A Comparison of Tumor Detection Approaches Based on K-means+GA, Watershed Clustering, and Otsu Threshold Method

Sonia Rezaei

Department of Radiology, School of Allied Medical Science, Shahid Beheshti University of Medical Science, Tehran, Iran

Objective: Tumor detection from medical images using image processing approaches is an essential step in the treatment of patients. In this study, three tumor detection approaches based on segmentation are proposed.

Methods: The first method uses genetic algorithm to optimize cluster centers of K-means clustering. The second method utilizes the Otsu threshold method to detect tumor areas. The third method is based on morphological operators and watershed clustering to detect tumor areas.

Result: The mentioned methods are applied to three images and the accuracy of results is assessed using the confusion matrix. Results show that the methods have an error about 2 to 6 percent in detecting tumor areas, which it indicates a high accurate results. Moreover, accuracy of K-means and Otsu methods in separating tumor and no-tumor areas is higher than that of the watershed method.

Conclusion: The three clustering methods are appropriate to separate tumor and no-tumor areas and their results are robust.

Introduction

The early detection of tumors, particularly in medical images of the brain, is vital for effective treatment and improved patient outcomes. Magnetic Resonance Imaging (MRI) is a suitable source for the detection of tumors, since it can separate tumor region from other sections [1]. However, interpreting MRI images can be challenging. Therefore, it is necessary to develop high accurate image processing techniques. It is very difficult to develop a comprehensive image processing technique to extract the tumor region, since characteristics of an image such as the contrast and size, shape, and location of tumor may be effective on the performance.

Soomro et al., (2022) provided a review paper regarding tumor detection using image processing from Magnetic Resonance Imaging (MRI). They investigated types of image segmentation for tumor detection and mentioned new cutting-edge methods for effective detection [2]. Abdalla and Esmail (2018) presented a solution based on artificial neural network method to detect tumor area.

They concluded that the proposed method can detect tumor area with an accuracy of about 99 % [3]. Tseng and Tang (2023) proposed a method to detect tumor area from image segmentation. First, image enhancement and segmentation are used to provide initial classification. Then, feature selection and the eXtreme Gradient Boosting classification approach are employed to detect tumor area. The accuracy of the proposed method is above 97 % that shows the proposed method is accurate [4].

Brain tumor detection using classification approaches is a solution to this end. There are some studies based on classification approaches in the previous studies [5-7]. Solanki et al. (2023) provided an overview regarding tumor detection using new classification methods such as Deep Learning, Transfer Learning, and Machine Learning models. According results, deep learning methods are an appropriate solution to segment MRI images [8]. Moreover, Rahman and Islam investigated convolutional neural networks to segment tumor areas. The proposed method with an accuracy above 97 % is robust to this end [9].

Clustering based tumor detection is an automatic manner to find region of interest. There is no need to use training samples. In some studies, clustering methods such as K-means, fuzzy C means, and improved ones are used to tumor detection [10-12]. In the mentioned methods, cluster centers regarding tumor and no-tumor classes are identified in different manners. A robust cluster center may result an accurate clustering output. For example, Babo et al., (2020) mentioned optimization algorithms can be helpful to achieve robust results in tumor detection [13]. Traditional methods of tumor detection often rely on the expertise of radiologists, who may face challenges in interpreting complex medical images. As such, there is a growing need for advanced computational image processing techniques that can enhance the accuracy and performance of tumor detection. Clustering techniques have become essential tools in medical image analysis, enabling the automatic grouping of similar pixels. K-means clustering, a widely used algorithm, partitions an image into K clusters based on pixel similarity, making it effective for initial tumor detection [14]. However, K-means has limitations, including sensitivity to image noise and the requirement for prior knowledge of the number of clusters, i.e., K [15]. To enhance K-means performance, Genetic Algorithms (GA) can be integrated to find optimized cluster centers, improving segmentation accuracy by evolving solutions based on a fitness or objective function [16].

Another way to find tumor area from image is to use thresholding approaches [17-20]. They are simple, rapid and accurate. Naser and Obaid (2018) presented a threshold based tumor detection approach. The accuracy of the proposed method is higher than 97% [21]. Sudharani et al., (2016) proposed a threshold technique based on histogram analysis to find tumor area. The proposed method can detect tumor area with an accuracy of 90 % [22]. Another promising method is Otsu's thresholding, which optimally separates pixels into two classes by minimizing intra-class variance [23]. This method is particularly effective for images with bimodal histograms and also can serve as a preprocessing step to enhance the performance of clustering algorithms.

Morphology based methods are another solution to detect tumor area from MRI images [24-26]. It is important to use appropriate stages in morphology methods to achieve accurate results. Moussaoui et al., (2023) used a clustering method based on watershed and morphological operators to detect tumor area. The proposed method has an accuracy of 99 % based on sensitivity measure [27]. Watershed segmentation offers a more sophisticated approach by considering the image as a topographic surface. This method effectively allocates tumor boundaries by identifying regions based on intensity gradients [28]. While watershed segmentation excels in separating tumor or notumor classes, it may be faced to over-segmentation, therefore post-processing techniques for refinement must be employed [29].

