Comparative Analysis of Atlas-Based Versus Machine Learning-Based Approach for Brain Tissue Segmentation in Medical Imaging

1st Lukas Studer
Medical Image Analysis Lab
University of Bern
Bern, Switzerland
lukas.studer1@students.unibe.ch

2nd Sophie Kamradt
Medical Image Analysis Lab
University of Bern
Bern, Switzerland
sophie.kamradt@students.unibe.ch

3rd Newton Kelvin Ollengo
Medical Image Analysis Lab
University of Bern
Bern, Switzerland
newton.ollengo@students.unibe.ch

***Abstract*—Medical Image Segmentation is a vital part of correctly diagnosing underlying conditions and treatment planning. The current gold standard lies in manual segmentation, which is tedious and time-consuming. Both an atlas-based and a machine-learning approach can be used to automate the process. To determine which shows superior performance, both approaches were compared against the ground truth and in their respective performances. For the machine learning approach, a random forest algorithm was used and the optimal parameters for the estimators and depth were determined. The labels of the atlas were created using 20 patient images with ground truth labels. These labels are then registered to the test images. An atlas is created using generated labels. The best results for the atlas based approach were generated at a probability of 0.35 while the best results for the machine-learning approach were generated at a depth of 30 with 70 estimators. Atlas based segmentation performed better than the machine-learning based approach according to the Hausdorff Distance. However, this is true, especially for structures with low anatomical variation, such as the amygdala, the hippocampus, and the thalamus. The white- and gray matter were better segmented by the machine-learning based approach, possibly due to their strong edges. Both segmentation approaches have their strong suit, and while the atlas based approach outperforms the machine learning approach only in regions with lower variability, it might be beneficial to combine the two approaches.**

***Index Terms*—Brain Tissue Segmentation, Atlas-Based Segmentation, Machine Learning-Based Segmentation, Random Forest Algorithm, Hausdorff Distance, Dice Coefficient, Medical Imaging**

I. Introduction

Medical image segmentation plays an important role in understanding and diagnosing underlying conditions. Particularly in the area of neurology has medical image segmentation been vital in diagnosing and consecutively treating an underlying condition. The choice of the segmentation algorithm and its precision therefore directly related to the correct diagnosis and hence prescribed choice of treatment. Any inaccurate delineation of the organ structures could lead to a wrong diagnosis or improper treatment [6] [7]. There exist several different segmentation algorithms and techniques, including atlas-based techniques or machine learning-based approaches. While atlas-based segmentation relies on a predefined dataset, the atlas, which serves as a reference for the segmentation of new images, in machine learning the model is trained on images where both the input and the output are given. The model then is trained and can afterward be applied to new images without any atlas. While the machine learning based approach demonstrates a certain flexibility and adaptivity, it also requires large amounts of data to be trained. The atlas-based segmentation requires much less data but is also less flexible and adaptable. We hypothesize that *Atlas-based segmentation consists of a powerful baseline for brain tissue segmentation when compared to an ML-based approach*. To compare both segmentation methods, both the DICE score and the Hausdorff distance have been used as a measure of comparison.

A. Related Work

Radiology relies more and more on automatic segmentation, as the process is tedious and time-consuming. Currently, the gold-standard is still manual segmentation, done by physicians [7]. In 2019 a study was conducted comparing an atlas-based vs. deep-learning-based approach to auto-segment cancerous liver tissue. The outcome of the study demonstrated better Hausdorff distances and higher DICE scores in the case of the deep learning framework [1]. More studies support the findings that deep learning outperforms the atlas-based approaches [4] [5]. Yaakub and their team even combined the two segmentation techniques, training a deep learning network on the atlases [8].

II. Materials and Methods

To compare whether the atlas-based approach outperforms the machine-learning-based approach, MRI images of ten patients were passed through both algorithms. The images were compared in the original native space. The general idea can be seen in figure 1.

