Automating the Koedam Parietal Atrophy Scale for Alzheimer's Using MRI Features and Clustering Techniques

Yasmin V. Oliveira, Ricardo J. Ferrari
Department of Computing
Federal University of São Carlos
São Carlos, SP, Brazil
Email: yasminvo@estudante.ufscar.br

Abstract—Early detection of Alzheimer's disease (AD) is crucial for effective intervention, and imaging biomarkers are pivotal in this process. The search for imaging biomarkers is important in diagnosing AD, offering a non-invasive and potentially early method to identify brain changes associated with the disease. These biomarkers can provide valuable insights into the progression of AD and aid in differential diagnosis, enabling the application of more effective treatment strategies. In this context, the Koedam visual scale for parietal atrophy is a valuable tool for assessing structural changes in the parietal lobe associated with AD. This study proposes an automated approach for the Koedam scale using attributes extracted from T1-weighted magnetic resonance imaging (MRI) combined with clustering techniques. Initially, a preprocessing pipeline is applied to the images to skull stripping, to mitigate noise and bias field effects and to define the ROI (parietal region). Subsequently, a finite mixture model is applied to segment the images into gray matter, white matter, and cerebrospinal fluid. The volume of each tissue is then utilized as a feature for clustering, effectively simulating the visual categorization of the Koedam scale. Our method, tested on 103 MRI images, demonstrates potential for automating the assessment of parietal atrophy, providing a more objective and efficient evaluation tool.

Keywords—Koedam scale, Alzheimer's disease; Clustering; MRI; Posterior atrophy

I. Introduction

Alzheimer's Disease (AD) is a neurodegenerative disorder and the most common form of dementia, accounting for approximately 60-80% of all cases [1]. Individuals with AD experience a progressive decline in their ability to retain new information due to damage in brain regions associated with memory formation. As the disease advances, cognitive abilities significantly decrease, leading to difficulties in daily activities such as walking and swallowing, ultimately resulting in death [2]. Cognitive impairment increases exponentially in older individuals, with a significant rise after age 65.

Mild Cognitive Impairment (MCI) refers to the early symptoms of cognitive decline that affect one or more areas without significantly impairing daily activities. The transition from MCI to dementia occurs when cognitive impairment interferes with the skills necessary for independence in daily activities [3]. With the increase in global life expectancy, chronic neurodegenerative diseases are becoming more prevalent. In the case of Alzheimer's, early diagnosis allows for the initiation of treatment as early as possible, thereby delaying the symptoms of the disease. Thus, accurate and effective computational methods for diagnosis become fundamental tools to accelerate the diagnosis of the disease [4].

The pathology of AD is characterized by the accumulation of extracellular senile plaques composed of filamentous aggregates of the beta-amyloid protein and intracellular neurofibrillary tangles predominantly formed by the tau protein. These plaques and tangles are particularly present in the cerebellar amygdala, hippocampus, and entorhinal cortex of the temporal lobe in the brains of AD patients [5] and are considered responsible for neuronal death and, consequently, neurodegeneration. Understanding the pathology and neurodegeneration process of AD enables the use of MRI imaging biomarkers to assist in the diagnosis and exclude other possible comorbidities of treatable dementias or suggest the presence of comorbidities that exacerbate dementia symptoms, such as cerebrovascular disease.

One way to diagnose AD using MRI images is by observing specific patterns of atrophy resulting from tissue loss. Recent studies using computational tools aim to develop automatic methods for classifying AD in MRI images [6], [7], [8]. MRI produces a 3D image, which physicians visualize as 2D images from selected slices in the axial, coronal, and sagittal planes. In diagnosing dementia using MRI images, radiologists systematically score global atrophy, focal atrophy, and vascular diseases (infarcts, white matter lesions, lacunes) [9]. This standardized assessment of MRI findings in a patient suspected of having a cognitive disorder includes the following scales: (a) Global Cortical Atrophy (GCA) scale, (b) Medial Temporal Lobe Atrophy (MTA) scale, (c) Koedam scale for parietal atrophy, and (d) Fazekas scale for white matter lesions and the search for strategic infarcts [8].

