Can Pyradiomics feature extraction significantly improve 3D brain MRI segmentation?

Daniel Kerber
Faculty of Medicine
University of Bern
Bern, Switzerland
daniel.kerber@students.unibe.ch

Diego Zeiter
Faculty of Medicine
University of Bern
Bern, Switzerland
diego.zeiter1@students.unibe.ch

Matteo Tagliabue
Faculty of Medicine
University of Bern
Bern, Switzerland
matteo.tagliabue@students.unibe.ch

***Abstract*—Several machine learning-based automatic segmentation algorithms have been developed to deal with the complex, laborious, and variability-prone task of manually segmenting the different regions of the brain tissue in 3D MRI. These algorithms use raw data images and features, including intensity, texture, shape, and statistical information, to create the desired segmentation. Among the different feature-extracting tools, one is Pyradiomics, which uses radiomics to extract high-dimensional data that are often not discernible to the human eye. This research hypothesizes that incorporating Pyradiomics in a healthy brain MRI segmentation pipeline can significantly enhance the accuracy and consistency of the inferred segmentations. Our research tests 54 features from three commonly used Pyradiomics classes within the medical image analysis (MIA) Lab brain MRI segmentation pipeline. These features are evaluated using the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) metrics. However, none of the results showed significant improvements in segmentation performance using an individual Pyradiomics feature. Consequently, we conclude that, concerning the tested features, Pyradiomics is best applied to unhealthy brain tissue. Future research might explore combining the best-performing Pyradiomics features or test the remaining Pyradiomics features, though substantial improvements in the segmentation of healthy brain tissue using this package remain uncertain.**

***Index Terms*—Segmentation, magnetic resonance imaging, brain, pyradiomics, feature extraction, segmentation performance, Dice coefficient, Hausdorff distance**

I. Introduction

Segmentation in medical image analysis allows for precise delineation of areas of interest, enabling quantitative and accurate diagnosis, treatment planning, and monitoring of various conditions. The current gold-standard segmentation is obtained manually through medical expert annotation. However, this method is laborious, time-consuming, and susceptible to intra- and inter-observer variability [1]. These drawbacks are particularly pronounced in 3D magnetic resonance imaging (MRI) of brains, where the intricate structures are divided across many image slices and exhibit similar tissue characteristics, which makes accurate segmentation more challenging than as in other distinctive structures, such as bones.

To address these limitations, several machine learning- and deep learning-based algorithms have been developed [2]. These algorithms leverage the raw images and features extracted from them, including intensity, texture, shape, and statistical information, to autonomously predict segmentation labels for unseen data. In this context, Pyradiomics [3], an open-source Python package for extracting radiomics features from medical imaging, has emerged as a significant tool, especially in oncology. It is based on the assumption that diseased tissue carries out information imperceptible to the human eye but visible in a transformed high-dimensional space. By extending this assumption, we hypothesize that Pyradiomics features can also be an important instrument to extract data from healthy brain tissue, thereby significantly enhancing the performance of healthy brain MRI segmentation.

The impact of our research lies in the potential to improve the accuracy while reducing the variability of the predicted segmentations so that ultimately, in a clinical setting, patients would get more reliable diagnosis and treatment planning. To achieve that, within the scope of this work, we used a dataset of magnetic resonance (MR) images of 30 unrelated healthy individuals and tested three commonly used Pyradiomics classes within the medical image analysis (MIA) Lab brain MRI segmentation pipeline. We evaluated these features using the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) metrics.

II. Materials and Methods

A. Dataset

The dataset includes 3 Tesla MR images from the head of 30 unrelated healthy individuals, extracted from the Human Connectome Project's database of healthy volunteers [4]. The data available for each individual included a T1-weighted (T1w) and a T2-weighted (T2w) MR image volume with bias field correction applied. Though the skull was retained, anonymization through defacing [5] was used. The two modalities provided in the original T1w subject space include the ground truth label maps and brain masks. Due to a lack of manual expert annotations, the ground truth labels provided were generated automatically with FreeSurfer 5.3 [6]. Additionally, the data was further enhanced with affine transformations to align the images with the given MNI152 atlas using nonlinear FNIRT, allowing for more comprehensive analysis and comparison. In this study, we randomly divided the dataset, in which 20 individuals were used for the model training and 10 for prediction testing.

