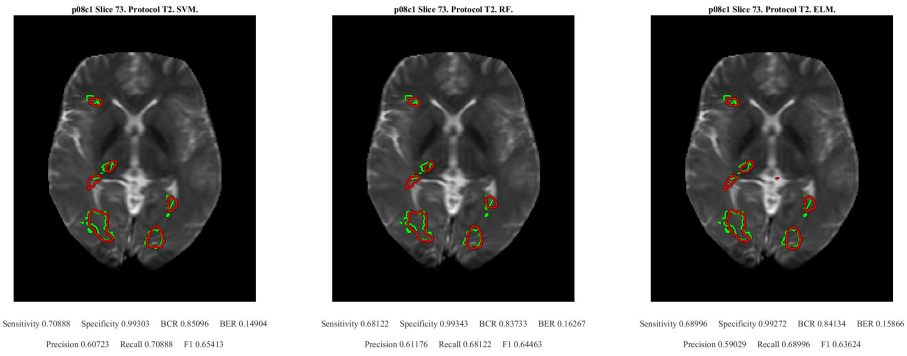


(b)

Figure 94: (a) Individual sequence Segmented regions using all classifiers trained using only Gray Level Run Length statistics from T2 MRI sequence for unseen segmentation using dataset p08c1 T2 sequence on slice 73. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.

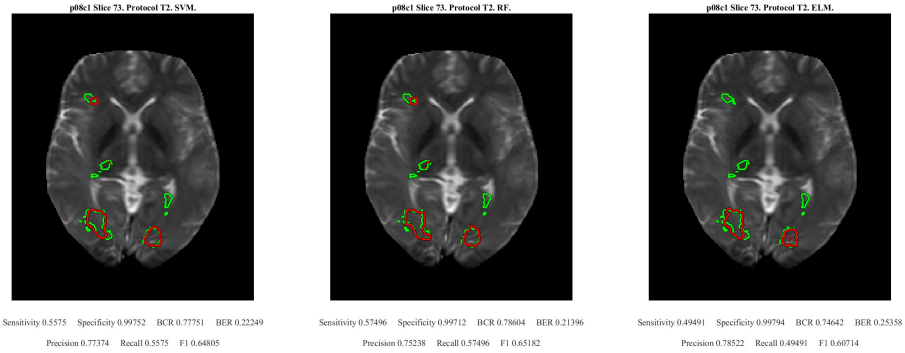


(a)

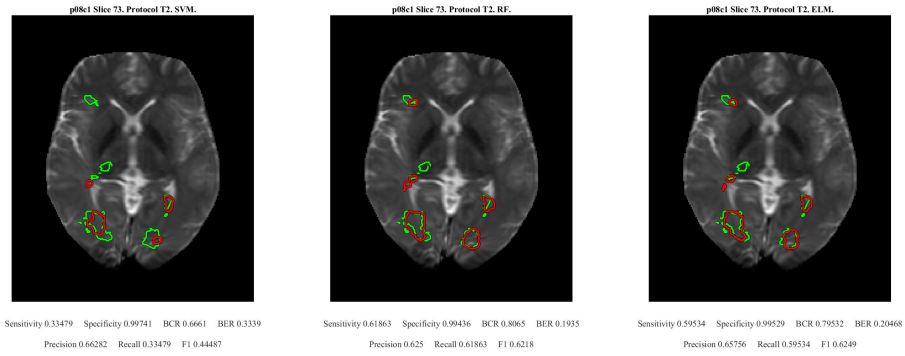


(b)

Figure 95: (a) Individual sequence Segmented regions using all classifiers trained using only Haralick Texture statistics from T2 MRI sequence for unseen segmentation using dataset p08c1 T2 sequence on slice 73. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.

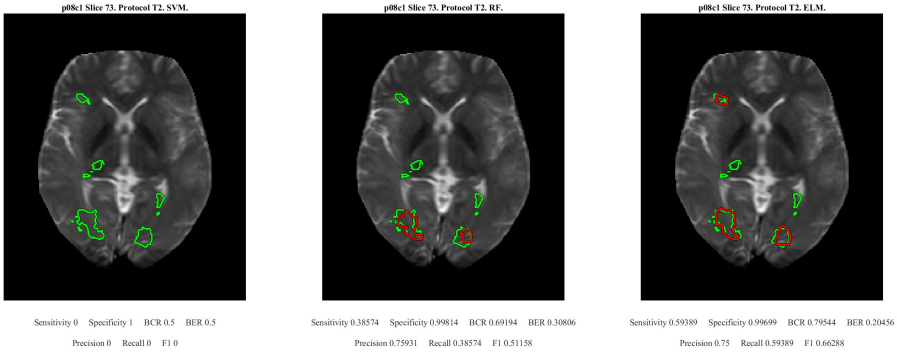


(a)

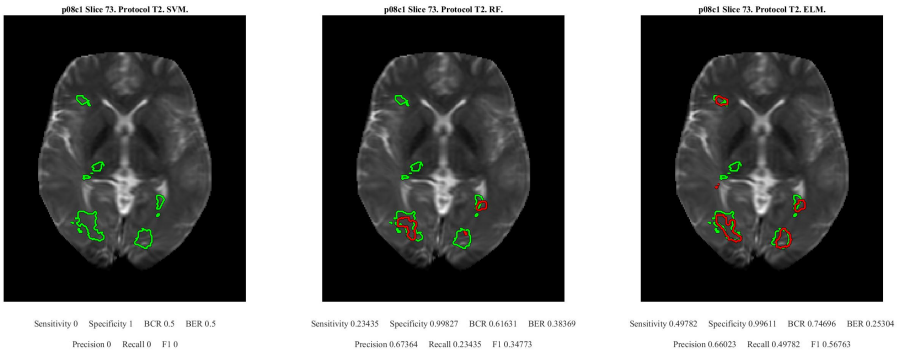


(b)

Figure 96: (a) Individual sequence Segmented regions using all classifiers trained on Histogram of Oriented Gradient statistics from T2 MRI sequence for unseen segmentation using dataset p08c1 T2 sequence on slice 73. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.



(a)



(b)

Figure 97: (a) Individual sequence Segmented regions using all classifiers trained using only Local Binary Pattern statistics from T2 MRI sequence for unseen segmentation using dataset p08c1 T2 sequence on slice 73. (b) depicts the same slice and dataset but using machine learning models trained on Multimodal MRI data.

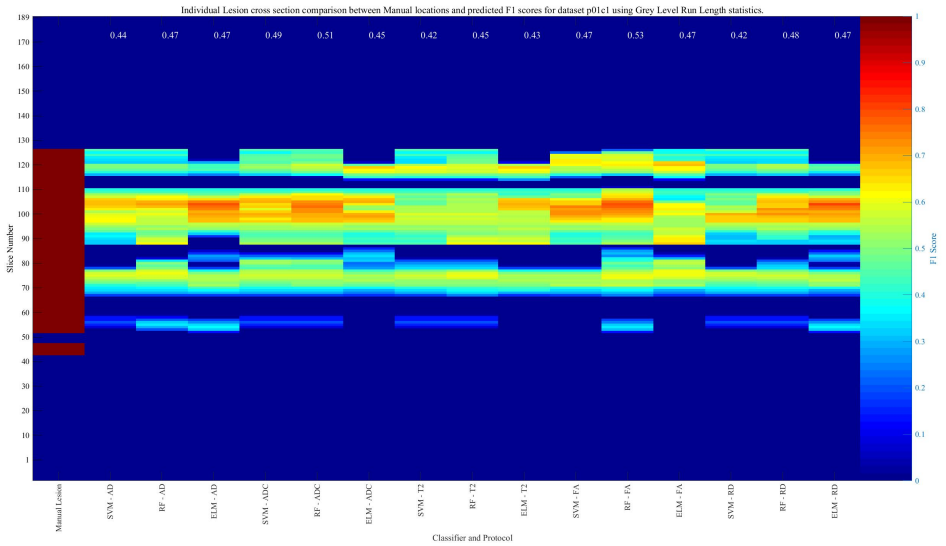


Figure 98: Cross sectional graph for p01c1 using Machine Learning models trained on Individual sequence Grey Level Run Length MRI data.