Besides the widely known and used medial temporal lobe atrophy, parietal atrophy also plays a significant role as a positive indicator in diagnosing AD. Precuneus atrophy (part of the superior parietal lobe) is particularly characteristic of AD in younger patients (presenile AD), who may have normal MTA scores [10]. The Koedam scale assesses parietal atrophy in sagittal, coronal, and axial planes, covering the enlargement of the posterior cingulate and parieto-occipital

sulci, in addition to parietal atrophy, including the precuneus. It ranges from 0 to 3, focusing on the integrity of these regions, with scores of 2 or higher indicating notable changes [11].

In this work, we introduce a novel method that automates the Koedam parietal atrophy scale using digital image processing and machine learning techniques. Manual evaluations, which often demand several minutes per image, are inherently laborious and time-consuming [12]. The time required for these assessments can differ based on the scale's complexity and the rater's expertise. While feasible in a clinical environment, this method becomes highly inefficient when processing extensive datasets containing thousands of images. Additionally, our method may serve as a second opinion for doctors, enhancing both the accuracy and confidence of their assessments.

By automating this process, we aim to standardize evaluations and improve diagnostic consistency. The paper is organized as follows: Section II describes the methodology and the image dataset used in this work, followed by results and discussions in Section III and conclusions in Section IV.

II. Methodology

The following sections outline each step of the proposed method as depicted in Figure 1.
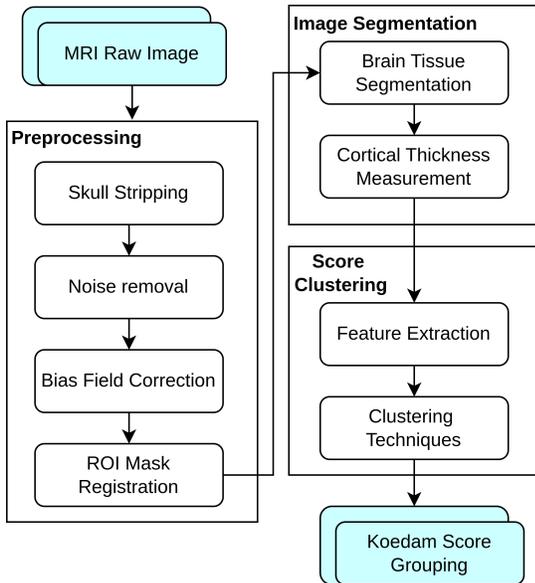

Fig. 1. Block diagram showing all steps of the proposed method.

A. Preprocessing

Preprocessing is a fundamental stage of our proposed method, preparing the images for brain tissue segmentation and cortical thickness measurement. The preprocessing steps, listed in the order they were executed, are described as follows.

1) Skull Stripping: The skull stripping process was performed using the ROBEX (Robust Brain Extraction) method [13]. This method combines a discriminative model and a generative model to achieve precise brain boundary detection.

The discriminative model employs a Random Forest classifier specifically trained to identify the brain boundary. Meanwhile, the generative model uses a point distribution model to ensure that the resulting brain segmentation is both plausible and anatomically consistent. This step focuses subsequent analysis on brain tissues, eliminating potential interference from surrounding structures.

2) Noise Removal: The Non-Local Means (NLM) denoising algorithm [14] was employed to reduce noise in the MRI images. The process starts with estimating local noise levels by analyzing intensity variance within local patches. Based on these estimates, the algorithm adaptively adjusts the filtering strength: increasing it in noisy areas to suppress noise more effectively and decreasing it in low-noise regions to preserve image details. The algorithm then applies this adjusted filter to each pixel by averaging its intensity with similar pixels in the local neighborhood, with weights determined by patch similarity. In this study, the algorithm parameters were set to a patch radius of 2 and a window radius of 5, as recommended by the authors after an exhaustive evaluation in MRI images [15].

3) Bias Field Correction: Bias field correction addresses low-frequency intensity variations in MRI images caused by inhomogeneities in the magnetic field, which can distort anatomical details and affect the accuracy of subsequent analyses. To correct these biases, the N4ITK algorithm [16] was used, which is an advanced version of the N3 algorithm. N4ITK improves upon its predecessor through enhanced B-spline fitting techniques and a hierarchical optimization approach, which together provide more accurate and robust correction of bias fields.