A. Medical Image Analysis Pipeline

Both approaches follow the general structure of a medical image analysis pipeline. The workflow of the medical image

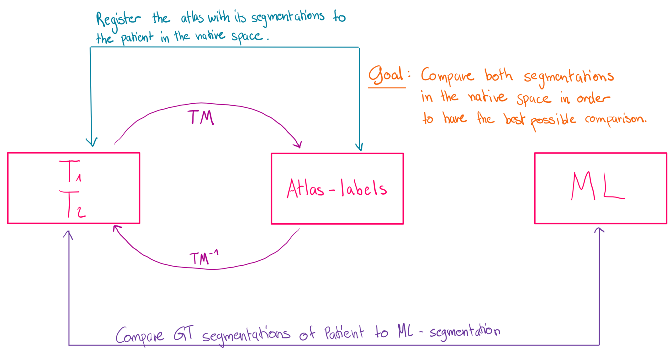

Fig. 1. Schematic of Comparison Approach

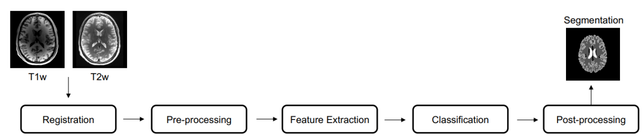

Fig. 2. medical image analysis pipeline

analysis pipeline can be seen in figure 2. Five critical steps are included; Registration, Preprocessing, Feature Extraction, Classification, and Post-Processing. However, not all of these steps were used in the two approaches.

B. Machine-Learning-based approach

The machine-learning algorithm is based on the *scikit-learn* library (version 1.0.2). It also relies on *SimpleITK* (version 2.1.1.2). The random forest classifier, a supervised learning algorithm created by Leo Breiman in 2001 [2]. Aligned with the medical image analysis pipeline, the images are loaded and preprocessed. This includes skullstripping, normalization, coordinate-, intensity-, and gradient-intensity feature extraction but not registration, since the ground truth is already in the native space. The features from the pre-processed images are extracted and used to build a random forest classifier model. The testing occurs by crawling through the testing image directories and pre-processing these by using the trained model. The trained model makes predictions and obtains probabilities which are stored within a CSV file. While post-processing is a part of the pipeline, the images were not post-processed.

C. Atlas-based approach

Using 20 patient images, a dictionary of the atlas labels was created. These labels were subsequently used to create the atlas. Before testing, the images were loaded and preprocessed. As part of the pre-processing, skullstripping, normalization, coordinate feature extraction, and registration to the native space was applied, but not intensity- or gradient intensity feature extraction. The combined transformed atlas labels are collected into a single label image and the resulting probabilities are stored in a CSV file. Within the probability interval, a specific threshold needs to be chosen to establish the tissue boundary.

D. Evaluation

For both, the atlas-based and the machine learning approach, the parameters needed finetuning to create the optimal results. In the case of machine learning, this required optimizing the parameters of the random forest algorithm and finding the best-suited amount of estimators and depth. For the atlas-based approach finding the best results for the best probability between 0 and 1. Both approaches were evaluated in the native space compared to the ground truth and were compared using the DICE score and the Hausdorff distance.

III. Results

The results are divided into machine learning segmentation and atlas label registration. These results were obtained on ten test images. The Hausdorff distance and the dice score were used to evaluate each case's results. These scores are visualized in a boxplot representing the performance of both approaches with optimized parameters. The process of parameter optimization can be seen in figure 3 and 4.

In a first step, optimal parameters for the machine learning approach were evaluated. To find optimal parameters for the estimators and the depth of the random forest classifier, first, the estimators were fixed to find a depth at which decent results were achieved. The result of this process can be seen in figure 3. Here the mean values of the Dice and Hausdorff scores for varying depths at a fixed number of estimators are visible. The estimators are set to a value of 7, while the depth starts at a value of 1 and goes up to a value of 70.

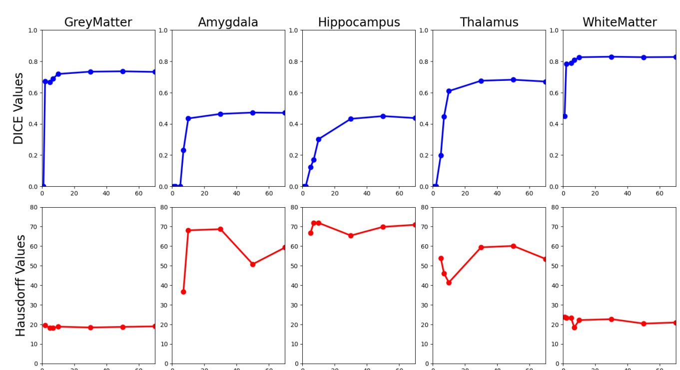

Fig. 3. Fixation of estimators at a value of 7 with varying depth

The optimization of the estimators can be seen in figure 4. As seen in figure 3, the performance of the machine learning segmentation no longer improves after the depth of 30. The estimators were then optimized at a fixed depth of 30. The estimators range from 1 to 100 with varying distances from each other. The best results were achieved with 70 estimators.