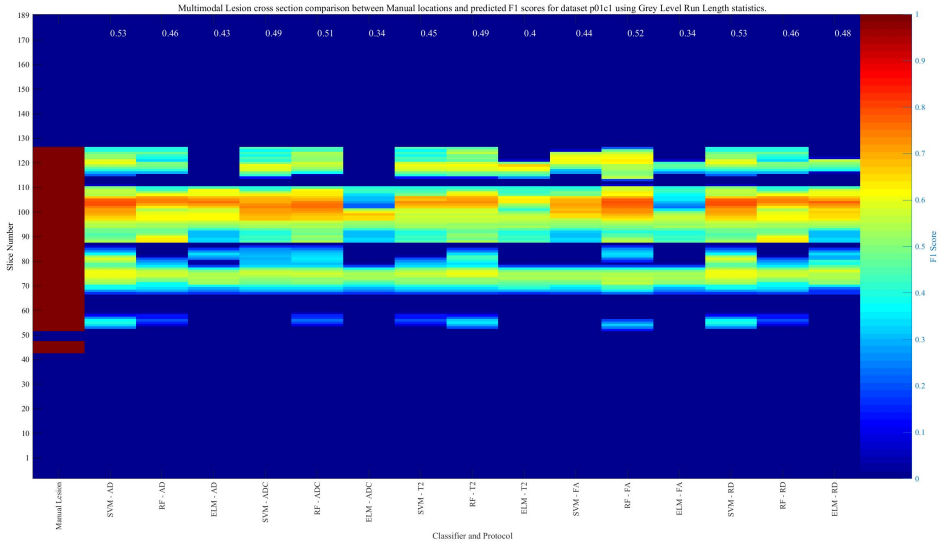


Figure 99: Cross sectional graph for p01c1 using Machine Learning models trained on Multimodal Grey Level Run Length.

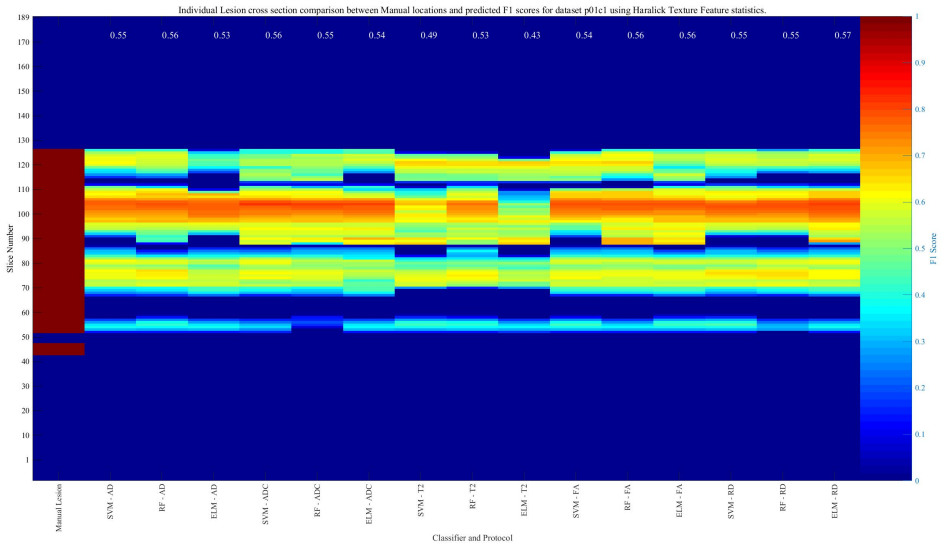


Figure 100: Cross sectional graph for p01c1 using Machine Learning models trained on Individual sequence Haralick Texture Feature MRI data.

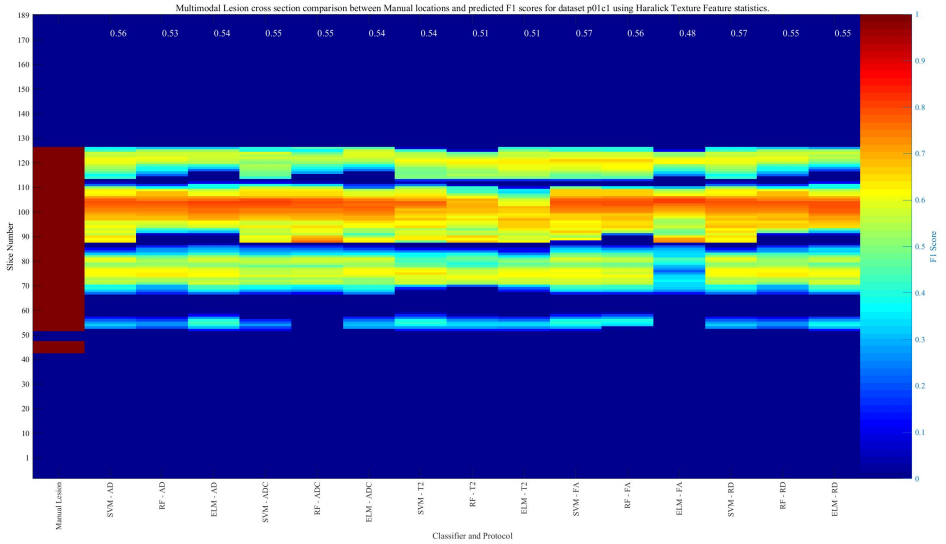


Figure 101: Cross sectional graph for p01c1 using Machine Learning models trained on Multimodal Haralick Texture Feature MRI data.

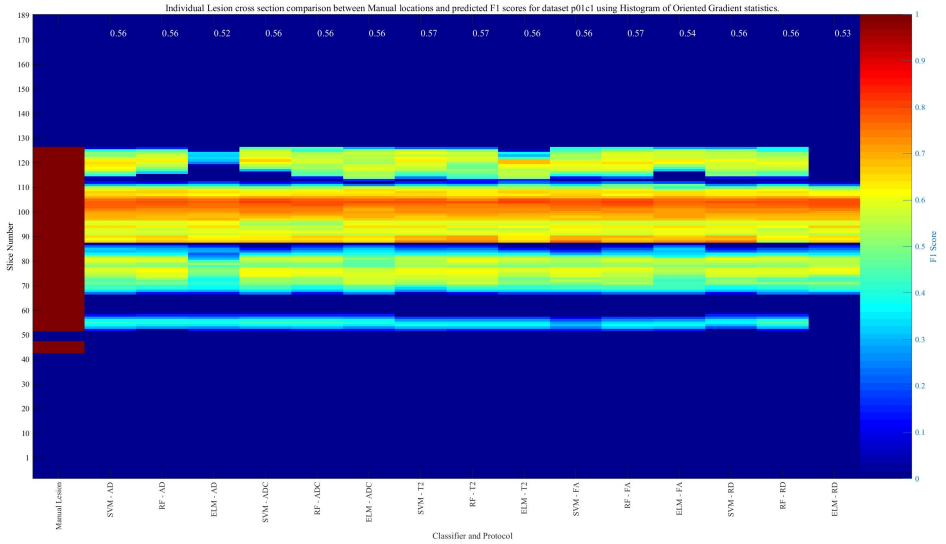


Figure 102: Cross sectional graph for p01c1 using Machine Learning models trained on Individual sequence Histogram of Oriented Gradient MRI data.

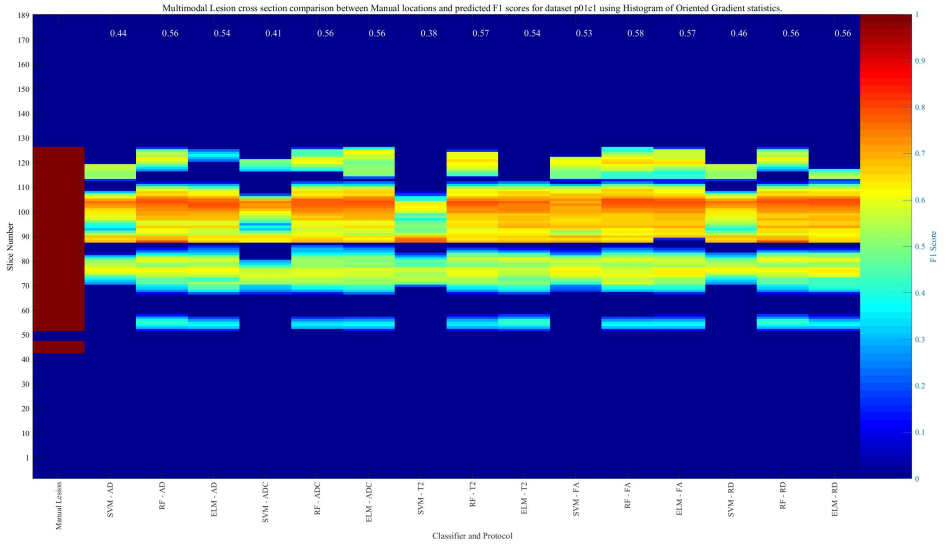


Figure 103: Cross sectional graph for p01c1 using Machine Learning models trained on Multimodal Histogram of Oriented Gradient MRI data.

