https://doi.org/10.48047/AFJBS.6.9.2024.5310-5314

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Comparative Analysis of Agglomerative Clusteringand K-means Clustering Algorithms for Brain Tumor Segmentation in MRI Images

[1]Sanjeev Gour, [2]Mamta Gour,[3]Abdul Razzak Khan Qureshi, [4]Mohsin Ali, [5]Ruby Bhatt, [6]Jitendra Choudhary

1 Asst. Professor, Medi-Caps University, Indore, M.P. India. sunj129@gmail.com
2 Asst. Professor, Medi-Caps University, Indore, M.P. India. mamtagour129@gmail.com
3 Asst. Professor, Medi-Caps University, Indore, M.P. India. dr.arqureshi786@gmail.com
4 Research Scholar, Medi-Caps University, Indore. India mohsin.ali@medicaps.ac.in
5 Asst. Professor, Medi-Caps University, Indore, M.P. India. profrubybhatt15@gmail.com
6 Associate Professor, Medi-Caps University, Indore, M.P. India. jitendra.scsit@gmail.com

Volume 6, Issue 9, May 2024

Received: 12 March 2024

Accepted: 02 April 2024

Published: 28 April 2024

doi: 10.48047/AFJBS.6.9.2024.5310-5314

Abstract—This research paper presents an analysis of clus- tering algorithms, specifically Agglomerative Clustering and K- means Clustering, applied to a dataset consisting of 100 MRI images of brain tumors. The aim of the study is to explore the effectiveness of these algorithms in segmenting brain tumor regions within the images.

The dataset comprises MRI images obtained from patientsdiagnosed with various types of brain tumors. Preprocessing techniques, including noise reduction, image enhancement, and normalization, were applied to ensure data quality and consis- tency. The images were then represented as feature vectors using appropriate image descriptors.

Two clustering algorithms, Agglomerative Clustering and K- means Clustering, were implemented and compared in terms of their ability to accurately cluster the tumor regions within the MRI images. Agglomerative Clustering, a hierarchical clusteringtechnique, merged similar data points to form clusters iteratively.K-means Clustering, on the other hand, partitioned the datapoints into a predefined number of clusters based on their proximity to cluster centroids.

visual inspections of the resulting clusters were conducted by domain experts to validate the accuracy and effectiveness of the algorithms.

The experimental results demonstrated that both Agglomer- ative Clustering and K-means Clustering achieved promising results in segmenting brain tumor regions within the MRI im- ages. However, each algorithm exhibited distinct characteristics in terms of cluster shape, size, and computational efficiency. Agglomerative Clustering demonstrated its ability to handle clusters of arbitrary shapes and sizes, while K-means Clustering performed well when the cluster shapes were relatively spherical and of similar sizes.

The findings of this research provide valuable insights into theapplication of clustering algorithms in the domain of medical imaging analysis. The results can potentially contribute to the development of more accurate and efficient techniques for brain tumor segmentation, assisting medical professionals in diagnosis, treatment planning, and monitoring of patients with brain tumors.

Index Terms—Image Clustering, K-means Clustering, Agglom-erative Clustering.

I. Introduction

Medical imaging plays a crucial role in modern healthcare, providing valuable insights for diagnosis, treatment planning,and monitoring of various medical conditions [5]. The ever- increasing volume of medical image data necessitates efficientmethods for organizing and analyzing these images. Cluster- ing, as an unsupervised learning technique, offers a promising approach to group similar images together, enabling effective exploration and interpretation of medical image datasets. In particular, the application of k-means clustering algorithms to medical image clustering has gained significant attention in recent years [1], [11]. Brain tumor segmentation plays a crucialrole in the field of medical imaging analysis as it aids in the diagnosis, treatment planning, and monitoring of patients withbrain tumors. Magnetic Resonance Imaging (MRI) is a widelyused modality for brain tumor imaging due to its excellent softtissue contrast and non-invasive nature. However, manually delineating tumor regions in MRI images is a labor-intensive and time-consuming task, prompting the need for automated segmentation techniques.