4) ROI Mask Registration: A mask of the region of interest (ROI) for this study was created using the CerebrA atlas [17], encompassing predefined anatomical regions such as the inferior parietal, superior parietal, precuneus, postcentral, and supramarginal areas. Affine and B-spline registration transformations were employed to align this ROI mask with each processed MRI image, ensuring consistent mapping across all images. The combined transformation from mapping the template atlas to the study images was applied to the ROI masks. To maintain fine anatomical details, the original images were kept fixed during the registration process. Finally, the registered masks were used to delineate the ROI in the original image for further segmentation and analysis.

B. Image Segmentation

This section details the methods employed in this study for image segmentation and the measurement of cortical thickness.

1) Brain Tissue Segmentation: This study aims to quantify the three primary brain tissues: white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). To achieve this, we utilized the Atropos algorithm from the ANTs framework [18]. Atropos employs a statistical model-based approach, using prior knowledge to differentiate between these three tissue types. Specifically, the algorithm applies a Gaussian Mixture Model (GMM) to represent the intensity distributions

of these tissues, allowing for probabilistic classification of each voxel based on the intensity values in the MRI images.

The process begins with initializing tissue probability maps, which can either be sourced from an atlas or provided by the user. In this study, we used the K-Means algorithm to generate these initial probability maps. These maps serve as prior probabilities for each tissue class, guiding the segmentation. Atropos refines these initial estimates through iterative expectation and maximization steps. During the expectation step, the algorithm computes the posterior probability of each voxel belonging to different tissue classes based on the current model parameters. In the maximization step, it updates these parameters to enhance the likelihood of the observed data, given the estimated class memberships.

One of the key features of Atropos is its ability to incorporate spatial priors through Markov Random Fields (MRF). This allows the algorithm to consider the spatial relationships between neighboring voxels, promoting smoother and more contiguous segmentation.

2) Cortical Thickness Measurement: The cortical thickness (CT) of each MR image was automatically determined using the diffeomorphic registration-based cortical thickness (DiReCT) algorithm [19]. This algorithm is grounded in the concept of diffeomorphic transformations, which are smooth and invertible mappings that preserve the structure of the data during the registration process. The core of DiReCT involves aligning brain images through these transformations to account for individual anatomical variations, thus ensuring that the cortical thickness measurements are precise and representative of the true anatomical structure.

The DiReCT algorithm operates in two main phases: initial alignment and thickness estimation. In the initial alignment phase, the algorithm employs a diffeomorphic registration approach to map the cortical surfaces of different subjects onto a common template. This process corrects for variations in brain shape and size, ensuring that corresponding cortical regions are properly aligned.

In the second phase, DiReCT estimates cortical thickness by analyzing the aligned cortical surfaces. It calculates the distance between the outer and inner cortical boundaries, providing a measure of thickness that is less sensitive to individual variability and alignment errors. By integrating diffeomorphic registration with accurate cortical thickness measurements, the DiReCT algorithm enhances the reliability of cortical thickness assessments.

C. Scores Clustering

In this section we briefly describe the MRI image-based features and the clustering techniques used in this study.

1) Feature Extraction: Feature extraction was conducted using the segmented brain image for tissue volumes and the DiReCT algorithm for CT measurements. These features were normalized by the total brain volume to account for variations in individual head sizes.

Figure 2 illustrates the correlation matrix computed to assess the relationships between the extracted image-based features.

Pearson's two-tailed test revealed significant correlations (p-value < 0.001) among "GM Vol" and "CT Mean". Based on these correlation results, we decided to use the three brain tissue volumes as features for the clustering algorithms.
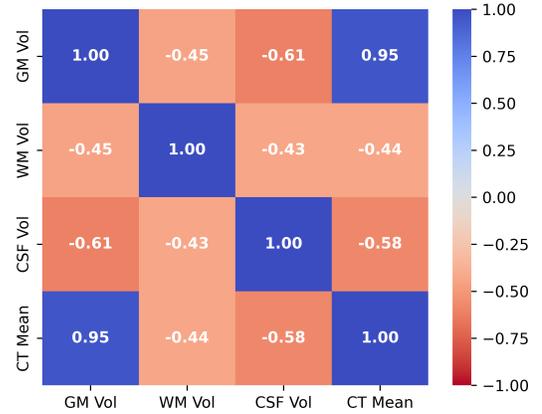

Fig. 2. Correlation matrix of CT normalized measure and the normalized GM, WM and CSF volumes.