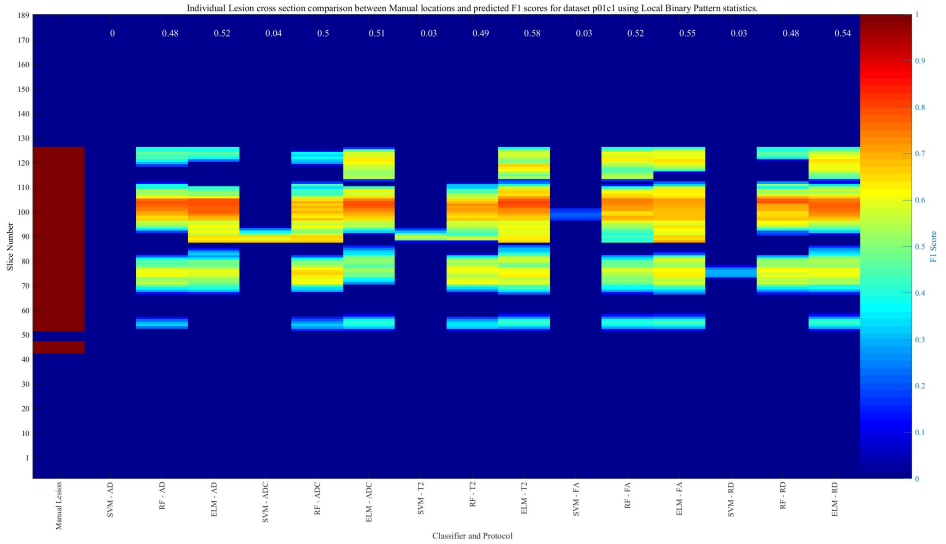


Figure 104: Cross sectional graph for p01c1 using Machine Learning models trained on Individual sequence Local Binary Pattern MRI data.

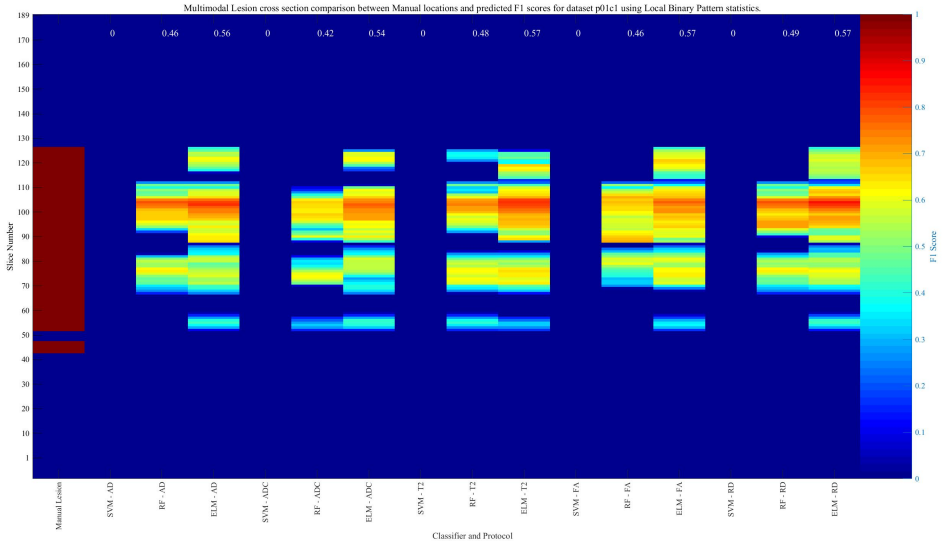


Figure 105: Cross sectional graph for p01c1 using Machine Learning models trained on Multimodal Local Binary Pattern MRI data.

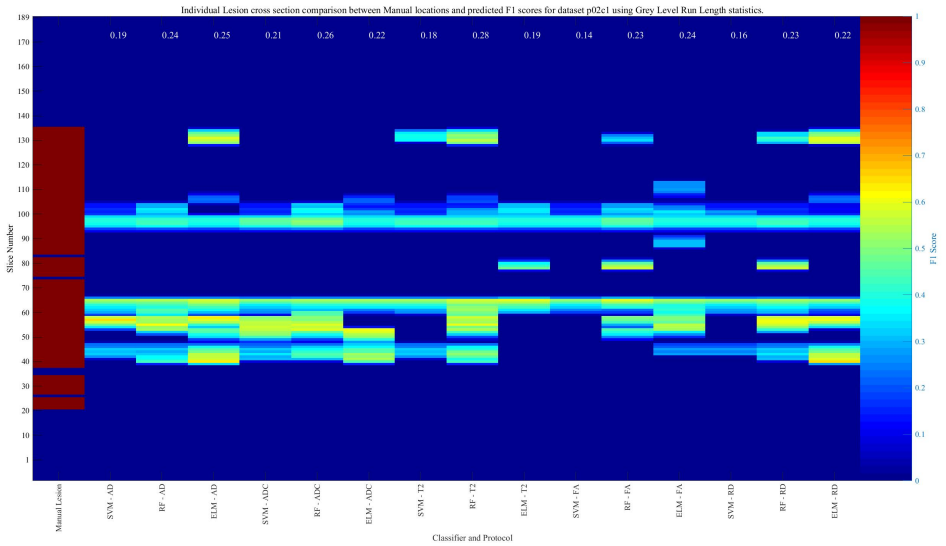


Figure 106: Cross sectional graph for p02c1 using Machine Learning models trained on Individual sequence Grey Level Run Length MRI data.

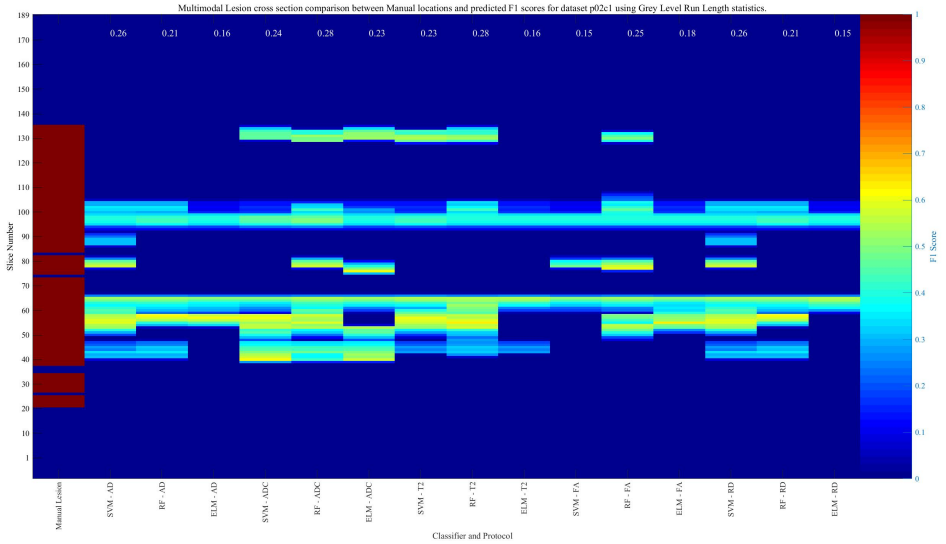


Figure 107: Cross sectional graph for p02c1 using Machine Learning models trained on Multimodal Grey Level Run Length MRI data.