To compare machine learning to the atlas approach, the optimal parameter for the atlas had to be found. The performance of the atlas label segmentation depends on the boundary level that has been chosen. The level takes a value between 0 and 1.

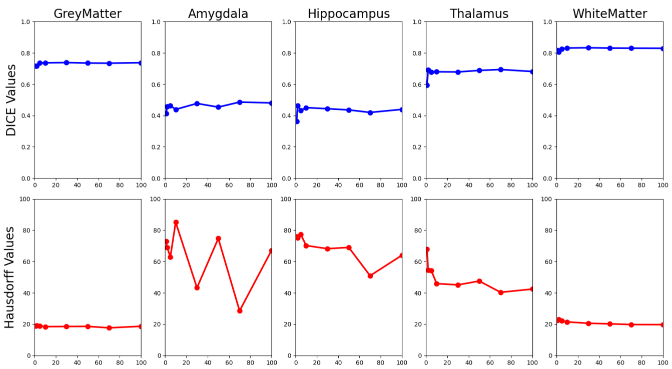

Fig. 4. Fixation of depth at a value of 30 with varying numbers for estimators

It defines a certainty at which we decide to have the boundary of a tissue. To be more precise, at a level of 0.4, we are 40 percent certain that the individual tissue is present. Having the boundary set at different values and comparing the resulting Dice and Hausdorff scores leads to the figure5. Analyzing the plots of the atlas label segmentation performance lead to the decision to take a boundary level of 0.35.

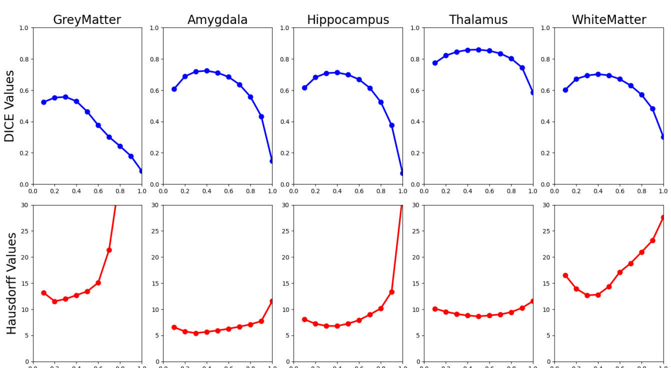

Fig. 5. Comparison of boundaries for atlas segmentation approach

Having all the parameters, influencing the performance of the machine learning and the atlas segmentation approaches, optimized, the two can be compared. The comparison of the two enables the conclusion of the present work. The figure comparing the two approaches can be seen in 6. The parameters used were 70 estimators and a depth of 30 for the ML approach, and a boundary level of 0.35 for the atlas approach.

The segmentation of the ML approach can be seen in figure 7. The segmentation was chosen randomly out of the test images and then visualized in ITKSnap.

The color scheme for the segmentation is as follows; green = GrayMatter, red = WhiteMatter, yellow = Hippocampus, light blue = Thalamus, blue = amygdala. The ground truth segmentation of the same patient can be seen in figure 8

IV. Discussion

Atlas-based segmentation compared to the ML approach, has its strength in the low Hausdorff distances. A low score

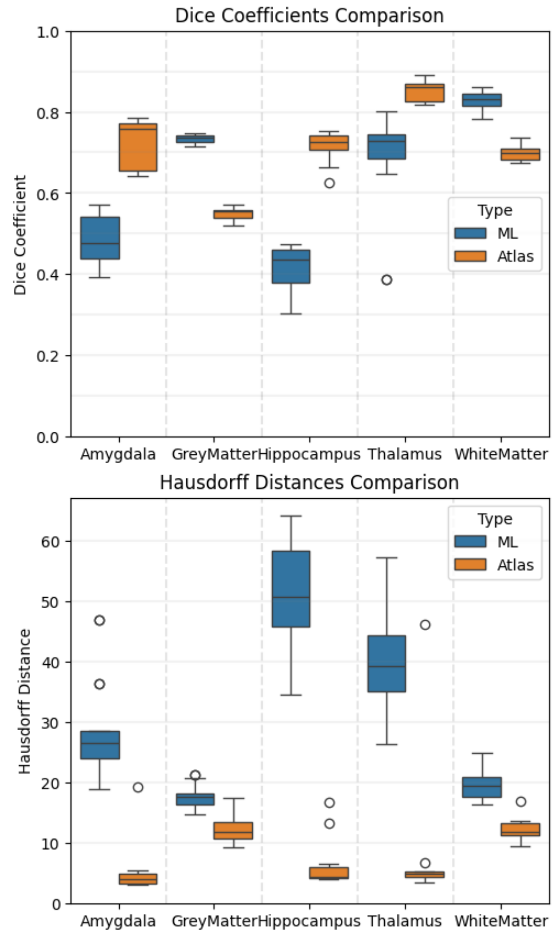

Fig. 6. Comparison of ML approach vs Atlas-label registration

in the Hausdorff measurement is a sign of stability against outliers. This is to be expected from an atlas label approach, as the labels act like a sort of ground truth of the standard brain. Registering these labels then ensures the prevention of outliers with big distances to the actual tissue. The low Hausdorff scores of the Atlas approach can be seen in figure 6.

