Zero-Shot Skull Stripping for Alzheimer's Disease Classification With the Segment Anything Model

Jean Christopher Pinheiro[1], **Vitória Chrissie de O. Pinheiro**[2] , **Rodrigo Silva**[1], **Gladston Moreira**[1], **Pedro H. L. Silva**[1], **Eduardo J. S. Luz**[1]

[1]Computing Department – Federal University of Ouro Preto (UFOP)
35.400-000 – Ouro Preto – MG – Brazil

[2]Medical School – Federal University of Minas Gerais (UFMG)
31.270-901 – Belo Horizonte – MG – Brazil

jean.christopher@aluno.ufop.edu.br

vitoriachrissie@ufmg.br

{rodrigo.silva,gladston,silvap,eduluz}@ufop.edu.br

Abstract. *This study explores magnetic resonance imaging segmentation for Alzheimer's disease by automating skull stripping using the Segment Anything Model (SAM), a zero-shot segmentation model. The challenge lies in selecting the correct mask generated by the SAM, for which we propose a heuristic based on templates to identify the optimal choice. This method presents a practical alternative to the traditional FMRIB Software Library Brain Extraction Tool. The effectiveness of our approach is indirectly assessed using Alzheimer's disease classification as a proxy task. Validation is conducted using the Alzheimer's Disease Neuroimaging Initiative dataset, demonstrating a 6% improvement in classification accuracy with the zero-shot approach.*

1. Introduction

Magnetic Resonance Imaging (MRI) is essential in medical diagnostics, particularly for identifying structural changes in Alzheimer's disease, aiding early diagnosis and treatment [Lee et al. 2019, Mofrad et al. 2021]. The presence of the skull in MRI scans can obscure brain tissue, necessitating effective skull stripping for accurate analysis [Oh et al. 2019, Pei et al. 2022, El-Baz et al. 2016]. Manual segmentation of MRI brain volumes is time-intensive, ranging from 15 minutes to 2 hours [Smith 2002], and impractical for large datasets, highlighting the need for automation [Varoquaux and Cheplygina 2022]. Skull stripping methods include traditional techniques, such as those in the FMRIB Software Library (FSL) [Jenkinson et al. 2012] and FreeSurfer [Perlaki et al. 2017, Quilis-Sancho et al. 2020], which are sensitive to noise and vary in efficiency based on the images and regions analyzed [Quilis-Sancho et al. 2020, Mohapatra et al. 2023]. However, recent machine learning approaches, particularly zero-shot learning models like the Segment Anything Model (SAM) [Kirillov et al. 2023], offer robust, noise-resistant results without the need for extensive data or computational resources [Azam and Tariq 2020].

The SAM [1] employs zero-shot segmentation, allowing image segmentation without

[1]https://segment-anything.com

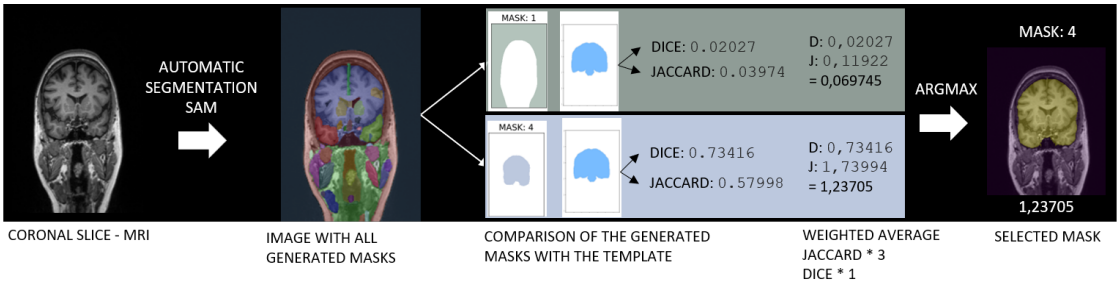

Figure 1. Workflow of the SAM for Alzheimer's disease classification: The process begins with a coronal MRI slice, where SAM segments the brain into regions, producing multiple masks. Selecting the right mask presents a significant challenge, and a heuristic is proposed here to address this issue.