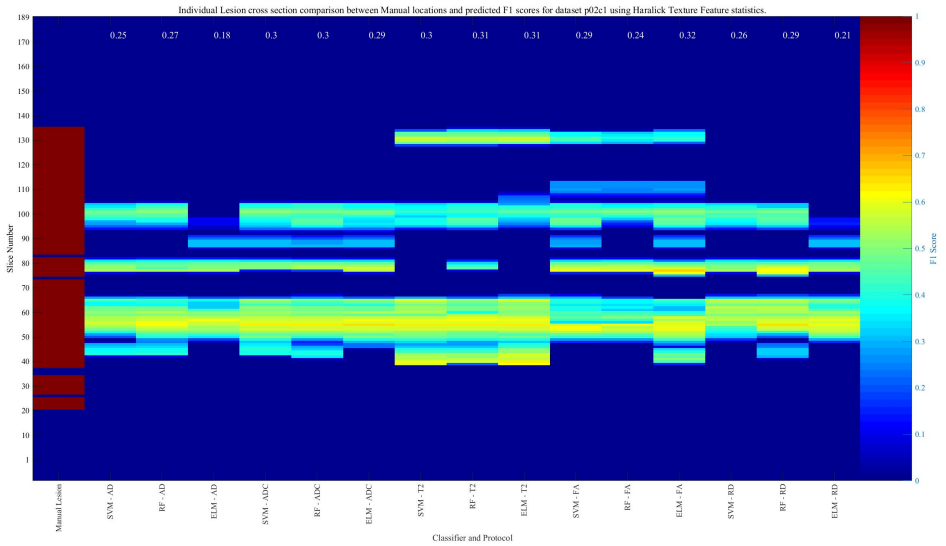


Figure 108: Cross sectional graph for p02c1 using Machine Learning models trained on Individual sequence Haralick Texture MRI data.

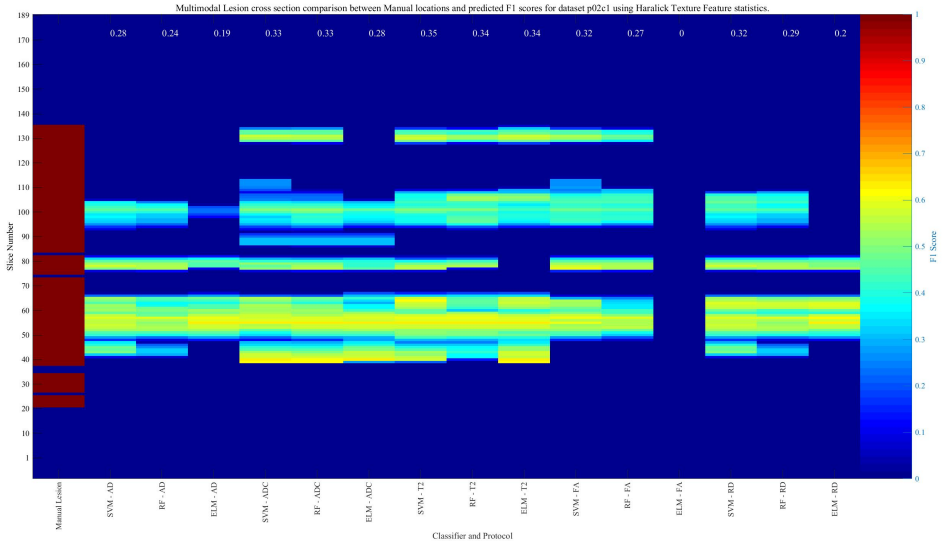


Figure 109: Cross sectional graph for p02c1 using Machine Learning models trained on Multimodal Haralick Texture MRI data.

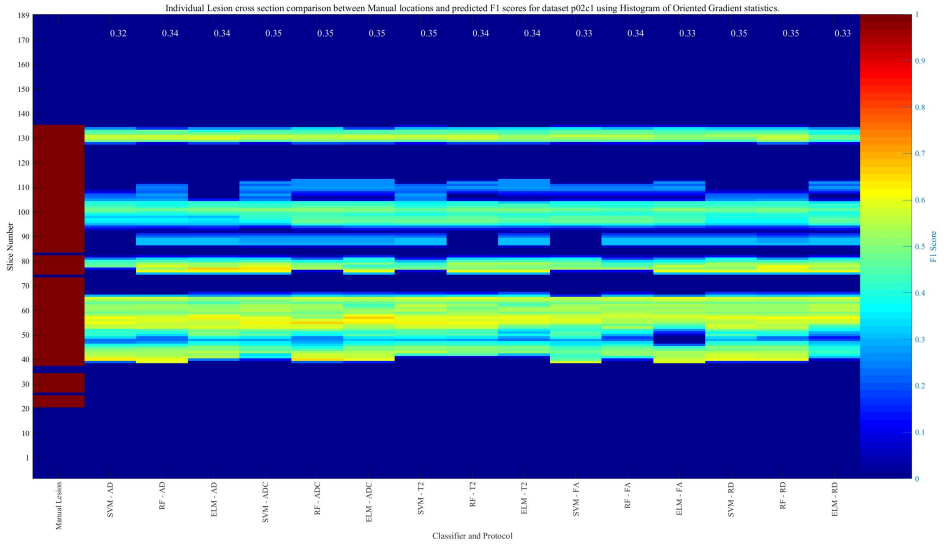


Figure 110: Cross sectional graph for p02c1 using Machine Learning models trained on Individual sequence Histogram of Oriented Gradient Texture MRI data.

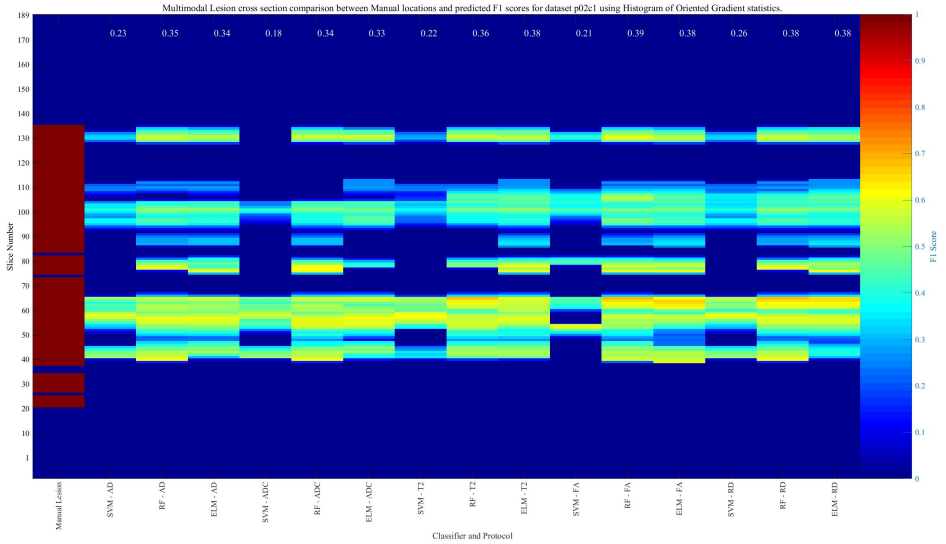


Figure 111: Cross sectional graph for p02c1 using Machine Learning models trained on Multimodal Histogram of Oriented Gradient Texture MRI data.

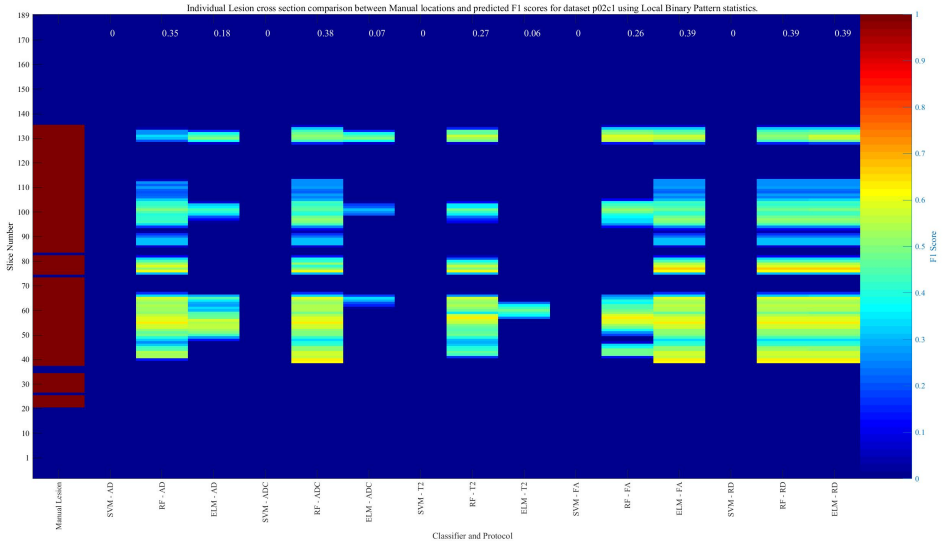


Figure 112: Cross sectional graph for p02c1 using Machine Learning models trained on Individual sequence Local Binary Pattern MRI data.