In contrast to the Atlas-based segmentation, the machine learning approach was able to show better segmentation performance in the areas of White matter and Gray matter. The reason behind the better segmentation outcomes for the two tissues may lie in the preprocessing methods and feature extraction employed within the machine learning approach. During preprocessing a brain mask is applied, stripping the image of everything but the brain tissue. The remaining image then shows a strong edge from the outer border of the Gray matter to the background. This edge then can easily be detected by the feature extraction algorithm, hence, facilitating the correct segmentation of the gray matter tissue. The machine learning approach similarly gains an advantage over the atlas approach when detecting white matter, as it again, excels in the presence of a strong edge. The presence or absence of a strong edge defines the performance of the approach. This can also be seen when looking at the results of the remaining three

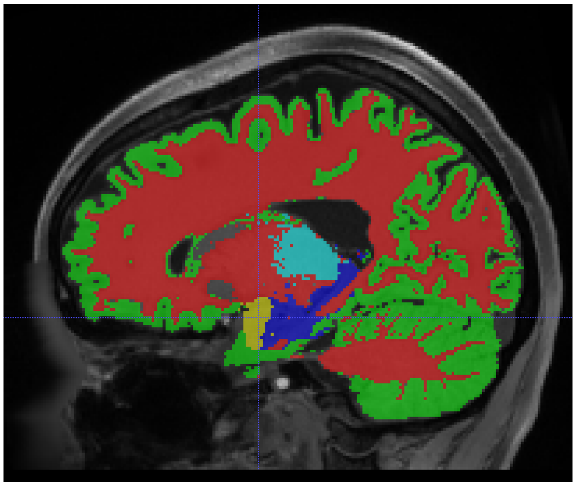

Fig. 7. Sagital view of the ML segmentation

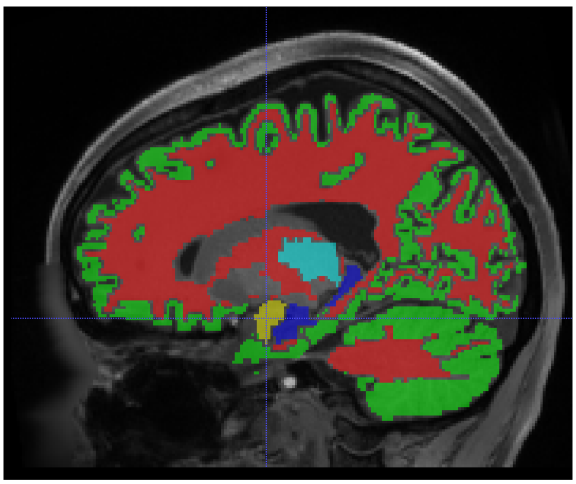

Fig. 8. Sagital view of the Ground truth segmentation

tissues, namely the Amygdala, Hippocampus, and Thalamus, where the strong edge is not present.

Another explanation for the obtained results would be the characteristics of the anatomy itself. While structures like the Amygdala, the Hippocampus, and the Thalamus are in all individuals more or less in the same position, the way the cerebral cortex is shaped with its sulci and gyri, is, however, very different for each person. Therefore relying on a segmentation approach, which uses the average of a structure with a big variability, is hence inferior regarding its performance for the given structure, compared to an approach that uses the characteristics of the given image.

The Hausdorff scores of the machine learning segmentation approach, show values that are ten-fold the one of the atlas approach, for some structures. The structures of concern are the Hippocampus and the Thalamus. A visual comparison of the two structures can be seen in figure9. These images show how the machine-learning approach overestimates the volume of the Hippocampus, the Thalamus, and the Amygdala. Moreover, the ML approach encounters difficulty in identifying these structures, leading to a segmentation mask that resembles a square shape. However, its performance may be explained through the amount of training images. Training on many more images could improve the performance.