specific pre-training. This generalization capability could enhance MRI image segmentation and help with Alzheimer's disease classification by effectively removing the skull, vertebrae, and neck from brain images. Traditional skull stripping methods often struggle with variability across scans and require significant adaptations, complicating clinical applications where robustness across brain morphologies is crucial [Kalavathi and Prasath 2016]. This study hypothesizes that SAM could offer a practical and competitive alternative to the FSL Brain Extraction Tool (BET) by providing effective neuroimaging analysis for Alzheimer's disease, reducing manual intervention, and better accommodating imaging variations. However, SAM generates multiple masks for the same image, and selecting the best mask can be challenging.

To test the proposed hypothesis, experiments are conducted using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset.[2] In this study, we address the challenge of selecting the most suitable mask from multiple SAM-generated candidates for the same image by introducing a heuristic for post-processing, named SMASH (SAM Mask Selection Heuristic), as illustrated in Figure 1. The experiments involve comparing masks from SAM+SMASH, masks from the FSL, and the original unmodified images across a sample of 303 individuals—204 without and 109 with Alzheimer's disease—to assess the effectiveness of skull stripping in classifying Alzheimer's disease, a proxy task. The classification task is performed using a convolutional neural network model. Qualitative analysis shows high-quality skull extraction by SAM+SMASH, while quantitative assessments reveal competitive performance against established methods. The application of SAM-enhanced masks, refined with SMASH, leads to improved classification accuracy: 70% using SAM masks compared to 66% with FSL-BET masks. These results demonstrate the efficacy of SAM in improving mask quality and, consequently, Alzheimer's disease classification accuracy.

2. Related Works

Traditional Methods. The FSL (FMRIB Software Library) and Freesurfer have been regarded as the gold standard in the domain of skull stripping [Mohapatra et al. 2023]. The FSL's Brain Extraction Tool (BET) and Freesurfer's segmentation utilities utilize sophisticated algorithms based on deformable models. These methods have been robust in handling the complexities associated with the task, though they each present unique

[2] adni.loni.usc.edu

advantages and limitations regarding computational efficiency and segmentation fidelity [Fatima et al. 2020]. For instance, while FSL is known for its faster processing times, Freesurfer is recognized for its detailed segmentation, particularly of subcortical structures, which are crucial for comprehensive neuroimaging analyses.

Deep Learning-based Methods. The advent of deep learning has introduced innovative approaches that potentially surpass the capabilities of traditional algorithms. Techniques leveraging 3D Convolutional Neural Networks and architectures such as U-Net have demonstrated exceptional precision and efficiency in segmenting brain tissues from MRI scans. These methods capitalize on large annotated datasets to train models that effectively generalize across various imaging conditions [Fatima et al. 2020, Kleesiek et al. 2016]. Nonetheless, these advanced techniques face significant hurdles in evaluation standardization, which is vital for their adaptation in clinical settings.

Several studies have used the ADNI dataset for Alzheimer's disease classification with slice-based approaches, achieving high accuracies with advanced CNN models: Farooq et al.[Farooq et al. 2017] reached 98.2% accuracy using a deep CNN for multi-class classification, Gunawardena et al.[Gunawardena et al. 2017] reported 96.6% with a ResNet-based model for binary classification, and Luo et al.[Luo et al. 2017] achieved 94.3% using a combined CNN and RNN approach. These studies employed various preprocessing techniques, making direct comparisons challenging due to the absence of standardized evaluation protocols.

In contrast, Mohapatra et al.[Mohapatra et al. 2023] evaluated the SAM and the FMRIB Software Library's Brain Extraction Tool (BET) for brain extraction from MRI images, using standardized preprocessing and focusing on whole-brain segmentation. Their study, which employed 45 MRI images from different datasets aligned to MNI152 space, assessed performance using several metrics, including the Dice coefficient and Jaccard Index. Our study differs by using raw MP-RAGE images from the ADNI dataset focused on slice-based segmentation near the hippocampus, without any preprocessing, to directly evaluate SAM's zero-shot segmentation capabilities in classifying Alzheimer's disease.