B. MIA Lab Pipeline

The Pyradiomics feature extraction was added to the MIA Lab pipeline [7] and evaluated based on segmentation metric performances. The pipeline is divided into five categories, including registration, pre-processing, feature extraction, classification, and post-processing. An overview of the workflow is presented below in Figure 1.

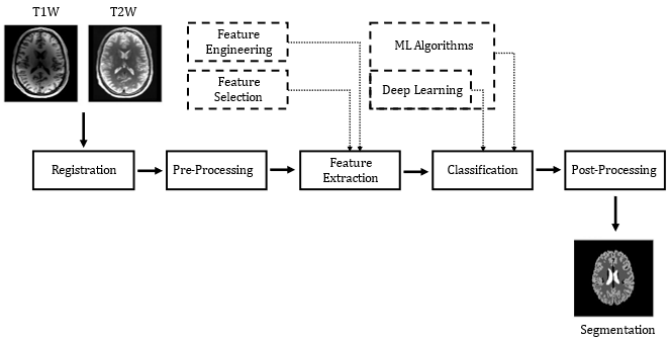

Fig. 1: Brain MRI segmentation procedure using the MIA Lab pipeline.

The registration process starts by aligning the T1w and T2w MR image slices with the provided atlas. In pre-processing, the brain masks were used to do skull-stripping to remove non-brain tissue from the data, and min-max normalization was applied to ensure pixel intensity values range between 0 and 1. For the baseline feature extraction, which is the pipeline area we want to improve using the Pyradiomics package, we kept the already implemented atlas coordinates, intensity, and gradient of intensity features. In the classification phase, we trained and tested a random forest classifier on how well it can infer the tissue class (amygdala, grey matter, hippocampus, thalamus, and white matter) for each given voxel based on the information extracted from the features. No post-processing was performed, considering that the scope of this project was to improve the performance of the model classifier in predicting the segmentation labels based on the features extracted and not through post-processing techniques.

C. Pyradiomics Features and Experiments

From the eight classes of features offered in the Pyradiomics package containing 120 features, we tested the classes gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), and first-order features (FOF) based on frequently used features from papers such as Chen et al. and Goya-Outi et al.

Chen et al. presented in their paper an automatic computer-aided diagnosis of gliomas that combined automatic segmentation and radiomics. Among their 20 selected Pyradiomics features were features from first-order, GLCM, and GLSZM, the most occurring ones [8]. Goya-Outi et al. aimed to propose rules to compute reliable textural indices from multimodal 3D brain MRI of patients with diffuse intrinsic pontine glioma. The Pyradiomics features used were from GLCM, GLSZM, and GLRLM. [9]. Although these studies focused on diseased brain tissues, we expect that the same features could also yield significant insights into differentiating the region labels of healthy brain tissue segmentations.

According to the official Pyradiomics documentation [3], FOF describes a basic statistical analysis of voxel intensities within defined mask regions, GLCM offers a second-order joint probability function analysis, and GLSZM focuses on quantifying gray level zones, i.e., number of connected voxels that share the same gray level intensity.

Our approach involved individually implementing and testing all 59 features in the MIA Lab brain MRI segmentation pipeline and evaluating its performance based on the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) metrics for each segmentation compared to the baseline, which does not include the Pyradiomics feature. In the end, only 54 features were incorporated into our analysis, as five features caused implementation challenges within the pipeline.

D. Evaluation Metrics

To assess the effectiveness of integrating Pyradiomics into the segmentation model, we employed two evaluation metrics: the Dice Similarity Coefficient (DSC) and the Hausdorff Distance (HD). The DSC measures the overlap between the predicted segmentation and the ground truth annotation, providing a quantitative indication of segmentation accuracy. Meanwhile, HD evaluates the maximum distance between segmented and reference boundary points, offering insights into spatial accuracy. Therefore, the desirable results should improve the DSC and HD to approximate the values to 1 and 0, respectively. These results were placed on different concise boxplots encompassing each feature experimented per label region of the brain, allowing a comprehensive comparison of the performances.

III. Results

A. Analysis of DSC and HD

When analyzing the DSC and HD values obtained for the five brain labels we tested, using each of the GLCM and GLSZM Pyradiomics features, we observed minimal or no changes compared to the baseline. FOFs were the ones that showed the greatest improvement, yet the values did not surpass a 9% reduction in HD for a single label. Consequently, we have focused on highlighting, with boxplots, only those results where significant changes were noted. All other results were added into comprehensive tables detailing all DSC and HD values for each label and feature in the Appendix. In these tables, cells indicating improved performance were marked in green, while those showing no improvement or deterioration remained white.