2) Clustering Techniques: For this research, we chose to evaluate the K-Means, Agglomerative Clustering, and Spectral Clustering algorithms to assess their suitability for the image-based features, given their different underlying principles. Each algorithm was set to identify four clusters, with each cluster corresponding to a different level of the Koedam score. The K-Means clustering, for instance, is a widely used partitioning method that divides a dataset into K distinct, non-overlapping clusters based on feature similarity. The algorithm initializes K centroids randomly or by using specific heuristics and assigns each data point to the nearest centroid. It then iteratively updates the centroids by calculating the mean of the points assigned to each cluster, reassigning data points to the new nearest centroids, and repeating the process until convergence. The objective of K-Means is to minimize the sum of squared distances between data points and their respective cluster centroids, thereby ensuring that points within the same cluster are as close as possible. The maximum number of iterations was set to 300 and the tolerance to 0.0001.

The other algorithm used in this study was the Agglomerative clustering, that is a hierarchical method that builds a nested cluster hierarchy by successively merging or splitting clusters based on a measure of similarity. It starts with each data point as a single cluster and iteratively merges the most similar clusters until a desired number of clusters (four in our case) is achieved or all points are in a single cluster. The similarity between clusters was defined as ward linkage, which minimizes the variance of the clusters being merged, and the metric was the Euclidean distance.

Spectral Clustering is an advanced technique that uses the eigenvalues (spectrum) of a similarity matrix to perform dimensionality reduction before applying a standard clustering algorithm. In our case, the K-Means is applied in the reduced space. It constructs a similarity graph where nodes represent

data points and edges reflect the similarity between points. The algorithm then computes the Laplacian matrix of the graph and its eigenvectors, using these eigenvectors to project the data into a lower-dimensional space where clusters are more distinct. Spectral Clustering is particularly effective for datasets with complex structures, such as non-convex clusters or clusters connected in a manifold. However, since our feature space already has low dimensionality, the full potential of this technique may not have been fully realized.

The aim of comparing the performance of these methods was to identify the most effective approach for accurately classifying different levels of atrophy.

D. Image Dataset

The dataset used in this study was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI), (adni. loni.usc.edu), specifically targeting individuals diagnosed with AD who were under 65 years of age. This age range was selected due to the distinctive characteristic of parietal atrophy in Early-Onset Alzheimer's Disease (EOAD) cases [20]. A total of 103 T1-weighted MRI images from ADNI patients fitting these criteria were collected. The images, which are approximately isotropic with voxel sizes of 1mm³ and dimensions of 256×256×256, were obtained using various MRI scanners with magnetic field strengths of 1.5 and 3 Tesla. To quantitatively assess the results of the clustering algorithms, the images were visually evaluated by the author (Y.V.O.) using the visual rating scale as described in Koedam's paper [11].

III. RESULTS AND DISCUSSIONS

In clinical practice, the Koedam scale assesses parietal atrophy in the coronal, axial, and sagittal planes, assigning a score for each. The grading is as follows: Grade 0 indicates closed sulci, Grade 1 shows mild widening and atrophy, Grade 2 indicates substantial widening and atrophy, and Grade 3 represents end-stage atrophy with pronounced sulci widening and knife-blade atrophy. If scores differ between planes, the highest score is used [11]. However, using volume data instead of 2D slice images can diminish the impact of high severity scores because it aggregates data from all planes, potentially resulting in an underestimation of atrophy severity. In the following discussions we refer to this as "volumetric effect". Despite these challenges, the analysis achieved favorable accuracy rates for cluster classification.