Clustering algorithms have been widely employed for imagesegmentation tasks, including brain tumor segmentation. These algorithms group similar pixels or regions together based on certain similarity measures. Among various clustering techniques, Agglomerative Clustering and K-means Clusteringhave shown promising results in different applications.

Agglomerative Clustering is a hierarchical clustering algo- rithm that iteratively merges similar data points or clusters to form larger clusters. It starts with each data point as a separatecluster and progressively merges the closest clusters based on their proximity. It has the advantage of being able to handle clusters of arbitrary shapes and sizes.

K-means Clustering, on the other hand, is a centroid-based partitioning algorithm that divides data points into K clusters, where K is a predetermined number. It aims to minimizethe within-cluster sum of squares by iteratively updating the cluster centroids and reassigning data points to the nearest centroid. K-means Clustering is computationally efficient and particularly suitable when the clusters are relatively spherical and of similar sizes.

In this research paper, we aim to compare the effective- ness of Agglomerative Clustering and K-means Clustering

algorithms for brain tumor segmentation in MRI images. We have collected a dataset consisting of 100 MRI images, each containing a brain tumor. The dataset covers a variety of braintumor types, sizes, and locations to ensure a comprehensive evaluation.

We will preprocess the MRI images, applying techniques such as noise reduction, image enhancement, and normaliza- tion to improve the quality and consistency of the data.

In this study, we will compare the performance of k-means clustering algorithms with alternative clustering algorithms commonly used in medical image analysis [12]. we aim to elucidate the potential benefits and challenges associated with utilizing k-means clustering algorithms for medical image analysis [7].

II. area of Interest

Brain tumors are abnormal growths in the brain that can have various types, sizes, and locations. Identifying and de- lineating the tumor regions in MRI images is essential for clinicians to assess the extent and characteristics of the tumor. Manual segmentation of brain tumors is a time-consumingand subjective process, highly dependent on the expertise of the radiologists. Therefore, there is a growing interest in developing automated segmentation techniques that canprovide consistent and reliable results.

Researchers work on adapting clustering algorithms to address the specific challenges and needs of these domains, contributing to specialized clustering approaches for domain- specific medical image datasets [6].

III. literature Review

Medical image clustering differs from clustering of other types of images in several ways due to the unique character- istics of medical images. Firstly, medical images come from various modalities such as X-ray, MRI, CT scan, ultrasound, etc. Each modality has its own specific characteristics and image acquisition processes. This means that clustering al- girths need to account for these differences in modalities while analyzing and grouping the images. Secondly, medical images often have higher dimensions compared to regularimages. They can be three-dimensional (3D) or even four- dimensional (4D) with temporal information. The clustering algorithms should be able to handle higher-dimensional data appropriately and capture meaningful patterns in the image volumes. The increased dimensionality poses a challenge in terms of computational complexity and the need for efficient algorithms [8]. Another important aspect is that medical im- ages are prone to noise and artifacts due to various factorssuch as imaging equipment limitations, patient movement,or inherent properties of the imaging technique. Clustering algorithms used for medical images should be robust enoughto handle and account for these noise and artifacts while extracting meaningful clusters [4]. The algorithms should be able to differentiate between true features and noise, ensuring the accuracy of the clustering results.In addition, medical images often require expert interpre- tation and annotation for accurate analysis. Clustering algo- rithms applied to medical images should consider incorpo- rating expert knowledge or domain-specific features to guide the clustering process and enhance the interpretability of the results. Expert annotations or clinical guidelines can be utilized to define the similarity measures or guide the clustering process based on specific medical criteria. Medical imagesare typically analyzed in the context of specific clinical tasks or applications. The clustering results need to be clinically relevant and meaningful for medical professionals. It is im- portant to consider the clinical implications and validate the clusters generated by the algorithms with domain experts. Theclustering should assist in identifying meaningful subgroupsof patients or anatomical structures that can aid in diagnosis, treatment planning, or monitoring.

Lastly, medical image datasets may have limited sample sizes or class imbalances, making it challenging to train and evaluate clustering algorithms effectively. Additionally, obtaining accurate ground truth labels for medical images can be time-consuming and costly. Therefore, unsupervisedor semi-supervised clustering approaches are often used in medical image analysis, where the algorithms exploit the inherent structures in the data without relying heavily on labeled examples [13].