3. Methodology

This section outlines the methodology and is divided into segmentation and classification phases. Initially, segmentation involves using the Segment Anything Model and the SMASH heuristic to remove the skull from MRI images and isolate the brain. This ensures that analyses are focused solely on brain tissue for accurate Alzheimer's disease detection. Subsequently, the brain-only images are used in the classification phase, where a Convolutional Neural Network differentiates between Alzheimer's patients and normal controls by identifying disease-specific features.

3.1. Alzheimer's Disease Neuroimaging Initiative Dataset

The dataset for this study was sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (`adni.loni.usc.edu`) [Petersen et al. 2010], initiated in 2003 as a public-private partnership led by Michael W. Weiner, MD. The primary goal of ADNI is to enable the integration of serial magnetic resonance imaging (MRI), positron emission tomography (PET), biological markers, and clinical and neuropsychological assessments to monitor the progression of mild cognitive impairment (MCI) and early-stage Alzheimer's

disease. For this work, 313 MRI files from ADNI were analyzed and divided between 204 participants in the normal control (NC) group and 109 participants in the Alzheimer's disease (AD) group.

This study deliberately excludes common preprocessing steps such as bias correction with the N3 algorithm, removal of unwanted structures, and resampling, despite the recognized benefits of preprocessing MRI for improving image quality [Shi et al. 2018, Suk et al. 2014]. The efficacy of segmentation methods applied directly to raw images is explored by focusing on segmentation techniques, allowing for an evaluation without preprocessing influences. Only raw image files obtained by magnetization-prepared rapid gradient echo (MP-RAGE) are used, ensuring that preprocessing variables do not affect the outcomes. All files are converted from Digital Imaging and Communications in Medicine (DICOM) format to Neuroimaging Informatics Technology Initiative (NIFTI) format using MRIcroGL software[3].

Data Preparation and Slices. We adopt a strategic approach to select specific slices from the coronal plane, focusing on those aligned with or near the hippocampus, which is paramount in Alzheimer's disease detection [Scheltens et al. 1992]. We opt for 10 slices chosen at regular intervals of two slices along the y-axis, commencing from slice number 122 and extending to 140. This decision is driven by the necessity for a direct and detailed visualization of the hippocampus, given its crucial role in disease pathology. Moving in increments of two slices, we aim to cover a broader range of anatomical variations, thereby ensuring a comprehensive representation of the hippocampal region across the selected slices.

3.2. Segmentation Methodology

This section details the methodology used for segmenting brain MRIs.

Segmentation with FSL BET. The FSL BET (Brain Extraction Tool) segments one 3D volume in NIfTI format at a time. To expedite processing, 30 files were segmented simultaneously using parallel programming with an intensity threshold of 0.6. This threshold effectively identified the region of interest, though it sometimes removed parts of the brain or failed to remove skull regions completely. Several parameters from the FSL BET user guide[4] were used, including the "robust" mode to handle intensity variations and the "mask" parameter to generate a segmentation mask. Importantly, the segmentation process did not involve preprocessing techniques, ensuring that both methods had an equal starting point. Consequently, several NIfTI format files were generated, including those displaying the region of interest and the corresponding masks, allowing for a direct comparison of segmentation quality and effectiveness under varying conditions without preprocessing enhancements.

SAM Mask Selection Heuristic: SMASH. To effectively segment selected slices using the SAM, it is essential to adapt the input data to the format expected by the model. Importantly, our methodology diverges from conventional uses of SAM as it does not utilize prompts or bounding boxes to direct the segmentation process. Instead, SAM can freely generate a comprehensive set of masks for each slice, ensuring a full exploration of

[3]https://www.nitrc.org/projects/mricrogl

[4]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET/UserGuide

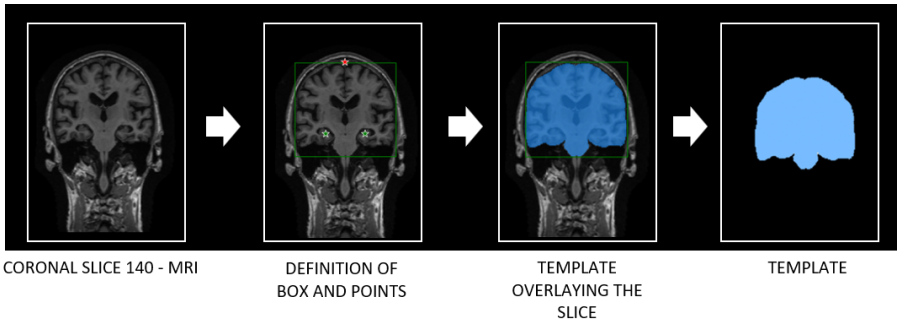

Figure 2. Steps in the Template Creation Process using Coronial Slice 140 from Individual *ADNI_007_S_1304*.