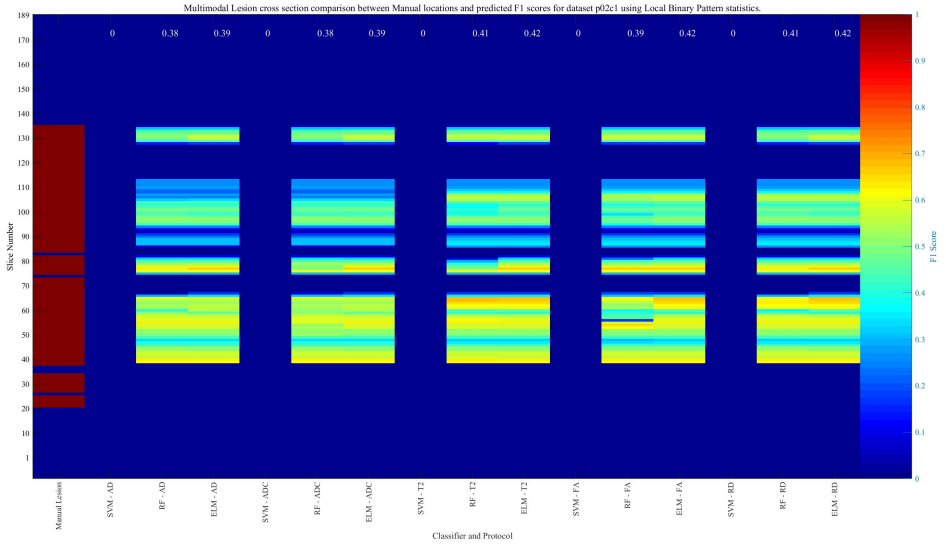


Figure 113: Cross sectional graph for p02c1 using Machine Learning models trained on Multimodal Local Binary Pattern MRI data.

D Results from DWI protocols

Table 7: Results table for each Feature Selection technique and Machine Learning Model using AD protocol data.

Method	Model	Data	F1	Sensitivity	Specificity	BER	Precision	Recall
Lasso	SVM	Individual	39.55 ± 8.9	61.73 ± 4.44	99.89 ± 0.04	19.19 ± 2.21	29.87 ± 9.44	61.73 ± 4.44
		Multimodal	44.82 ± 7.36	61.17 ± 3.32	99.89 ± 0.04	19.47 ± 1.65	36.03 ± 9.1	61.17 ± 3.32
	RF	Individual	39.59 ± 10.06	61.77 ± 3.56	99.89 ± 0.04	19.17 ± 1.77	30.06 ± 10.7	61.77 ± 3.56
		Multimodal	36.17 ± 11.44	60.99 ± 3.08	99.88 ± 0.04	19.56 ± 1.54	26.78 ± 11.24	60.99 ± 3.08
	ELM	Individual	42.16 ± 7.14	61.18 ± 2.75	99.89 ± 0.04	19.47 ± 1.37	32.66 ± 7.89	61.18 ± 2.75
		Multimodal	38.69 ± 15.93	63.33 ± 4.35	99.89 ± 0.04	18.39 ± 2.17	30.33 ± 14.79	63.33 ± 4.35
mRMR	SVM	Individual	34.67 ± 17.17	53.21 ± 23.6	99.88 ± 0.04	16.3 ± 7.3	26.23 ± 14.21	53.21 ± 23.6
		Multimodal	42.99 ± 7.23	61.39 ± 3.61	99.89 ± 0.04	19.36 ± 1.79	33.65 ± 8.4	61.39 ± 3.61
	RF	Individual	38.41 ± 11.63	60.96 ± 2.81	99.88 ± 0.04	19.58 ± 1.4	29.29 ± 12.02	60.96 ± 2.81
		Multimodal	38 ± 10.01	61.84 ± 2.21	99.88 ± 0.04	19.14 ± 1.1	28.32 ± 10.42	61.84 ± 2.21
	ELM	Individual	41.21 ± 6.4	61.95 ± 3.02	99.89 ± 0.03	19.08 ± 1.51	31.37 ± 7.42	61.95 ± 3.02
		Multimodal	42.27 ± 7.16	60.76 ± 3.49	99.89 ± 0.04	19.67 ± 1.73	33.01 ± 8.29	60.76 ± 3.49

Table 8: Results table for each Feature Selection technique and Machine Learning Model using ADC protocol data.

Method	Model	Data	F1	Sensitivity	Specificity	BER	Precision	Recall
Lasso	SVM	Individual	40.3 ± 7.23	61.26 ± 2.99	99.89 ± 0.04	19.43 ± 1.49	30.61 ± 8.06	61.26 ± 2.99
		Multimodal	45.01 ± 8.16	60.74 ± 3.14	99.9 ± 0.04	19.68 ± 1.56	36.57 ± 10.09	60.74 ± 3.14
	RF	Individual	40.14 ± 9.56	60.62 ± 2.39	99.89 ± 0.04	19.74 ± 1.19	30.91 ± 10.66	60.62 ± 2.39
		Multimodal	39.62 ± 9.98	61.47 ± 3.04	99.89 ± 0.04	19.32 ± 1.51	30.14 ± 10.68	61.47 ± 3.04
	ELM	Individual	33.74 ± 15.8	61.71 ± 2.45	99.88 ± 0.03	19.21 ± 1.21	25.15 ± 13.79	61.71 ± 2.45
		Multimodal	37.24 ± 8.31	61.05 ± 4.74	99.88 ± 0.04	19.54 ± 2.36	27.45 ± 8.83	61.05 ± 4.74
mRMR	SVM	Individual	35.46 ± 18.04	53.3 ± 23.58	99.88 ± 0.04	16.26 ± 7.23	27.34 ± 15.58	53.3 ± 23.58
		Multimodal	40.38 ± 8.97	60.77 ± 3.39	99.89 ± 0.04	19.67 ± 1.69	30.95 ± 9.58	60.77 ± 3.39
	RF	Individual	40.46 ± 10.65	61.44 ± 2.31	99.89 ± 0.04	19.34 ± 1.15	31.31 ± 11.69	61.44 ± 2.31
		Multimodal	41.09 ± 8.46	61.47 ± 1.97	99.89 ± 0.04	19.32 ± 0.98	31.61 ± 9.54	61.47 ± 1.97
	ELM	Individual	40.93 ± 7.45	61.43 ± 3.8	99.89 ± 0.04	19.34 ± 1.89	31.2 ± 8.06	61.43 ± 3.8
		Multimodal	36.02 ± 8.85	61.36 ± 4.07	99.88 ± 0.04	19.38 ± 2.02	26.16 ± 8.42	61.36 ± 4.07

Table 9: Results table for each Feature Selection technique and Machine Learning Model using FA protocol data.