As illustrated in Figure 3, the accuracy rates for K-Means were 84.61% for Cluster 0, 36.09% for Cluster 1, 68.75% for Cluster 2, and 85.71% for Cluster 3. For Spectral Clustering, the accuracy rates were 75.00% for Cluster 0, 35.71% for Cluster 1, 58.33% for Cluster 2, and 100% for Cluster 3. Finally, for Agglomerative Clustering, the accuracy rates were 46.15% for Cluster 0, 36.36% for Cluster 1, 71.87% for Cluster 2, and 100% for Cluster 3. These accuracy rates are calculated as the percentage of images classified with a specific Koedam score that are present in their respective clusters.
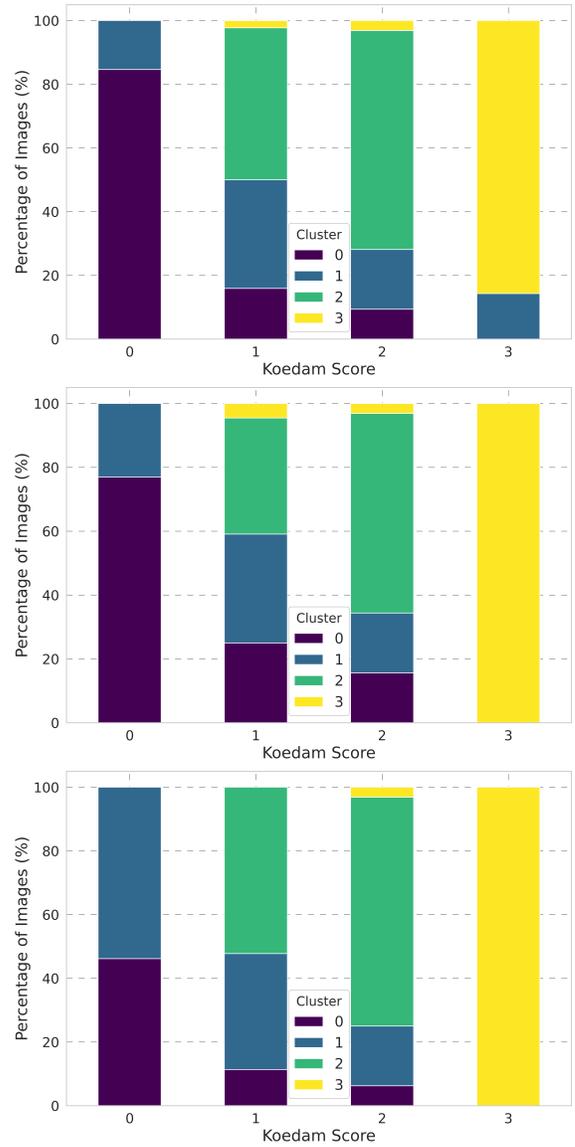

Fig. 3. Comparative of three clustering techniques applied to processed images using four features. Top: K-Means distribution of each Koedam Clustering score across clusters, Middle: Spectral Clustering distribution of each Koedam score across clusters, Bottom: Agglomerative Clustering distribution of each Koedam score across clusters.

In Figure 3 (Top and Middle plots), the higher accuracy observed for Cluster 0 can be attributed to better convergence of the algorithms, as images in this cluster consistently had a score of 0 across all three planes. Due to the visual rating scale used, any image with a higher score in any slice was classified into a higher category, ensuring clear distinction among the clusters. The images with score 0 that appear in Cluster 1 may be attributed to inaccuracies in the ground truth classification. The best accuracy, 84.61%, for Cluster 0 was achieved with the K-Means.

Moreover, Cluster 1 in Figure 3 showed significant dispersion, likely due to the volumetric effect. Conversely, images scored as 1 that appeared in Clusters 2 and 3 may indicate

segmentation errors or classification inaccuracies. The best result for Cluster 1 was achieved with Agglomerative Clustering, which had a 36.36% accuracy rate, highlighting the challenges in classifying this cluster due to subtle atrophy patterns. Cluster 2 had better accuracy than Cluster 1 across all clustering algorithms and better than Cluster 0 with Agglomerative Clustering. Images scored as 2 were present in both Clusters 0 and 1, mainly due to the volumetric effect. The highest accuracy for Cluster 2 was 71.87%, achieved with Agglomerative Clustering. Cluster 3 demonstrated the highest accuracy, often exhibiting homogeneous atrophy across all planes, with 100% accuracy in both Agglomerative and Spectral Clustering.

Considering the average accuracy rates of each clustering method, K-Means achieved 68.3%, Spectral Clustering 65.3%, and Agglomerative Clustering 63.6%. The average is significantly impacted by the low accuracy of Cluster 1.