possible segmentations without predefined areas of interest (ROI).

SAM is capable of high-precision object identification within any given image, demonstrating a significant *zero-shot* generalization capability for segmentation tasks [Kirillov et al. 2023]. The architecture of SAM is detailed in Figure 3 and includes the following key components:

- **Image Encoder:** This component processes each image once, handling high-resolution images efficiently. It utilizes a Vision Transformer (ViT) pre-trained with a Masked Autoencoder (MAE), optimizing image representation by focusing on essential regions [Dosovitskiy et al. 2021, He et al. 2021, Kirillov et al. 2023].

$$E_I = \text{ViT}(\text{MAE}(I)). \tag{1}$$

- **Prompt Encoder:** Typically, this would process sparse or dense prompts to refine the segmentation focus. However, this step is bypassed in our approach to allow SAM's architecture to generate several segmentations without external cues autonomously.

$$E_P = \begin{cases} \text{CLIP}(P) & \text{if } P \text{ is text (not used)} \\ \text{Conv}(P) & \text{if } P \text{ is a mask (not used)} \end{cases} \tag{2}$$

- **Mask Decoder:** It decodes the image features combined with any prompts into a set of actionable masks. Given our approach bypasses prompt inputs, the decoder focuses solely on the features derived from the image encoder, producing a diverse array of segmentation masks.

$$\{M_i\} = \text{Transformer}(E_I, E_P \text{ (unused)}). \tag{3}$$

After the inference process, SAM generates multiple masks for each slice. Our methodology involves reviewing these masks to select the most closely resembles a predefined template, utilizing a heuristic based on similarity metrics.

Methodological Steps for SMASH:

1. **Mask Generation:** SAM utilizes its deep learning architecture to generate multiple masks for each MRI slice. This process accommodates variations in brain anatomy and imaging characteristics, essential for precise segmentation.

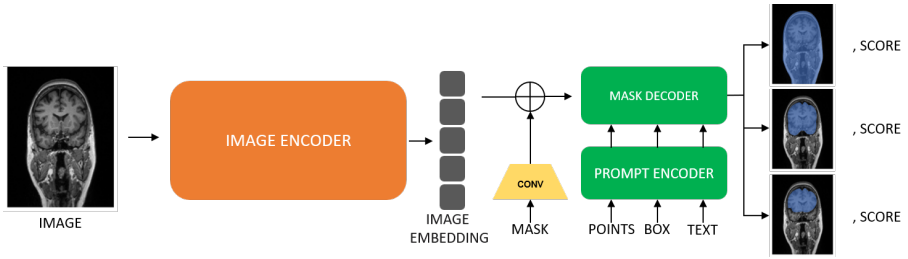

Figure 3. SAM architecture. Adapted from [Kirillov et al. 2023].

2. **Definition of Similarity Metrics:**
   To evaluate the similarity between each generated mask (m) and the template (T), two metrics are used: the Dice Metric, defined as $D(m,T) = \frac{2 \times |m \cap T|}{|m| + |T|}$, which quantifies the precision of overlap between the mask and the template, and the Jaccard Metric, expressed as $J(m,T) = \frac{|m \cap T|}{|m \cup T|}$, which emphasizes the total coverage of the template by the mask.
3. **Calculation of Combined Score:** The combined score for each mask is calculated using the equation:

$$P(m,T) = 3 \times J(m,T) + D(m,T), \tag{4}$$

   wherein a weight of 3 is assigned to the Jaccard metric to prioritize masks with greater area overlap, and a weight of 1 is given to the Dice metric to evaluate the accuracy of the overlap.
4. **Selection of Optimal Mask:** The optimal mask is selected based on the highest combined score:

$$m^* = \arg\max_{m \in M} P(m,T), \tag{5}$$

   where M is the set of all masks produced by SAM for a specific slice. This step ensures the mask selection that most accurately matches the template, aiding precise analysis of critical brain regions.