The Atlas-based approach performed well on structures of low variability. All of the training images were of healthy patients, it can be assumed that the atlas-based approach would struggle severely when the brain anatomy is altered due to pathologies such as brain lesions or atrophies. A multi-atlas approach could aid in accounting for the anatomical variations [3].

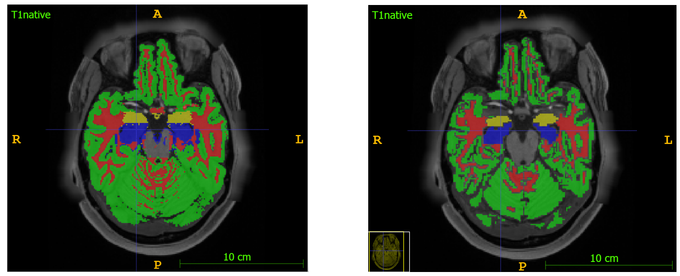

Fig. 9. Comparison of the segmentations in the transversal plane. Left = ML prediction, right = Ground truth

The dice scores of the atlas-based approach are on average superior to the ones of the machine learning approach. However, when setting an acceptance level at a score of 0.7 for dice, both approaches succeeded only in three out of five classes.

V. Conclusion

Both segmentation approaches have their strengths and weaknesses. The machine learning-based segmentation approach has shown solid results in the segmentation of the tissues White matter and Gray matter. The dice score for these two tissues has averaged a value of 0.78, indicating good coverage between prediction and ground truth. For structures with less variability like the Hippocampus, the Amygdala, and the Thalamus, the Atlas-based approach has shown superiority, in both dice score and Hausdorff distance.

The hypothesis that Atlas-based segmentation consists of a powerful baseline for brain tissue segmentation when compared to an ML-based approach, can hence, not entirely be supported. To fully support the hypothesis it had to be changed and specified for which tissues it would be used. An alternative hypothesis would be; Atlas-based segmentation, for regions with a lower variability, consists of a powerful baseline for brain tissue segmentation when compared to an ML-based approach.

To improve the segmentation results, the two approaches could be combined, as was done in the work of Yaakub et al. [8]. A majority voting between the two, giving the structures adapted weights according to the dice scores seen in figure6, would likely lead to improved results. The strengths of both approaches could be combined by integrating them in the stated manner.

ACKNOWLEDGMENT

We would like to thank our tutors, Amith Kamath, Michael Müller, and Shelley Shu for their support throughout this project.

REFERENCES

[1] Sang Hoon Ahn et al., "Comparative clinical evaluation of atlas and deep-learning-based auto-segmentation of organ structures in liver cancer," vol. 14, no. 1, Nov. 2019, doi: https://doi.org/10.1186/s13014-019-1392-z.

[2] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

[3] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. Bach Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," Computer Methods and Programs in Biomedicine, vol. 104, no. 3, pp. e158–e177, Dec. 2011, doi: https://doi.org/10.1016/j.cmpb.2011.07.015.

[4] W. Chen et al., "Deep learning vs. atlas-based models for fast auto-segmentation of the masticatory muscles on head and neck CT images," Radiation Oncology, vol. 15, no. 1, Jul. 2020, doi: https://doi.org/10.1186/s13014-020-01617-0.

[5] M. Costea et al., "Comparison of atlas-based and deep learning methods for organs at risk delineation on head-and-neck CT images using an automated treatment planning system," Radiotherapy and Oncology, vol. 177, pp. 61–70, Dec. 2022, doi: https://doi.org/10.1016/j.radonc.2022.10.029.

[6] B.W.K. Schipaanboord et al., "An Evaluation of Atlas Selection Methods for Atlas-Based Automatic Segmentation in Radiotherapy Treatment Planning," IEEE Transactions on Medical Imaging, vol. 38, no. 11, pp. 2654–2664, Nov. 2019, doi: https://doi.org/10.1109/tmi.2019.2907072.

[7] M. K. Singh and K. K. Singh, "A Review of Publicly Available Automatic Brain Segmentation Methodologies, Machine Learning Models, Recent Advancements, and Their Comparison," Annals of Neurosciences, p. 097275312199017, Mar. 2021, doi: https://doi.org/10.1177/0972753121990175.

[8] S. N. Yaakub, R. A. Heckemann, S. S. Keller, C. J. McGinnity, B. Weber, and A. Hammers, "On brain atlas choice and automatic segmentation methods: a comparison of MAPER and FreeSurfer using three atlas databases," Scientific Reports, vol. 10, no. 1, Feb. 2020, doi: https://doi.org/10.1038/s41598-020-57951-6.