Method	Model	Data	F1	Sensitivity	Specificity	BER	Precision	Recall
Lasso	SVM	Individual	32.84 ± 15.87	53.94 ± 23.88	99.88 ± 0.04	15.94 ± 7.11	24.04 ± 12.64	53.94 ± 23.88
		Multimodal	46.03 ± 7.32	59.92 ± 3.92	99.9 ± 0.03	20.09 ± 1.95	38 ± 8.77	59.92 ± 3.92
	RF	Individual	40.21 ± 10.72	61.65 ± 2.92	99.89 ± 0.04	19.23 ± 1.45	30.96 ± 11.27	61.65 ± 2.92
		Multimodal	40.83 ± 12.29	60.05 ± 3.78	99.89 ± 0.04	20.03 ± 1.88	32.5 ± 12.88	60.05 ± 3.78
	ELM	Individual	34.87 ± 17.98	53.57 ± 23.7	99.88 ± 0.04	16.12 ± 7.17	26.71 ± 15.57	53.57 ± 23.7
		Multimodal	30.93 ± 13.64	61.7 ± 4.39	99.87 ± 0.04	19.21 ± 2.18	21.94 ± 11.9	61.7 ± 4.39
mRMR	SVM	Individual	43.36 ± 6.74	60.75 ± 3.57	99.89 ± 0.04	19.68 ± 1.77	34.22 ± 7.98	60.75 ± 3.57
		Multimodal	41.66 ± 6.31	61.82 ± 4.71	99.89 ± 0.04	19.15 ± 2.34	31.92 ± 6.69	61.82 ± 4.71
	RF	Individual	40.89 ± 10.84	61.15 ± 2.81	99.89 ± 0.04	19.48 ± 1.4	31.88 ± 11.74	61.15 ± 2.81
		Multimodal	39.54 ± 10.93	60.82 ± 3.23	99.89 ± 0.04	19.65 ± 1.61	30.35 ± 11.39	60.82 ± 3.23
	ELM	Individual	30.97 ± 17.04	52.19 ± 23.28	99.88 ± 0.04	16.81 ± 7.62	22.93 ± 14.32	52.19 ± 23.28
		Multimodal	34.3 ± 10.2	60.7 ± 4.91	99.88 ± 0.04	19.71 ± 2.44	24.79 ± 10.01	60.7 ± 4.91

Table 10: Results table for each Feature Selection technique and Machine Learning Model using RD protocol data.

Method	Model	Data	F1	Sensitivity	Specificity	BER	Precision	Recall
Lasso	SVM	Individual	39.39 ± 9.28	61.77 ± 3.97	99.89 ± 0.04	19.17 ± 1.97	29.76 ± 9.78	61.77 ± 3.97
		Multimodal	46.16 ± 7.02	60.67 ± 3.8	99.9 ± 0.04	19.72 ± 1.89	37.97 ± 8.86	60.67 ± 3.8
	RF	Individual	39.43 ± 9.74	61.72 ± 3.05	99.89 ± 0.04	19.2 ± 1.52	29.85 ± 10.43	61.72 ± 3.05
		Multimodal	37.83 ± 11.62	60.35 ± 3.56	99.88 ± 0.04	19.88 ± 1.78	28.77 ± 11.49	60.35 ± 3.56
	ELM	Individual	39.63 ± 11.13	61.93 ± 3.25	99.89 ± 0.04	19.09 ± 1.62	30.2 ± 11.23	61.93 ± 3.25
		Multimodal	43.07 ± 7.96	61.04 ± 3.42	99.89 ± 0.04	19.54 ± 1.7	34.04 ± 9.54	61.04 ± 3.42
mRMR	SVM	Individual	36.01 ± 17.22	54.09 ± 23.92	99.88 ± 0.04	15.86 ± 7.05	27.47 ± 14.23	54.09 ± 23.92
		Multimodal	44.45 ± 6.61	60.71 ± 3.89	99.89 ± 0.04	19.7 ± 1.93	35.57 ± 7.73	60.71 ± 3.89
	RF	Individual	39.12 ± 11.24	61.02 ± 2.84	99.89 ± 0.04	19.54 ± 1.41	29.96 ± 11.61	61.02 ± 2.84
		Multimodal	38.19 ± 9.92	61.92 ± 2.2	99.88 ± 0.04	19.1 ± 1.1	28.49 ± 10.35	61.92 ± 2.2
	ELM	Individual	40.91 ± 8.03	60.71 ± 3.38	99.89 ± 0.04	19.7 ± 1.68	31.45 ± 8.63	60.71 ± 3.38
		Multimodal	40.77 ± 8.52	60.93 ± 3.94	99.89 ± 0.04	19.59 ± 1.96	31.37 ± 9.49	60.93 ± 3.94

Table 11: Results table for each Feature Extraction technique and Machine Learning Model using AD protocol data.

Features	Data	Model	F1	Sensitivity	Specificity	BER	Precision	Recall
First Order	Individual	SVM	40.73 ± 11.61	62.18 ± 1.8	99.89 ± 0.04	18.96 ± 0.9	31.72 ± 12.7	62.18 ± 1.8
		RF	42.12 ± 9.72	63.48 ± 2.08	99.89 ± 0.04	18.32 ± 1.03	32.43 ± 10.69	63.48 ± 2.08
		ELM	40.27 ± 12.2	61.52 ± 2.34	99.89 ± 0.03	19.29 ± 1.16	31.66 ± 13.75	61.52 ± 2.34
	Multimodal	SVM	40.42 ± 12.15	62.23 ± 2.78	99.89 ± 0.04	18.94 ± 1.39	31.5 ± 13.49	62.23 ± 2.78
		RF	40.41 ± 10.55	64.64 ± 1.58	99.89 ± 0.04	17.74 ± 0.78	30.42 ± 10.95	64.64 ± 1.58
		ELM	29.69 ± 11.26	65.62 ± 3.46	99.87 ± 0.04	17.25 ± 1.73	20.18 ± 9.85	65.62 ± 3.46
GLRL	Individual	SVM	36.03 ± 11.67	60.96 ± 3.05	99.88 ± 0.04	19.58 ± 1.52	26.7 ± 11.52	60.96 ± 3.05
		RF	39.15 ± 10.29	61.1 ± 2.49	99.89 ± 0.03	19.51 ± 1.24	29.79 ± 10.89	61.1 ± 2.49
		ELM	37.25 ± 10.57	61.63 ± 2.91	99.88 ± 0.04	19.25 ± 1.45	27.72 ± 10.92	61.63 ± 2.91
	Multimodal	SVM	40.28 ± 10.7	62.2 ± 2.81	99.89 ± 0.04	18.96 ± 1.4	30.94 ± 11.8	62.2 ± 2.81
		RF	36.95 ± 11.44	61 ± 2.73	99.88 ± 0.04	19.56 ± 1.36	27.59 ± 11.49	61 ± 2.73
		ELM	34.81 ± 11.9	61.49 ± 3.56	99.88 ± 0.04	19.31 ± 1.77	25.39 ± 11.49	61.49 ± 3.56
Haralick	Individual	SVM	42.18 ± 10.7	62.69 ± 2.69	99.89 ± 0.04	18.71 ± 1.34	32.98 ± 11.64	62.69 ± 2.69
		RF	42.41 ± 10.47	62.51 ± 3.01	99.89 ± 0.04	18.8 ± 1.5	33.23 ± 11.65	62.51 ± 3.01
		ELM	38.59 ± 12.66	63.57 ± 2.44	99.88 ± 0.04	18.27 ± 1.22	29.15 ± 12.48	63.57 ± 2.44
	Multimodal	SVM	43.11 ± 10.39	62.21 ± 2.56	99.89 ± 0.04	18.95 ± 1.27	34.19 ± 11.99	62.21 ± 2.56
		RF	41.02 ± 10.48	63.4 ± 3.27	99.89 ± 0.04	18.36 ± 1.63	31.28 ± 10.89	63.4 ± 3.27
		ELM	38.7 ± 12.67	61.93 ± 2.33	99.89 ± 0.04	19.09 ± 1.16	29.63 ± 13.12	61.93 ± 2.33
HoG	Individual	SVM	44.28 ± 8.5	62.01 ± 2.84	99.89 ± 0.04	19.05 ± 1.41	35.2 ± 10.12	62.01 ± 2.84
		RF	44.53 ± 7.98	61.86 ± 2.84	99.89 ± 0.04	19.12 ± 1.41	35.44 ± 9.55	61.86 ± 2.84
		ELM	42.59 ± 6.87	61.85 ± 2.44	99.89 ± 0.04	19.13 ± 1.21	32.92 ± 7.66	61.85 ± 2.44
	Multimodal	SVM	26.68 ± 11.16	68.22 ± 7.93	99.86 ± 0.05	15.96 ± 3.97	17.04 ± 8.41	68.22 ± 7.93
		RF	44.16 ± 8.18	61.95 ± 2.82	99.89 ± 0.04	19.08 ± 1.4	34.95 ± 9.56	61.95 ± 2.82
		ELM	42.98 ± 7.51	61.41 ± 3.1	99.89 ± 0.04	19.35 ± 1.54	33.59 ± 8.58	61.41 ± 3.1
LBP	Individual	SVM	0.71 ± 1.22	16.56 ± 28.46	99.83 ± 0.07	6.02 ± 10.4	0.36 ± 0.62	16.56 ± 28.46
		RF	37.65 ± 5.74	64.97 ± 4.43	99.88 ± 0.05	17.57 ± 2.21	26.71 ± 5.32	64.97 ± 4.43
		ELM	40.18 ± 11.83	62.85 ± 3.71	99.89 ± 0.04	18.63 ± 1.86	30.77 ± 11.66	62.85 ± 3.71
	Multimodal	SVM	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		RF	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		ELM	44.04 ± 7.39	61 ± 3.22	99.89 ± 0.04	19.56 ± 1.6	35.02 ± 8.85	61 ± 3.22

Table 12: Results table for each Feature Extraction technique and Machine Learning Model using ADC protocol data.