B. GLCM Features

For GLCM, out of the 21 features analyzed, none showed significant changes in DSC compared to the baseline. However, nearly all features improved the HD marginally across all labels. For example, the amygdala mean HD (see Appendix A) was reduced by 1.9% for twelve different GLCM features.

However, none of these features stood out as the best, given that all of them resulted in an equal improvement for the thalamus, see Figure 2 below, in which the mean HD was reduced by 2.7%. For the other labels, improvements in mean HD were relatively minor, below 1%.

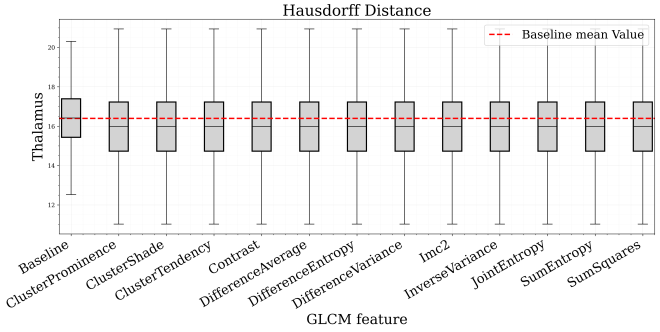

Fig. 2: Best-performing GLCM features for the thalamus.

C. GLSZM Features

The GLSZM features did not significantly alter the DSC values. The amygdala DSC improved by 1.9% using nine of the GLSZM features; meanwhile, the thalamus DSC got worse by 1.3% for any of the sixteen features, and the other labels were unchanged. However, similar to GLCM, some improvements in mean HD were observed in the amygdala and thalamus labels, but also some deterioration of the grey matter and hippocampus. For instance, the HD of the thalamus was reduced by up to 5.3% with six different features out of the 16. However, the most impactful features for the amygdala, reducing the mean HD by 6.4%, including Zone Percentage, Small Area Emphasis, and Small Area High and Low Gray Level Emphasis, as depicted in Figure 3 below, caused the thalamus HD to increase by 1.6%. Similarly, features that marginally improved the thalamus and amygdala caused a worsening of the HD in the hippocampus of up to 2.1% and grey matter of 4.8%; see Appendix B.

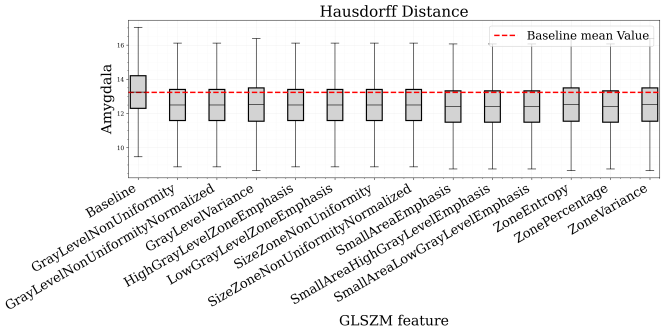

Fig. 3: Best-performing GLSZM features for the amygdala.

Moreover, despite minor improvements in most label HD variabilities, the thalamus robustness significantly decreased as the variability rose nearly sixfold (from 1.4 to 9.5) compared to the baseline for all the GLSZM features.

D. First-order Features (FOF)

Lastly, the FOF group enhanced the DSC in the amygdala and hippocampus without causing the deterioration of the other labels. The Median feature led the DSC improvements, increasing hippocampus DSC by 5.9% and amygdala by 1.9%, with RootMeanSquared closely behind. Skewness was the only feature to boost white matter DSC by 1.2%. Moreover, FOF improved HD for every label except the thalamus, where the mean HD increased by 15.5% (see Appendix C). Energy and TotalEnergy were the top performers, enhancing amygdala HD by 7.9%, as depicted in Figure 4, and improving white matter by 7.1% and grey matter by 3.8%. Median and Root Mean Squared were next in line, optimizing white and grey matter by 8.8% and 4.0%, respectively, while having a lesser effect on the amygdala and no change in the hippocampus and thalamus.