Figure 4 illustrates a scatter plot of the image attributes showing the formation of four distinct clusters obtained with the K-Means algorithm. Examining this plot, it can be noticed that Cluster 3 is the most distinct cluster, following by Cluster 2. Clusters 0 and 1 are the ones with the highest overlapping. This visual verification is in according to the results in Figure 3. Additionally, it can be noticed that Cluster 3 contains the highest amount of CSF, which aligns with the increased atrophy observed, as CSF volume tends to rise with greater atrophy. Conversely, Cluster 0, characterized by minimal atrophy, maintains a lower CSF volume.
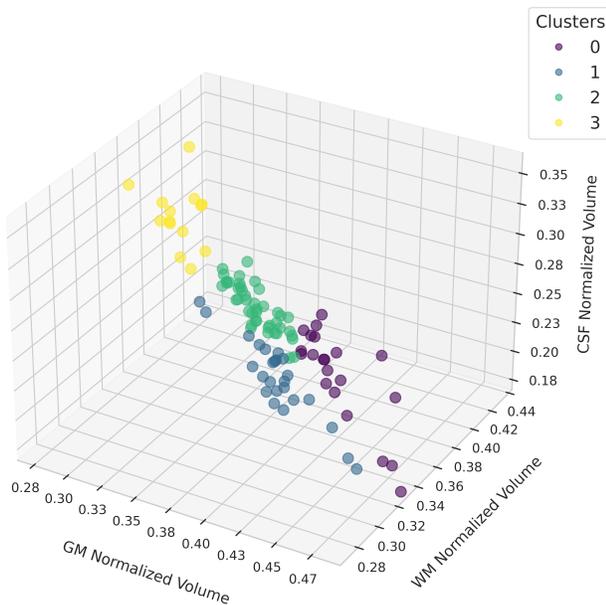

Fig. 4. Scatter plot of the normalized brain tissues clustered by the K-Means algorithm.

To evaluate the compactness and separation of the clusters generated by the three methods, silhouette scores were computed. K-Means achieved a score of 0.38, while both Spectral Clustering and Agglomerative Clustering had scores of 0.35.

These relatively low scores indicate less-than-ideal separation between clusters, with considerable overlap as shown in Figure 4. Additionally, the cohesion was significantly affected by high-dispersion clusters, such as Clusters 0 and 1.

To evaluate the method's effectiveness in clustering MR images based on parietal atrophy, we identified the images closest to each cluster centroid, as shown in Figure 5. This analysis confirms that the clustering method effectively differentiates between atrophy levels in the study images. Moreover, for the ROI marked in red, atrophy and corresponding CSF volumes increase progressively across the clusters, with Cluster 0 showing preserved areas and Cluster 3 indicating extreme atrophy.
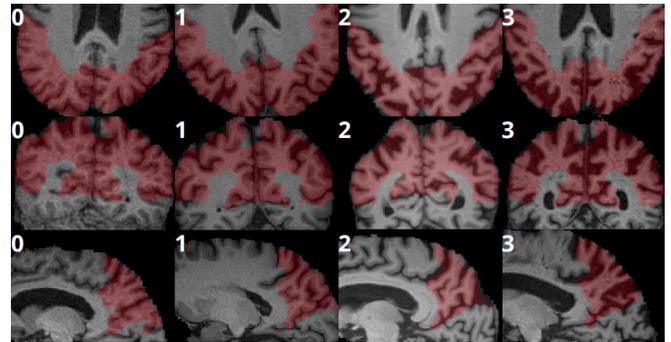

Fig. 5. Examples of images nearest to the centroid of their respective cluster groups. The colored brain regions highlight areas used for the analysis.