In this methodology, a template representing the desired characteristics of the final mask is created, as shown in Figure 2. Initially, a manual mask is crafted based on slice 140 of an MRI scan from "ADNI_007_S_1304," a 75-year-old female diagnosed with Alzheimer's, under the guidance of a specialist. The use of a single template is based on the assumption that SAM can automatically generate various masks, including one that closely resembles the brain without the skull. Therefore, having an approximate template should suffice to identify the most similar mask, allowing the template to underpin the proposed heuristic. This approach is classified as a greedy heuristic because it relies on making the best local decisions (selecting the mask with the highest score) to achieve a globally satisfactory result. The formalization of the SMASH heuristic is depicted in Algorithm 1.

3.3. Classification Methodology

The challenge in the quantitative evaluation of masks generated for the segmentation of magnetic resonance imaging lies in the absence of validated masks available in public datasets (*ground truth*). Consequently, the quality of these masks is indirectly assessed

Algorithm 1 Heuristic for Selecting Masks Segmented by SAM (SMASH)

Data: Set of MRI slices, SAM model, template T
Result: Selected masks M^*
begin
  foreach *slice* s *in slices* **do**
    // Step 1: Segment slices with SAM
    $M \leftarrow$ SAM.segment(s) // Generate masks $M = \{m_1, m_2, \ldots, m_k\}$
    // Step 2: Initialize lists of metrics and scores
    $scores \leftarrow []$ // List of scores
    $dice\_coefficients \leftarrow []$ // List of Dice coefficients
    $jaccard\_similarities \leftarrow []$ // List of Jaccard similarities
    foreach *mask* m *in* M **do**
      // Step 3: Calculate Dice and Jaccard metrics
      $D(m,T) \leftarrow \frac{2\times|m\cap T|}{|m|+|T|}$ // Dice metric
      $J(m,T) \leftarrow \frac{|m\cap T|}{|m\cup T|}$ // Jaccard metric
      // Store the metrics
      $dice\_coefficients \leftarrow dice\_coefficients \cup [D(m,T)]$
      $jaccard\_similarities \leftarrow jaccard\_similarities \cup [J(m,T)]$
      // Step 4: Calculate the combined score with fixed weights
      $P(m,T) \leftarrow \frac{1\times D(m,T)+3\times J(m,T)}{2}$
      $scores \leftarrow scores \cup [P(m,T)]$
    end
    // Step 5: Select the mask with the highest score
    $selected\_mask\_index \leftarrow \arg\max(scores)$
    $m^* \leftarrow M[selected\_mask\_index]$
    // Store the optimal mask for the current slice
    $M^* \leftarrow M^* \cup \{m^*\}$
  end
  return M^*
end

through their performance in secondary tasks. In this study, we adopted the strategy of training a neural network to investigate the impact of segmentation on the classification of images related to Alzheimer's disease. For this purpose, a Convolutional Neural Network (CNN) is implemented, given its recognized efficacy in the literature for such analyses.

Neural Network Architecture. The neural network architecture in this study is developed through an experimental, incremental process. It begins with a basic network to establish a performance baseline and assess computational complexity. Through iterative evaluation and adjustment, the architecture gradually incorporates more layers and neurons, closely monitoring their impact on performance and computational costs. This refinement process primarily evaluates the network's performance on raw data, focusing on segmentation. After extensive adjustments, the final architecture, as shown in Figure 4, is finalized. This study centers on Alzheimer's classification as a case study for validating the MRI slice segmentation method, exploring advanced neural architectures for Alzheimer's classification beyond its scope.