Features	Data	Model	F1	Sensitivity	Specificity	BER	Precision	Recall
First Order	Individual	SVM	40.83 ± 10.04	62.73 ± 1.71	99.89 ± 0.04	18.69 ± 0.87	31.46 ± 11.3	62.73 ± 1.71
		RF	42.88 ± 9.74	63.28 ± 1.86	99.89 ± 0.04	18.41 ± 0.92	33.58 ± 11.22	63.28 ± 1.86
		ELM	40.86 ± 10.91	62.24 ± 1.18	99.89 ± 0.04	18.93 ± 0.59	31.86 ± 12.54	62.24 ± 1.18
	Multimodal	SVM	40.17 ± 9.11	63.49 ± 1.96	99.89 ± 0.04	18.31 ± 0.98	30.26 ± 9.66	63.49 ± 1.96
		RF	41.23 ± 11.95	63.24 ± 1.82	99.89 ± 0.04	18.43 ± 0.91	32 ± 12.79	63.24 ± 1.82
		ELM	42.02 ± 10.85	61.7 ± 3.18	99.89 ± 0.04	19.21 ± 1.59	33.11 ± 12.28	61.7 ± 3.18
GLRL	Individual	SVM	37.34 ± 11.29	60.82 ± 3.38	99.88 ± 0.04	19.65 ± 1.68	28.04 ± 11.57	60.82 ± 3.38
		RF	40.85 ± 10.42	61.1 ± 3.03	99.89 ± 0.04	19.51 ± 1.51	31.72 ± 11.4	61.1 ± 3.03
		ELM	36.51 ± 10.63	60.5 ± 3.19	99.88 ± 0.04	19.81 ± 1.59	26.99 ± 10.5	60.5 ± 3.19
	Multimodal	SVM	37.45 ± 10.72	61.43 ± 2.8	99.88 ± 0.04	19.34 ± 1.4	27.99 ± 11.21	61.43 ± 2.8
		RF	40.88 ± 9.65	61.33 ± 2.21	99.89 ± 0.04	19.39 ± 1.1	31.64 ± 10.71	61.33 ± 2.21
		ELM	32.4 ± 10.99	60.34 ± 4.75	99.88 ± 0.03	19.89 ± 2.36	22.99 ± 10.49	60.34 ± 4.75
Haralick	Individual	SVM	41.94 ± 10.76	62.79 ± 2.36	99.89 ± 0.04	18.66 ± 1.17	32.77 ± 12.37	62.79 ± 2.36
		RF	42.95 ± 10.04	62.35 ± 2.33	99.89 ± 0.04	18.88 ± 1.15	33.87 ± 11.66	62.35 ± 2.33
		ELM	41.14 ± 10.03	63.46 ± 2.89	99.89 ± 0.04	18.32 ± 1.44	31.52 ± 11.28	63.46 ± 2.89
	Multimodal	SVM	43.59 ± 8.85	62.7 ± 2.19	99.89 ± 0.04	18.7 ± 1.08	34.3 ± 10.54	62.7 ± 2.19
		RF	43.53 ± 9.15	63.53 ± 2.54	99.89 ± 0.04	18.29 ± 1.26	34.02 ± 10.65	63.53 ± 2.54
		ELM	40.03 ± 10.11	63.76 ± 3.6	99.89 ± 0.04	18.18 ± 1.8	30.18 ± 10.73	63.76 ± 3.6
HoG	Individual	SVM	44.37 ± 8.2	61.84 ± 2.89	99.89 ± 0.04	19.13 ± 1.44	35.32 ± 9.78	61.84 ± 2.89
		RF	43.91 ± 8.4	61.39 ± 2.7	99.89 ± 0.04	19.36 ± 1.34	34.95 ± 10.09	61.39 ± 2.7
		ELM	44.42 ± 7.96	61.07 ± 2.95	99.89 ± 0.04	19.52 ± 1.47	35.63 ± 9.96	61.07 ± 2.95
	Multimodal	SVM	23.89 ± 9.48	69.87 ± 5.53	99.86 ± 0.06	15.14 ± 2.77	14.74 ± 6.97	69.87 ± 5.53
		RF	43.93 ± 8.36	62.44 ± 2.52	99.89 ± 0.04	18.83 ± 1.25	34.62 ± 9.89	62.44 ± 2.52
		ELM	42.59 ± 8.83	61.65 ± 3.64	99.89 ± 0.04	19.23 ± 1.82	33.27 ± 10.25	61.65 ± 3.64
LBP	Individual	SVM	0.89 ± 1.65	22.33 ± 38.15	99.83 ± 0.07	3.15 ± 5.4	0.45 ± 0.85	22.33 ± 38.15
		RF	37.72 ± 6.89	64.07 ± 4.58	99.88 ± 0.05	18.02 ± 2.28	27.12 ± 6.75	64.07 ± 4.58
		ELM	34.99 ± 13.77	64.17 ± 4.17	99.88 ± 0.05	17.98 ± 2.07	25.69 ± 11.86	64.17 ± 4.17
	Multimodal	SVM	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		RF	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		ELM	43.28 ± 7.59	62.96 ± 4.26	99.89 ± 0.04	18.58 ± 2.13	33.6 ± 8.43	62.96 ± 4.26

Table 13: Results table for each Feature Extraction technique and Machine Learning Model using FA protocol data.