Given these, clustering algorithms applied to medical images need to be tailored and adapted to handle the specific challenges and characteristics of medical imaging data [2]. Researchers and practitioners often develop specialized clus- tering methods or modify existing algorithms to accommodatethese unique aspects of medical image clustering. This allows for more accurate and clinically relevant clustering results, enabling improved medical image analysis and decision- making. Image clustering and retrieval have a wide rangeof applications, including image organization, recommendation systems, and image search engines. The use of image mining techniques can significantly improve the accuracy and efficiency of these applications, enabling users to find and organize large collections of images more effectively. In this paper, we will discuss the various image mining techniques used for image clustering and retrieval and their applicationsin different fields. Some importance techniques are Imageretrieval, Object recognition, Image organization and Data visualization.

A. Image Clustering

Image clustering include improved organization and search-ability of large image datasets, faster retrieval of relevant im- ages, and the ability to identify patterns and trends within the data. Despite the benefits of image clustering, there are some limitations to the approach. One limitation is the subjective nature of the clustering process, as the choice of featuresand clustering algorithm can significantly impact the resulting clusters [9]. Another limitation is the requirement for large amounts of annotated data to train clustering algorithms, whichcan be time-consuming and expensive. Additionally, clustering

algorithms may struggle with highly diverse datasets, where images may not belong to clear-cut clusters. Finally, the useof image clustering may not be suitable for applications that require fine-grained categorization or object recognition, as these tasks require more sophisticated algorithms and a deeperunderstanding of the image content. It’s worth noting that the choice of features and the number of clusters (K) are crucial parameters that require careful consideration and tuning basedon your specific dataset and objectives. Experimenting with different feature sets, clustering parameters, and evaluation metrics can help you optimize the clustering results. This study contributes to the understanding of the applicability and effectiveness of k-means clustering algorithms in the realm of medical image analysis. The findings aim to guide researchersand practitioners in choosing appropriate clustering techniquesand provide valuable insights for the development of advancedclustering methods tailored specifically for medical image datasets.

IV. DATA COLLECTION

In order to get the data for our research area There are sev- eral resources where we can find data for medical image clus- tering. Kaggle: Kaggle (https://www.kaggle.com/) is a popularplatform for data science competitions. It offers a wide range of datasets, including medical imaging datasets. we can searchfor medical imaging datasets on Kaggle and find ones that suityour clustering requirements. In order to process the medical data we uses the Python It provides several libraries that are commonly used for processing medical images. SimpleITKis a powerful library for medical image analysis. It providesa simple and efficient interface to perform common image processing tasks, such as image segmentation, registration,filtering, and feature extraction [3], [10]. It supports various medical image formats and is widely used in the research and medical imaging communities.

V. METHODOLOGY

We have used number of python libraries to process the medical images, these libraries offer a range of functionalitiesand can be combined to perform advanced medical imageprocessing tasks.

A. Agglomerative Clustering

we obtain a dataset consisting of 100 MRI images withbrain tumors. Ensure the dataset covers a diverse range of tumor types, sizes, and locations to ensure a comprehensive evaluation. after obtaining the data set, we apply preprocessing techniques to enhance the quality and consistency of the MRI images. Techniques may include noise reduction, image denoising, intensity normalization, and image enhancement. Now data set is ready as input for clustering algorithms.

Implement the Agglomerative Clustering algorithm.

Apply Agglomerative Clustering to the feature vectors of the MRI images to obtain a hierarchical clustering structure.