Finally, the mean scores for the Functional Activities Questionnaire (FAQ), Mini Mental State Examination (MMSE), and Clinical Dementia Rating (CDR) were calculated for each cluster. FAQ assesses daily living activities, MMSE evaluates cognitive function, and CDR measures dementia severity. Higher FAQ and CDR scores correlate with greater decline, while lower MMSE scores indicate more severe decline. A strong linear relationship was observed between FAQ scores and clusters identified using Spectral Clustering, with scores of 14.0, 16.35, 19.50, and 22.36 for Clusters 0 through 3, respectively. MMSE scores, also analyzed with Spectral Clustering, were 21.89, 19.54, 20.72, and 19.21, while CDR scores using K-Means were 0.86, 0.83, 0.92, and 0.95. These represent the best results across the methods tested. EOAD patients with parietal atrophy may initially present with non-cognitive symptoms, such as functional, motor, or visual issues [21], which FAQ effectively identifies. The lack of linearity in MMSE and CDR scores, particularly in Cluster 1, is attributed to high data variability within that cluster.

IV. CONCLUSIONS

This study presents an automated approach to the Koedam scale for assessing parietal atrophy in Alzheimer's disease using MR image features and clustering techniques. By leveraging T1-weighted MRI images, we developed a method to quantify WM, GM, and CSF volumes, which served as input features for clustering algorithms. The application of the Atropos algorithm within the ANTs framework allowed

for robust segmentation of brain tissues, while the DiReCT algorithm provided accurate cortical thickness measurements.

The clustering analysis demonstrated varying degrees of accuracy across different clusters. Cluster 3 is the most distinct cluster, following by Cluster 2. Clusters 0 and 1 are the ones with the highest overlapping, likely due to the volumetric effect and subtle differences in atrophy patterns. Despite these challenges, the overall performance of the clustering algorithms was promising, with each method showing strengths in different clusters. The silhouette scores, although modest, indicated the need for further refinement in separating clusters more distinctly.

The study also highlighted the relationship between clinical scores and the identified clusters. A strong linear correlation was observed between FAQ scores and the clusters, validating the method's potential in differentiating levels of functional decline. MMSE and CDR scores provided additional insights, though their non-linearity in certain clusters pointed to the inherent variability in AD presentations.

In conclusion, while this research is ongoing, our automated approach offers a novel and objective tool for assessing parietal atrophy in AD, aligning with the Koedam visual scale. By enhancing the accuracy and efficiency of atrophy classification, this method holds promise for improving early diagnosis and intervention strategies in clinical practice. Future work will focus on addressing the volumetric effect, refining clustering techniques, and validating the approach with larger and more diverse datasets to further establish its clinical utility.

V. Acknowlegments

References

[1] A. Association, "Alzheimer's Disease Facts and Figures," Tech. Rep. 4, 2023.

[2] A. Atri, "The Alzheimer's Disease Clinical Spectrum: Diagnosis and Management," *Medical Clinics*, vol. 103, no. 2, pp. 263–293, 2019.

[3] J. Smid, A. Studart-Neto, K. César-Freitas, M. Dourado, R. Kochhann, B. Barbosa, L. Schilling, M. Balthazar, N. Frota, L. de Souza, P. Caramelli, P. Bertolucci, M. Chaves, S. Bruck, R. Nitrini, E. Resende, and F. Vale, "Declínio cognitivo subjetivo, comprometimento cognitivo leve e demência – diagnóstico sindrômico: recomendações do Departamento Científico de Neurologia Cognitiva e do Envelhecimento da Academia Brasileira de Neurologia," *Dement Neuropsychol*, vol. 16, no. 3, pp. 1–17, 2022.

[4] K. Dhana, O. H. Franco, E. M. Ritz, C. N. Ford, P. Desai, K. R. Krueger, T. M. Holland, A. Dhana, X. Liu, N. T. Aggarwal, D. A. Evans, and K. B. Rajan, "Healthy lifestyle and life expectancy with and without Alzheimer's dementia: population based cohort study," *BMJ (Clinical research ed.)*, vol. 377, no. e068390, 2022.

[5] Y. Rao, B. Ganaraja, B. Murlimanju, T. Joy, A. Krishnamurthy, and A. Agrawal, "Hippocampus and its involvement in Alzheimer's disease: a review," *3 Biotech*, vol. 12, no. 55, pp. 1–10, 2022.

[6] B. Dubois, C. von Arnim, N. Burnie, S. Bozeat, and J. Cummings, "Biomarkers in Alzheimer's disease: role in early and differential diagnosis and recognition of atypical variants," *Alzheimer's Research & Therapy*, vol. 15, no. 175, pp. 1–13, 2023.