Classification Metrics. To clarify the impact of the proposed methodology, we computed three key metrics: accuracy, precision, and F1-score, each ranging from zero to one. An accuracy of zero indicates that all classifications are incorrect, whereas an accuracy of one denotes that all classifications are correct. Accuracy is determined by dividing the

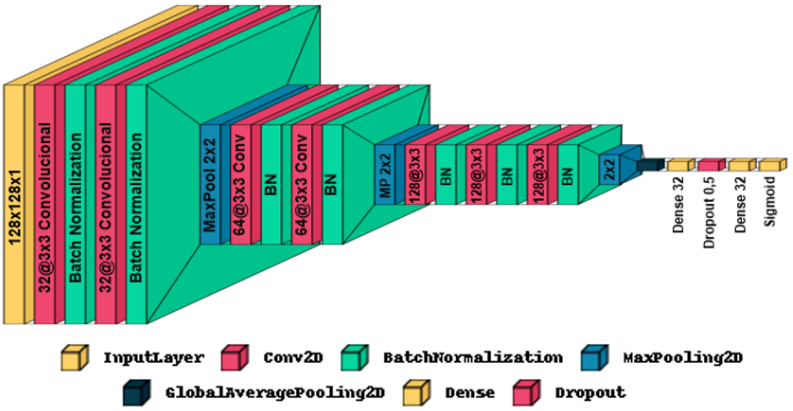

Figure 4. Diagram of the neural network architecture for Alzheimer's classification. It includes an input layer, several convolutional layers (Conv2D) with Batch Normalization and MaxPooling2D, a Global Average Pooling layer for dimension reduction, Dense layers with Dropout for regularization, and a sigmoid activation function for binary output.

number of correct classifications by the total number of samples. Precision measures the proportion of correctly classified Alzheimer's disease cases among all instances predicted as Alzheimer's disease. The F1-score, which is the harmonic mean of precision and recall, balances the precision with recall, where recall represents the ratio of correctly predicted Alzheimer's disease cases to the total number of actual Alzheimer's disease cases.

4. Results and Discussion

Two types of experiments address this work's central hypothesis. Experiment #1 investigates the SAM for segmenting medical images alongside the SMASH heuristic for mask selection. This experiment also explores whether the outcomes achieved with SAM+SMASH are on par with those obtained using the FSL BET in the task of segmenting brain MRIs without preprocessing. The analysis is qualitative since the ADNI dataset does not provide ground truth masks.

In Experiment #2, a model is trained using images segmented by FSL BET and SAM+SMASH to classify them in Alzheimer's disease problems. This task serves as quantitative validation for the segmentation provided by SAM. The comparison is made solely against FSL BET since it is considered the gold standard for MRI image segmentation. Furthermore, due to the lack of standardization in the literature, direct comparisons with other techniques applied to the ADNI dataset are not feasible due to differences in evaluation protocols and preprocessing techniques.

4.1. Experimental Settings

The experiments are conducted on a machine equipped with an AMD Ryzen Threadripper, 120GB RAM, and a GeForce RTX 3090. Libraries such as Nibabel, Numpy, and OpenCV, along with FSL software version 6.0 and Python's subprocess package, are used for handling Nifti files and other tasks. The original implementation of SAM is used [5].

4.2. Experiment #1 - Qualitative Evaluation

To analyze the results from the SAM+SMASH method, a subset of data from the ADNI dataset is selected to evaluate its successes and challenges. The resulting masks are overlaid

[5]https://github.com/facebookresearch/segment-anything

on the original images for visual comparison, with masks from the FSL BET tool displayed in blue and those from SAM+SMASH in green. This comparison, illustrated in Figure 5, highlights the differences in segmentation coverage: BET tends to be more conservative, while SAM often provides broader coverage, potentially capturing more brain tissue.

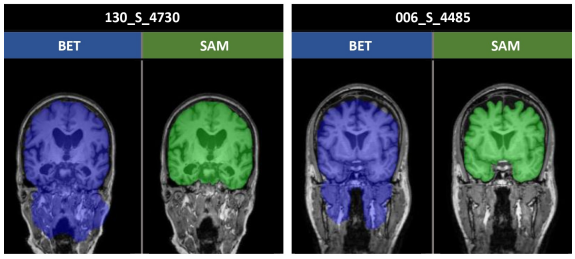

Figure 5. Visual Comparison of MRI Segmentation Results between FSL BET and SAM: This figure displays a side-by-side comparison of MRI slice segmentation results using the FSL Brain Extraction Tool (BET) and the Segment Anything Model (SAM). Each pair of images represents a single slice from the ADNI dataset, labeled by its identifier. BET results are illustrated in blue, while SAM results are shown in green.