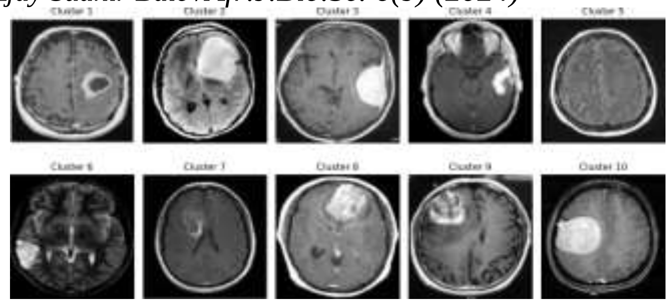

Fig. 1. Agglomerative Clustering

B. K-means Clustering

Implement the K-means Clustering algorithm. Select the number of clusters (K) based on prior knowledge or through techniques such as elbow method or silhouette analysis. ApplyK-means Clustering to the feature vectors of the MRI images to obtain cluster assignments.

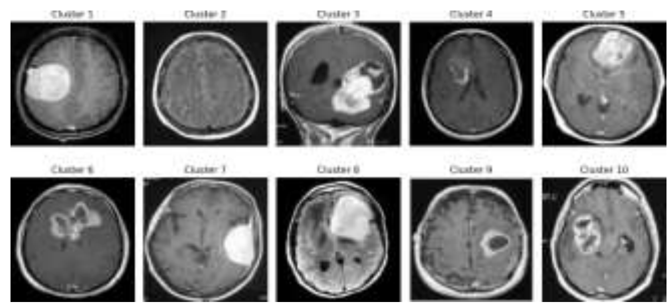

Fig. 2. K-mean Clustering

Assess the performance of both clustering algorithms using evaluation metrics specific to clustering tasks.

Compare the performance of Agglomerative Clustering andK-means Clustering.

Visually inspect the clustering results and compare them with ground truth tumor segmentations. Assess the visual qual-ity of the segmented tumor regions and determine the abilityof the algorithms to accurately delineate tumor boundaries. In order to evaluate the quality and performance of the clusteringalgorithm we used silhouette score, adjusted Rand index, and completeness score. Each of these metrics provides different insights into different aspects of the clustering results.

1. The silhouette score measures the quality of clustering byassessing how well each sample fits into its assigned cluster compared to other clusters. The silhouette score ranges from -1 to 1, with higher values indicating better-defined and more compact clusters. A high silhouette score suggests that the samples are well-clustered and that the clustering algorithm has separated the data points effectively.

2. The adjusted Rand index measures the similarity betweenthe clustering labels obtained from the algorithm and the true labels (ground truth). A higher ARI score suggests better agreement between the clustering and true labels, indicating the clustering algorithm's ability to accurately assign datapoints to their correct clusters.

3. The completeness score evaluates how well the clusteringalgorithm assigns all data points from the same true class to thesame cluster. The completeness score ranges from 0 to 1, witha value of 1 indicating that all samples of a class are assigned to the same cluster. A higher completeness score suggests thatthe clustering algorithm has captured the entire data points belonging to a true class within a single cluster, which isdesirable. These metrics help you evaluate the effectiveness ofthe algorithm in capturing the underlying structure and patternsin your data.

VI. RESULTS AND DISCUSSIONS

The comparative analysis of Agglomerative Clustering and K-means Clustering for brain tumor segmentation in MRIimages yielded insightful results. The evaluation metrics and visual inspection findings provide a comprehensive under-standing of the performance of both algorithms in accurately delineating tumor regions. Based on the provided performancemeasures for the medical images dataset, here's a comparison of the performance between the K-Means Clustering and Agglomerative Clustering algorithms:

Silhouette Score: K-Means Clustering: 0.42324639 Ag- glomerative Clustering: 0.406898704 The K-Means Clustering algorithm achieves a slightly higher silhouette score comparedto Agglomerative Clustering, indicating that the K-Means algorithm provides better-defined and more compact clusters for the given dataset.

Adjusted Rand Index: K-Means Clustering: -0.000921217 Agglomerative Clustering: -0.011762397 Both clustering al- gorithms have low adjusted Rand index scores, indicating that the clustering labels generated by the algorithms do not havea significant agreement with the true labels. However, theK-Means Clustering algorithm has a slightly higher adjusted Rand index score compared to Agglomerative Clustering.

Completeness Score: K-Means Clustering: 0.005225998 Agglomerative Clustering: 0.005620278 Both clustering algo-rithms have low completeness scores, indicating that they fail to capture all the data points belonging to the same true class within a single cluster. The scores are quite similar for both algorithms.