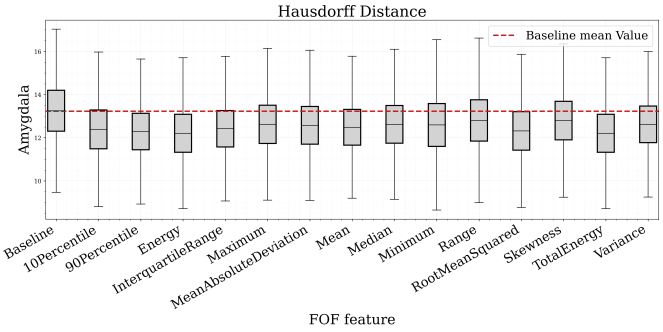

Fig. 4: Best-performing first-order features for the amygdala.

IV. Discussion

The findings of our study provided valuable insights into the efficacy of using numerous Pyradiomics features in improving the precision and robustness of an automatic pipeline for the segmentation of the different regions of healthy brains. The lack of significant changes in DSC values and various changes in HD across the 54 individually extracted features could suggest that Pyradiomics has a greater focus and influence on delineating the boundaries of the different labels instead of the spatial similarity between segmented and ground truth data. This could be useful for applications such as brain tumor boundary location and other oncology-related issues, where Pyradiomics is often applied to differentiate between healthy and diseased tissue. These results using healthy patients can be explained, given that the principal assumption of Pyradiomics was not satisfied. Its features extract quantitative information through data transformation techniques from otherwise not perceptible diseased tissue properties. Thus, Pyradiomics cannot differentiate between the properties of the various parts of a healthy brain, at least when only considering the amygdala, hippocampus, thalamus, grey matter, and white matter. A possible explanation for the differences between the similar results from GLCM/GLSZM and FOF is that both gray level-based matrix features analyze texture and patterns, whereas FOF is purely based on voxel intensities and does not account for spatial factors.

Moreover, one of the limitations of this study was that only default parameters were used for every feature setup, which hinders the analysis of the full capabilities of Pyradiomics in healthy brain tissue. However, we suspect that even if fine-tuning the features, it would not significantly improve the metrics, considering that the changes with the default parameters were insignificant to the point where there were nearly no changes in DSC and a maximum reduction of HD of 8.8% for two of the 54 features across five brain labels at the expense of an increase in computational cost for the training of the model. Additionally, several features that improved some of the labels' HD caused an increase in HD in others. Similarly, our results do not necessarily encourage combining multiple Pyradiomics techniques to test the model, given that the performance of most features was either constant with the baseline or marginally better ($< 3\%$) for some labels. Therefore, given the results obtained through the metrics HD and DSC, the hypothesis that Pyradiomic features, specifically using GLCM, GLSZM, and FOF, could significantly improve the performance of the segmentation pipeline for healthy brain tissue was not satisfied. To further investigate this hypothesis, one could optimize the feature parameters or also test the different classes such as 2D/3D shape-based, Gray Level Run Length Matrix, Neighbouring Gray Tone Difference Matrix, and Gray Level Dependence Matrix, and a combination of multi-classes and features.

V. Conclusion

Our investigation into the efficacy of Pyradiomics features (GLCM, GLSZM, and FOF) extraction of healthy brain tissue in MR images did not significantly enhance the segmentation performance between the different labels, therefore failing to prove our hypothesis. The marginal improvements observed, especially in HD, suggest that Pyradiomics are more suited when dealing with diseased tissue, e.g., to find the brain tumor boundary location. Despite using almost half of all features available and relying on their default parameters, the results indicated limited potential for enhancing healthy brain tissue segmentation. However, future research may explore the optimization of these parameters or investigate the other Pyradiomics classes or whether training the model using multi-feature combinations of the best-performing techniques would improve the metrics.

References

[1] M. Bach Cuadra, J. Favre, and P. Omoumi, "Quantification in Musculoskeletal Imaging Using Computational Analysis and Machine Learning: Segmentation and Radiomics," eng, *Seminars in Musculoskeletal Radiology*, vol. 24, no. 1, pp. 50–64, Feb. 2020, ISSN: 1098-898X. DOI: 10.1055/s-0039-3400268.

[2] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 449–459, Aug. 1, 2017, ISSN: 1618-727X. DOI: 10.1007/s10278-017-9983-4. [Online]. Available: https://doi.org/10.1007/s10278-017-9983-4 (visited on 12/13/2023).

[3] J. J. M. van Griethuysen, A. Fedorov, C. Parmar, *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Research*, vol. 77, no. 21, e104–e107, 2017. DOI: 10.1158/0008-5472.CAN-17-0339. [Online]. Available: https://doi.org/10.1158/0008-5472.CAN-17-0339.