4.3. Experiment #2 - Quantitative Results

This experiment quantitatively evaluates three different approaches for inputting images into a CNN: (i) raw images (RAW), (ii) images pre-segmented by FSL BET, and (iii) slices segmented by SAM refined by a template selection heuristic (SMASH). A learning rate of 0.0001, early stopping with patience of 30, a total of 200 epochs, a batch size of 36, and the ADAM optimizer are used for network training. These approaches are trained and evaluated using a binary classification (normal or Alzheimer's) with images from the ADNI dataset to determine the impact of segmentation on Alzheimer's disease classification. Results are presented in Table 1, analyzing the benefits of each segmentation technique and their effects on diagnostic accuracy, it is possible to observe that the strategy employing SAM+SMASH overcome the other strategies. The primary goal of this work is to assess the impact of segmentation rather than achieving state-of-the-art classification accuracy.

Table 1. Quantitative results for the three approaches of experiment 2. NC = normal control group; AD = Alzheimer's disease group.

Metrics	RAW		FSL - BET		SAM	
	NC	AD	NC	AD	NC	AD
F1 Score	0.6834	0.2723	0.7707	0.3895	**0.7978**	**0.4418**
Precision Score	0.6410	0.3210	0.6976	0.5403	**0.7165**	**0.6435**
Accuracy	0.5587		0.6667		**0.7032**	

The confusion matrix related to the reported accuracy in Table 1 is presented in Figure 6.

5. Conclusion

This study explored the potential of the SAM for zero-shot segmentation of brain MRI images to improve Alzheimer's disease classification. Our approach involved selecting

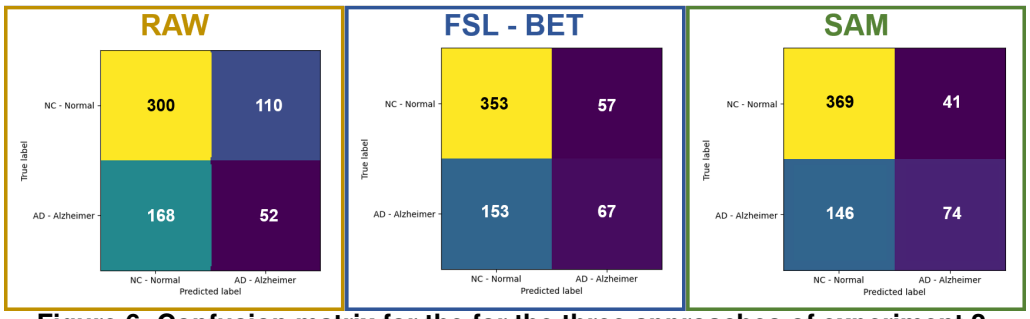

Figure 6. Confusion matrix for the for the three approaches of experiment 2.

relevant slices, creating a template for mask evaluation, segmenting images with SAM, using a template-based heuristic (SMASH) for mask selection, and employing a CNN for classification. The experiments demonstrated SAM's feasibility for brain MRI segmentation, achieving improved classification performance within the ADNI dataset (77% ACC vs 66% ACC with FSL BET).

However, the reliance on a one-size-fits-all template has limitations due to biological variability, which may result in the loss of masks. Future research should explore using multiple masks and different scales and potentially a range of heuristics for better template selection to enhance segmentation accuracy. Despite these limitations, the method offers significant advantages like simplicity and ease of implementation, making it adaptable to various medical segmentation contexts.

Acknowledgments

The authors would also like to thank the *Fundação de Amparo a Pesquisa do Estado de Minas Gerais* (FAPEMIG, grants APQ-01518-21, APQ-01647-22), *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq, grants 307151/2022-0, 308400/2022-4) and *Universidade Federal de Ouro Preto* (PROPPI/UFOP) for supporting the development of this study.