Features	Data	Model	F1	Sensitivity	Specificity	BER	Precision	Recall
First Order	Individual	SVM	39.49 ± 13.75	61.28 ± 3.21	99.89 ± 0.04	19.42 ± 1.6	30.95 ± 13.71	61.28 ± 3.21
		RF	42.36 ± 11.18	62.65 ± 3.38	99.89 ± 0.04	18.73 ± 1.68	33.52 ± 12.64	62.65 ± 3.38
		ELM	40.42 ± 13.76	61.09 ± 4.37	99.89 ± 0.04	19.51 ± 2.18	31.92 ± 13.7	61.09 ± 4.37
	Multimodal	SVM	44.13 ± 14.18	59.79 ± 2.44	99.9 ± 0.03	20.15 ± 1.22	37.3 ± 15.22	59.79 ± 2.44
		RF	46.17 ± 11.18	60.87 ± 2.29	99.9 ± 0.03	19.61 ± 1.14	38.85 ± 13.09	60.87 ± 2.29
		ELM	41.95 ± 14.72	60.61 ± 3.49	99.89 ± 0.03	19.75 ± 1.74	34.05 ± 14.14	60.61 ± 3.49
GLRL	Individual	SVM	36.25 ± 13.07	60.9 ± 4.29	99.88 ± 0.04	19.61 ± 2.14	27.12 ± 12.5	60.9 ± 4.29
		RF	40.75 ± 11.19	61.2 ± 2.99	99.89 ± 0.04	19.46 ± 1.49	31.74 ± 12.03	61.2 ± 2.99
		ELM	35.76 ± 10.85	60.12 ± 3.98	99.88 ± 0.04	20 ± 1.98	26.46 ± 11.03	60.12 ± 3.98
	Multimodal	SVM	37.97 ± 14.53	59.61 ± 4.41	99.89 ± 0.03	20.25 ± 2.2	29.97 ± 15.96	59.61 ± 4.41
		RF	42.63 ± 11.96	60.75 ± 2	99.9 ± 0.03	19.68 ± 1	34.71 ± 14.53	60.75 ± 2
		ELM	30.23 ± 10.83	60.34 ± 4.84	99.87 ± 0.04	19.9 ± 2.4	20.98 ± 9.86	60.34 ± 4.84
Haralick	Individual	SVM	43.07 ± 9.06	61.5 ± 3.2	99.89 ± 0.04	19.3 ± 1.59	34.25 ± 10.99	61.5 ± 3.2
		RF	43.09 ± 11.02	62.47 ± 3.11	99.89 ± 0.04	18.82 ± 1.55	34.36 ± 12.44	62.47 ± 3.11
		ELM	44.89 ± 8.68	61.25 ± 3.26	99.9 ± 0.04	19.42 ± 1.62	36.43 ± 10.56	61.25 ± 3.26
	Multimodal	SVM	48.61 ± 9.1	59.91 ± 1.85	99.91 ± 0.03	20.09 ± 0.92	41.95 ± 11.55	59.91 ± 1.85
		RF	46.7 ± 10.27	59.82 ± 1.64	99.9 ± 0.03	20.14 ± 0.82	39.65 ± 12.4	59.82 ± 1.64
		ELM	22.24 ± 20.34	53.69 ± 24.35	99.86 ± 0.04	16.07 ± 7.63	15.86 ± 16.52	53.69 ± 24.35
HoG	Individual	SVM	44.28 ± 8.37	61.97 ± 2.83	99.89 ± 0.04	19.07 ± 1.41	35.13 ± 9.81	61.97 ± 2.83
		RF	44.31 ± 8.74	62.75 ± 3.11	99.89 ± 0.04	18.68 ± 1.55	34.94 ± 10.16	62.75 ± 3.11
		ELM	43.41 ± 7.88	62.08 ± 2.85	99.89 ± 0.04	19.01 ± 1.42	33.94 ± 9.01	62.08 ± 2.85
	Multimodal	SVM	34.11 ± 12.09	66.02 ± 3.73	99.87 ± 0.05	17.05 ± 1.87	23.87 ± 11.07	66.02 ± 3.73
		RF	48.54 ± 7.76	60.6 ± 2.09	99.9 ± 0.03	19.75 ± 1.04	41.16 ± 10.12	60.6 ± 2.09
		ELM	48.28 ± 7.2	60.26 ± 2.97	99.9 ± 0.03	19.92 ± 1.48	40.73 ± 8.8	60.26 ± 2.97
LBP	Individual	SVM	1.93 ± 2.15	52.3 ± 38.3	99.83 ± 0.07	9.62 ± 9.53	0.99 ± 1.12	52.3 ± 38.3
		RF	36.13 ± 9.04	66.83 ± 2.84	99.88 ± 0.05	16.65 ± 1.41	25.48 ± 9.1	66.83 ± 2.84
		ELM	43.65 ± 6.61	61.46 ± 3.08	99.89 ± 0.04	19.32 ± 1.53	34.25 ± 7.87	61.46 ± 3.08
	Multimodal	SVM	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		RF	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		ELM	45.66 ± 8.33	59.86 ± 2.83	99.9 ± 0.04	20.12 ± 1.41	37.71 ± 10.26	59.86 ± 2.83

Table 14: Results table for each Feature Extraction technique and Machine Learning Model using RD protocol data.

Features	Data	Model	F1	Sensitivity	Specificity	BER	Precision	Recall
First Order	Individual	SVM	40.22 ± 11.17	62.14 ± 2.28	99.89 ± 0.04	18.98 ± 1.14	31.13 ± 12.17	62.14 ± 2.28
		RF	41.47 ± 10.41	63.65 ± 2.23	99.89 ± 0.04	18.23 ± 1.11	31.83 ± 11.39	63.65 ± 2.23
		ELM	40.22 ± 11.35	61.41 ± 2.44	99.89 ± 0.03	19.35 ± 1.21	31.33 ± 12.67	61.41 ± 2.44
	Multimodal	SVM	44.56 ± 10.63	60.42 ± 2.45	99.9 ± 0.03	19.84 ± 1.23	36.72 ± 13.52	60.42 ± 2.45
		RF	45.24 ± 10.24	62.75 ± 1.48	99.9 ± 0.03	18.68 ± 0.74	36.56 ± 11.83	62.75 ± 1.48
		ELM	35.28 ± 20.96	52.83 ± 23.59	99.89 ± 0.03	16.49 ± 7.5	29 ± 20.3	52.83 ± 23.59
GLRL	Individual	SVM	35.76 ± 12.56	60.4 ± 3.36	99.88 ± 0.04	19.86 ± 1.68	26.67 ± 12.22	60.4 ± 3.36
		RF	39.5 ± 10.51	60.86 ± 2.85	99.89 ± 0.04	19.62 ± 1.42	30.25 ± 10.97	60.86 ± 2.85
		ELM	38.4 ± 10.78	61.27 ± 3.49	99.89 ± 0.03	19.42 ± 1.74	28.92 ± 10.98	61.27 ± 3.49
	Multimodal	SVM	42.03 ± 12.71	61.14 ± 2.47	99.89 ± 0.03	19.48 ± 1.23	34.14 ± 15.53	61.14 ± 2.47
		RF	38.75 ± 13.68	59.95 ± 1.92	99.89 ± 0.03	20.08 ± 0.97	30.76 ± 15.68	59.95 ± 1.92
		ELM	35.66 ± 13.59	61.56 ± 3.57	99.88 ± 0.04	19.28 ± 1.78	26.63 ± 13.42	61.56 ± 3.57
Haralick	Individual	SVM	42.55 ± 10.24	62.77 ± 2.18	99.89 ± 0.04	18.67 ± 1.08	33.21 ± 11.25	62.77 ± 2.18
		RF	43.01 ± 9.6	62.69 ± 2.63	99.89 ± 0.04	18.71 ± 1.31	33.65 ± 10.63	62.69 ± 2.63
		ELM	39.37 ± 13.65	62.45 ± 2.64	99.89 ± 0.04	18.83 ± 1.31	30.6 ± 14.77	62.45 ± 2.64
	Multimodal	SVM	47.7 ± 9.68	60.32 ± 2.33	99.9 ± 0.03	19.89 ± 1.16	40.55 ± 12.14	60.32 ± 2.33
		RF	45.97 ± 9.64	62.02 ± 2.75	99.9 ± 0.03	19.04 ± 1.37	37.4 ± 10.98	62.02 ± 2.75
		ELM	42.45 ± 12.29	60.73 ± 3.23	99.89 ± 0.03	19.69 ± 1.62	33.89 ± 12.69	60.73 ± 3.23
HoG	Individual	SVM	44.4 ± 8.08	62.3 ± 2.71	99.89 ± 0.04	18.9 ± 1.34	35.2 ± 9.7	62.3 ± 2.71
		RF	44.16 ± 8.19	61.69 ± 2.83	99.89 ± 0.04	19.21 ± 1.4	35.11 ± 9.9	61.69 ± 2.83
		ELM	43.03 ± 7.28	62.46 ± 3.97	99.89 ± 0.04	18.82 ± 1.98	33.32 ± 8.21	62.46 ± 3.97
	Multimodal	SVM	30.64 ± 10.74	67.83 ± 6	99.87 ± 0.05	16.15 ± 3.01	20.27 ± 8.42	67.83 ± 6
		RF	48.73 ± 7.72	60.09 ± 2.3	99.91 ± 0.03	20 ± 1.14	41.63 ± 9.92	60.09 ± 2.3
		ELM	47.65 ± 7.17	60.09 ± 3.52	99.9 ± 0.03	20.01 ± 1.76	39.87 ± 8.6	60.09 ± 3.52
LBP	Individual	SVM	1.22 ± 1.81	30.8 ± 40.49	99.83 ± 0.07	6.06 ± 9.88	0.63 ± 0.94	30.8 ± 40.49
		RF	38.74 ± 5.62	64.68 ± 5.22	99.88 ± 0.05	17.72 ± 2.61	27.91 ± 5.32	64.68 ± 5.22
		ELM	42.68 ± 7.68	62.61 ± 3.51	99.89 ± 0.04	18.75 ± 1.75	33.03 ± 8.8	62.61 ± 3.51
	Multimodal	SVM	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		RF	0 ± 0	0 ± 0	99.83 ± 0.07	0 ± 0	0 ± 0	0 ± 0
		ELM	48.61 ± 6	60.05 ± 3.44	99.9 ± 0.03	20.02 ± 1.72	41.07 ± 7.17	60.05 ± 3.44