[4] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, and K. Ugurbil, "The WU-minn human connectome project: An overview," *NeuroImage*, vol. 80, pp. 62–79, Oct. 15, 2013, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.05.041. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3724347/ (visited on 12/13/2023).

[5] M. Milchenko and D. Marcus, "Obscuring surface anatomy in volumetric imaging data," *Neuroinformatics*, vol. 11, no. 1, pp. 65–75, Jan. 2013, ISSN: 1559-0089. DOI: 10.1007/s12021-012-9160-3.

[6] B. Fischl, D. H. Salat, E. Busa, *et al.*, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, Jan. 31, 2002, ISSN: 0896-6273. DOI: 10.1016/s0896-6273(02)00569-x.

[7] S. Pereira, A. Pinto, J. Oliveira, A. M. Mendrik, J. H. Correia, and C. A. Silva, "Automatic brain tissue segmentation in MR images using random forests and conditional random fields," *Journal of Neuroscience Methods*, vol. 270, pp. 111–123, Sep. 1, 2016, ISSN: 1872-678X. DOI: 10.1016/j.jneumeth.2016.06.017.

[8] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "Computer-aided grading of gliomas combining automatic segmentation and radiomics," *International journal of biomedical imaging*, vol. 2018, 2018.

[9] J. Goya-Outi, F. Orlhac, R. Calmon, *et al.*, "Computation of reliable textural indices from multimodal brain mri: Suggestions based on a study of patients with diffuse intrinsic pontine glioma," *Physics in Medicine & Biology*, vol. 63, no. 10, p. 105 003, 2018.

VI. Appendix

Appendix A. Complete set of GLCM Results

GLCM	DICE										HDRFDST									
	Amygdala		Hippocampus		Thalamus		WhiteMatter		GreyMatter		Amygdala		Hippocampus		Thalamus		WhiteMatter		GreyMatter	
	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
Baseline	0.52	0.03	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.23	1.40	13.47	1.07	16.39	1.44	3.87	0.35	2.49	0.64
Autocorrelation	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
ClusterProminence	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
ClusterShade	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
ClusterTendency	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
Contrast	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
Correlation	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
DifferenceAverage	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
DifferenceEntropy	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
DifferenceVariance	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
Id	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
Idm	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
Idmn	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
Idn	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
Imc1	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
Imc2	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
InverseVariance	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
JointEnergy	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
JointEntropy	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
MaximumProbability	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
SumEntropy	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
SumSquares	0.52	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63

TABLE I: All results for the five brain labels using every GLCM feature individually.

Appendix B. Complete set of GLSZM Results

GLSZM	DICE										HDRFDST									
	Amygdala		Hippocampus		Thalamus		WhiteMatter		GreyMatter		Amygdala		Hippocampus		Thalamus		WhiteMatter		GreyMatter	
	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
Baseline	0.52	0.03	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.23	1.40	13.47	1.07	16.39	1.44	3.87	0.35	2.49	0.64
GrayLevelNonUniformity	0.53	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.47	1.34	13.73	0.99	15.52	9.39	3.88	0.34	2.61	0.62
GrayLevelNonUniformityNormalize	0.53	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.47	1.34	13.73	0.99	15.52	9.39	3.88	0.34	2.61	0.62
GrayLevelVariance	0.52	0.04	0.50	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.50	1.43	13.75	1.00	15.67	9.52	3.88	0.34	2.56	0.61
HighGrayLevelZoneEmphasis	0.53	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.47	1.34	13.73	0.99	15.52	9.39	3.88	0.34	2.61	0.62
LargeAreaEmphasis	0.53	0.04	0.50	0.02	0.75	0.02	0.81	0.02	0.71	0.01	12.81	1.42	13.47	0.95	17.02	9.63	3.91	0.34	2.53	0.62
LargeAreaHighGrayLevelEmphasis	0.53	0.04	0.50	0.02	0.75	0.02	0.81	0.02	0.71	0.01	12.81	1.42	13.47	0.95	17.02	9.63	3.91	0.34	2.53	0.62
LargeAreaLowGrayLevelEmphasis	0.53	0.04	0.50	0.02	0.75	0.02	0.81	0.02	0.71	0.01	12.81	1.42	13.47	0.95	17.02	9.63	3.91	0.34	2.53	0.62
LowGrayLevelZoneEmphasis	0.53	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.47	1.34	13.73	0.99	15.52	9.39	3.88	0.34	2.61	0.62
SizeZoneNonUniformity	0.53	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.47	1.34	13.73	0.99	15.52	9.39	3.88	0.34	2.61	0.62
SizeZoneNonUniformityNormalized	0.53	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.47	1.34	13.73	0.99	15.52	9.39	3.88	0.34	2.61	0.62
SmallAreaEmphasis	0.52	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.38	1.34	13.55	1.01	16.65	9.22	3.88	0.34	2.49	0.61
SmallAreaHighGrayLevelEmphasis	0.52	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.38	1.35	13.55	1.01	16.65	9.22	3.88	0.34	2.49	0.61
SmallAreaLowGrayLevelEmphasis	0.52	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.38	1.35	13.55	1.01	16.65	9.22	3.88	0.34	2.49	0.61
ZoneEntropy	0.52	0.04	0.50	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.50	1.43	13.75	1.00	15.67	9.52	3.88	0.34	2.56	0.61
ZonePercentage	0.52	0.04	0.51	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.38	1.35	13.55	1.01	16.65	9.22	3.88	0.34	2.49	0.61
ZoneVariance	0.52	0.04	0.50	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.50	1.43	13.75	1.00	15.67	9.52	3.88	0.34	2.56	0.61