References

Azam, H. and Tariq, H. (2020). Skull stripping using traditional and soft-computing approaches for magnetic resonance images: a semi-systematic meta-analysis. *Machine Graphics and Vision*, 29(1/4):33–53.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale.

El-Baz, S., Soliman, F., and Gimel'farb, G. (2016). A survey of brain mri image segmentation methods and the issues involved. *Magnetic Resonance Imaging*, 34(10):1300–1318.

Farooq, A., Anwar, S., Awais, M., and Rehman, S. (2017). A deep cnn based multi-class classification of alzheimer's disease using mri. In *2017 IEEE International Conference on Imaging systems and techniques (IST)*, 1–6. IEEE.

Fatima, A., Shahid, A. R., Raza, B., Madni, T. M., and Janjua, U. I. (2020). State-of-the-art traditional to the machine-and deep-learning-based skull stripping techniques, models, and algorithms. *Journal of Digital Imaging*, 33:1443–1464.

Gunawardena, K., Rajapakse, R., and Kodikara, N. D. (2017). Applying convolutional neural networks for pre-detection of alzheimer's disease from structural mri data. In *2017 24th international conference on mechatronics and machine vision in practice (M2VIP)*, 1–7. IEEE.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. (2021). Masked autoencoders are scalable vision learners.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2):782–790.

Kalavathi, P. and Prasath, V. S. (2016). Methods on skull stripping of mri head scan images—a review. *Journal of digital imaging*, 29:365–379.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026.

Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., and Biller, A. (2016). Deep mri brain extraction: A 3d convolutional neural network for skull stripping. *NeuroImage*, 129:460–469.

Lee, G., Nho, K., Kang, B., et al. (2019). Predicting alzheimer's disease progression using multi-modal deep learning approach. *Scientific Reports*.

Luo, S., Li, X., and Li, J. (2017). Automatic alzheimer's disease recognition from mri data using deep learning method. *Journal of Applied Mathematics and Physics*, 5(9):1892–1898.

Mofrad, S. A., Lundervold, A. J., Vik, A., et al. (2021). Cognitive and mri trajectories for prediction of alzheimer's disease. *Scientific Reports*, 11:2122.

Mohapatra, S., Gosai, A., and Schlaug, G. (2023). Sam vs bet: A comparative study for brain extraction and segmentation of magnetic resonance images using deep learning.

Oh, K., Chung, Y. C., Kim, K. W., et al. (2019). Classification and visualization of alzheimer's disease using volumetric convolutional neural network and transfer learning. *Scientific Reports*, 9:18150.

Pei, L., Ak, M., Tahon, N. H., Zenkin, S., Alkarawi, S., Kamal, A., Yilmaz, M., Chen, L., Er, M., Ak, N., and Colen, R. (2022). A general skull stripping of multiparametric brain mris using 3d convolutional neural network. *Scientific Reports*, 12:10826.

Perlaki, G., Horvath, R., Nagy, S. A., Bogner, P., Doczi, T., Janszky, J., and Orsi, G. (2017). Comparison of accuracy between fsl's first and freesurfer for caudate nucleus and putamen segmentation. *Scientific Reports*, 7(1):2418.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack Jr, C., Jagust, W. J., Shaw, L. M., Toga, A. W., et al. (2010). Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209.

Quilis-Sancho, J., Fernandez-Blazquez, M. A., and Gomez-Ramirez, J. (2020). A comparative analysis of automated mri brain segmentation in a large longitudinal dataset: Freesurfer vs. fsl. *bioRxiv*.

Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H. C., Vermersch, P., Kuiper, M., Steinling, M., Wolters, E. C., and Valk, J. (1992). Atrophy of medial temporal lobes on MRI in "probable" alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry*, 55(10):967–972.

Shi, J., Zheng, X., Li, Y., Zhang, Q., and Ying, S. (2018). Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, 22(1):173–183.

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.*, 17(3):143–155.

Suk, H.-I., Lee, S.-W., and Shen, D. (2014). Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101:569–582.

Varoquaux, G. and Cheplygina, V. (2022). Machine learning for medical imaging: methodological failures and recommendations for the future. *npj Digital Medicine*, 5(1):48.