[7] X. Feng, F. Provenzano, and S. Small, "A deep learning MRI approach outperforms other biomarkers of prodromal Alzheimer's disease," *Alzheimer's Research & Therapy*, vol. 14, no. 45, pp. 1–11, 2022.

[8] T. Klyucherev, P. Olszewski, A. Shalimova, V. Chubarev, V. Tarasov, M. Attwood, S. Syvänen, and H. Schiöth, "Advances in the development of new biomarkers for Alzheimer's disease," *Translational Neurodegeneration*, vol. 11, no. 25, pp. 1–24, 2022.

[9] J. Wu, K. Zhao, Z. Li, D. Wang, Y. Ding, Y. Wei, H. Zhang, and Y. Liu, "A systematic analysis of diagnostic performance for Alzheimer's disease using structural MRI," *Psychoradiology*, vol. 2, no. 1, pp. 1–9, 2022.

[10] G. Karas, P. Scheltens, S. Rombouts, R. van Schijndel, M. Klein, B. Jones, W. van der Flier, H. Vrenken, and F. Barkhof, "Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study," *Neuroradiology*, vol. 49, no. 12, pp. 967–976, 2007.

[11] E. Koedman, M. Lehmann, W. van der Flier, P. Scheltens, Y. Pijnenburg, N. Fox, F. Barkhof, and M. Wattjes, "Visual assessment of posterior atrophy development of a MRI rating scale," *European Radiology*, vol. 21, pp. 2618–2625, 2011.

[12] L.-O. Wahlund, P. Julin, J. Lindqvist, and P. Scheltens, "Visual assessment of medial temporal lobe atrophy in demented and healthy control subjects: correlation with volumetry," *Psychiatry Research: Neuroimaging*, vol. 90, no. 3, pp. 193–199, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925492799000165

[13] J. E. Iglesias, C. Liu, P. M. Thompson, and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1617–1634, 2011.

[14] J. Manjón, P. Coupé, L. Martí-Bonmatí, D. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *Journal of Magnetic Resonance Imaging*, vol. 31, no. 1, pp. 192–203, 2010.

[15] J. Manjón, J. Carbonell-Caballero, J. Lull, L. García-Martí, G. Martí-Bonmatí, and M. Robles, "MRI denoising using Non-Local Means," *Medical Image Analysis*, vol. 12, pp. 514–523, 2008.

[16] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: improved N3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[17] A. Manera, M. Dadar, V. Fonov, and D. Collins, "CerebrA, registration and manual label correction of Mindboggle-101 atlas for MNI-ICBM152 template," *Scientific Data*, vol. 7, no. 1, pp. 1–9, 2020.

[18] B. Avants, N. Tustison, J. Wu, P. Cook, and J. Gee, "An Open Source Multivariate Framework for n-Tissue Segmentation with Evaluation on Public Data," *Neuroinformatics*, vol. 9, no. 4, pp. 381–400, 2011.

[19] M. Rebsamen, C. Rummel, M. Reyes, R. Wiest, and R. Mckinley, "Direct cortical thickness estimation using deep learning-based anatomy segmentation and cortex parcellation," *Human Brain Mapping*, vol. 41, pp. 4804–4814, 2021.

[20] M. Mendez, "Early-Onset Alzheimer's Disease," *Neurologic Clinics*, vol. 35, no. 2, pp. 263–281, 2017.

[21] L. Apostolova, P. Aisen, A. Eloyan, A. Fagan, K. Fargo, T. Tatiana Foroud, C. Gatsonis, L. Grinberg, C. Jack Jr., J. Kramer, R. Koeppe, W. Kukull, M. Murray, K. Nudelman, M. Rumbaugh, A. Toga, P. Vemuri, A. Trullinger, L. Iaccarino, G. Day, N. Graff-Radford, L. Honig, D. Jones, J. Masdeu, M. Mendez, E. Musiek, C. Onyike, E. Rogalski, S. Salloway, D. Wolk, T. Wingo, M. Carrillo, B. Dickerson, G. Rabinovici, and the LEADS Consortium, "The Longitudinal Early-onset Alzheimer's Disease Study (LEADS): Framework and methodology," *Alzheimer's & Dementia*, vol. 17, pp. 2043–2055, 2021.