Based on these performance measures, the K-Means Clus- tering algorithm performs slightly better in terms of silhouette score and adjusted Rand index compared to Agglomerative Clustering for the given medical images dataset. However, it'simportant to note that the performance may vary depending onthe specific dataset and its characteristics. It's recommended toconsider other evaluation metrics and perform further analysisto make a more comprehensive assessment of the cluster-ing algorithms' performance. The results showed that both Agglomerative Clustering and K-means Clustering achieved reasonably high scores in terms of clustering agreement. However, there were some variations in the performance of thealgorithms across different tumor types, sizes, and locations.Future research directions could involve exploring the com-bination of both algorithms or incorporating other advanced clustering techniques to leverage their respective advantages. Additionally, investigating the impact of different distancemetrics and linkage criteria in Agglomerative Clustering and exploring adaptations of K-means Clustering for handling irregular cluster shapes could further improve the segmentationaccuracy.

VII. Conclusion

Compare the performance of Agglomerative Clustering and K-means Clustering in terms of accuracy, efficiency, and robustness. Analyze the strengths and limitations of each algorithm for brain tumor segmentation.

Silhouette Score indicates that the K-Mean algorithm pro- duced clusters that are more compact and well-separated than the Agglomerative clustering algorithm. While Adjusted Rand Index gives A negative ARI score suggests that the clustering results are not significantly correlated with the ground truth labels. Also, both algorithms achieved very low completeness scores. so overall the K-mean clustering gives better results.

Future research could explore the combination of these clustering algorithms with other machine learning approaches to further enhance their performance and applicability in clinical settings.

References

[1] Nasir Ahmed. Recent review on image clustering. *IET Image Process- ing*, 9(11):1020–1032, 2015.

[2] Tamalika Chaira. A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images. *Applied soft computing*,11(2):1711–1717, 2011.

[3] Guy Barrett Coleman and Harry C Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 67(5):773–785, 1979.

[4] Nameirakpam Dhanachandra, Khumanthem Manglem, and Yambem Jina Chanu. Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. *Procedia ComputerScience*, 54:764–771, 2015.

[5] Maria Fayez, Soha Safwat, and Ehab Hassanein. Comparative study of clustering medical images. In *2016 SAI Computing Conference (SAI)*, pages 312–318. IEEE, 2016.

[6] Subbiahpillai Neelakantapillai Kumar, Alfred Lenin Fred, and Paul Se- bastin Varghese. An overview of segmentation algorithms for the analysis of anomalies on medical images. *Journal of Intelligent Systems*,29(1):612–625, 2018.

[7] Hongfu Liu, Ming Shao, Sheng Li, and Yun Fu. Infinite ensemblefor image clustering. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages1745–1754, 2016.

[8] Takayasu Moriya, Holger R Roth, Shota Nakamura, Hirohisa Oda, Kai Nagara, Masahiro Oda, and Kensaku Mori. Unsupervised segmentation of 3d medical images based on clustering and deep representation learn-ing. In *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10578, pages 483–489.SPIE, 2018.

[9] Dzung L Pham, Chenyang Xu, and Jerry L Prince. Current methods in medical image segmentation. *Annual review of biomedical engineering*, 2(1):315–337, 2000.

[10] Jadwiga Rogowska. Overview and fundamentals of medical imagesegmentation. *Handbook of medical imaging, processing and analysis*, pages 69–85, 2000.

[11] Neeraj Sharma and Lalit M Aggarwal. Automated medical image segmentation techniques. *Journal of medical physics/Association of Medical Physicists of India*, 35(1):3, 2010.

[12] Siti Noraini Sulaiman and Nor Ashidi Mat Isa. Adaptive fuzzy-k-means clustering algorithm for image segmentation. *IEEE Transactions on Consumer Electronics*, 56(4):2661–2668, 2010.

[13] Jianlong Wu, Keyu Long, Fei Wang, Chen Qian, Cheng Li, Zhouchen Lin, and Hongbin Zha. Deep comprehensive correlation mining for image clustering. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8150–8159, 2019.