TABLE II: All results for the five brain labels using every GLSZM feature individually.

Appendix C. Complete set of FOF Results

FOF	DICE										HDRFDST									
	Amygdala		Hippocampus		Thalamus		WhiteMatter		GreyMatter		Amygdala		Hippocampus		Thalamus		WhiteMatter		GreyMatter	
	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
Baseline	0.517	0.03	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.23	1.40	13.47	1.07	16.39	1.44	3.87	0.35	2.49	0.64
10Percentile	0.521	0.04	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.37	1.33	13.10	1.02	18.93	1.99	3.88	0.39	2.48	0.69
90Percentile	0.518	0.03	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.27	1.25	13.21	0.91	17.65	3.46	3.75	0.35	2.41	0.54
Energy	0.525	0.03	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.19	1.30	13.33	0.87	17.04	0.91	3.60	0.41	2.43	0.62
Entropy	0.520	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.98	1.40	13.37	1.00	15.95	1.84	3.86	0.34	2.47	0.63
InterquartileRange	0.522	0.04	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.40	1.24	13.49	1.58	16.66	0.84	3.94	0.28	2.55	0.63
Kurtosis	0.523	0.04	0.50	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.84	1.32	13.55	0.95	16.39	1.00	3.96	0.36	2.65	0.68
Maximum	0.520	0.03	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.60	1.30	13.34	1.02	17.69	3.10	3.82	0.35	2.53	0.63
MeanAbsoluteDeviation	0.524	0.03	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.47	1.22	14.04	0.92	16.84	1.00	3.84	0.35	2.43	0.62
Mean	0.518	0.04	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.56	1.29	13.17	0.83	17.14	0.79	3.68	0.42	2.55	0.63
Median	0.525	0.04	0.54	0.04	0.76	0.03	0.81	0.02	0.71	0.01	12.60	1.29	13.57	0.93	18.35	1.89	3.53	0.39	2.39	0.62
Minimum	0.523	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.57	1.46	13.23	1.09	16.96	1.08	3.80	0.36	2.57	0.69
Range	0.516	0.03	0.52	0.03	0.75	0.02	0.81	0.02	0.71	0.01	12.78	1.41	13.11	0.98	18.25	1.23	3.75	0.37	2.54	0.60
RootMeanSquared	0.527	0.03	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.30	1.32	13.67	0.95	16.89	1.06	3.62	0.43	2.39	0.61
Skewness	0.518	0.04	0.51	0.03	0.76	0.03	0.82	0.01	0.71	0.01	12.77	1.32	14.66	0.63	16.55	0.97	3.84	0.38	2.80	0.80
TotalEnergy	0.525	0.03	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.19	1.30	13.33	0.87	17.04	0.91	3.60	0.41	2.43	0.62
Uniformity	0.519	0.04	0.51	0.03	0.76	0.03	0.81	0.02	0.71	0.01	13.03	1.45	13.43	1.04	16.36	1.90	3.87	0.35	2.47	0.63
Variance	0.521	0.04	0.52	0.03	0.76	0.03	0.81	0.02	0.71	0.01	12.60	1.25	12.86	0.97	17.79	0.82	3.82	0.33	2.55	0.63

TABLE III: All results for the five brain labels using every first-order feature individually.