In this paper, we prepare a comparison regarding the accuracy of three tumor detection methods based on K-means and Genetic Algorithm (GA) clustering, Otsu thresholding method, and watershed segmentation approach.

Materials and Methods

Data used

In this study, three MRI images are used to assess the proposed method. Figure 1 shows the

employed images in this study. The images consist of one channel.

Figure 1. The Original MRI Images and Ground Truth Used in This Study.

To assess the performance of the proposed method, a ground truth image according to Figure 1 was generated by an expert. The generated images highlight tumor regions with a blue color in the images.

Methodology

In this study, three segmentation methods to detect tumor areas from images are employed. These methods are popular in image processing applications and employ different ways for image segmentation. Therefore, it is interesting to compare them. The methods are applied in three main steps including: 1) image segmentation, 2) find tumor segment, 3) accuracy assessment. The following section provides more detailed information of each method (Figure 2).

Figure 2. The Proposed Workflow Used in This Study.

Method1: K-means+GA clustering

K-means is a well-known clustering method in image processing, which was used in previous studies [30, 31]. In the K-means clustering method, cluster centers are estimated in an iterative approach. The centers in the K-means clustering method are calculated using the mean of pixel values in each cluster. To improve cluster center estimation, various approaches were used in the literature [32, 33]. Genetic Algorithm (GA) is one of approaches that has been used to estimate cluster centers in previous studies [34, 35]. In this paper, the following steps are used in method 1 to detect tumor areas:

Step1: Read image

Step2: Set GA parameters including number of iterations, number of clusters, number of populations, mutation, crossover and elitism rates

Step3: Run GA to find optimum cluster centers

Step4: Assign each pixel to nearest and optimumcluster center based on K-means clustering

Step5: Find cluster number regarding tumor area

Step6: Assign pixels in a region to a segment

Step7: Extract area and circularity of segments

Step8: To remove noise segments, tumor segment is detected using two threshold values on area and circularity Step9: Accuracy assessment on tumor segment detected.

Method2: Otsu clustering

Otsu clustering is a familiar method to segment images based on threshold methods [36, 37]. This method uses threshold values to cluster pixels of an image. It finds appropriate threshold values based on the image's histogram. Otsu clustering was used in previous studies in different image processing applications such as tumor detection [38-40]. For applying Otsu clustering in Method2, the following steps are considered:

Step1: Read image

Step2: Assign each pixel to a cluster based on Otsu clustering

Step3: Find cluster number regarding tumor area

Step4: Assign pixels in a region to a segment

Step5: extract area and circularity of segments

Step6: to remove noise segments, tumor segment is detect using two threshold values on area and circularity

Step7: Accuracy assessment on tumor segment detected

Method3: Watershed Clustering

Watershed clustering is a technique commonly used in image segmentation and clustering that is based on the concept of watersheds in geography [41, 42]. It is often employed in the context of image analysis and processing, especially for separating distinct objects or regions within an image. To apply watershed clustering to tumor detection, pre-processing steps are required. To this end, morphological operators are used to prepare image to enter this technique [42-44]. The steps involved in method3 are presented below:

Step1: Read image

Step2: Convert RGB image to gray scale one

Step3: Enhance Image based on histogram

Step4: Apply erode on image of Step3

Step5: Reconstruct a new image using images of steps 3 and 4

Step6: Apply dilate on image of Step4

Step7: Reconstruct a new image using images of steps 5 and 6

Step8: Perform Watershed clustering

Step9: Modify output of watershed clustering using local minimums

Step10: Accuracy assessment

Accuracy Assessment

Accuracy assessment is an essential step in image processing algorithms. In this study, the accuracy of detecting tumor and non-tumor regions should be evaluated. To this end, a ground truth image that delineates tumor and no-tumor regions is used. The ground truth images were generated by an expert. It is necessary to compare the outputs of the three methods with the ground truth images. To obtain a statistical measure, a confusion matrix is utilized. Some statistics such as overall, user, and producer accuracies can be derived from the confusion matrix, which have been employed in previous studies [45-48].

Results and Discussion

Figure 3 shows the outputs of K-Means+GA method for the three images. As shown, the initial output of the method can detect tumor areas in all three images (Figure 3 a-c).

Figure 3. The Result of K-means+GA Clustering.

After creating the binary image, it becomes evident that non-tumor areas are also detected, since their digital numbers are similar to tumor areas (Figure 3 d-f). To remove undesired areas, we segment connected pixels in to similar groups. Then, area and circularity of each segment are calculated.

Subsequently, segments with an area higher than 300 pixels and a circularity greater than 0.5 are considered as tumor areas. As seen in Figure 3 (e-h), the threshold values are appropriate to detect tumor areas in this study. A visual analysis among the detected tumor areas and ground truth images shows that the first method is robust in detecting regions of interest.

The output of the second method i.e., Otsu thresholding is presented in Figure 4.

Figure 4. The Result of Otsu Thresholding Method in Tumor Detection.

According to Figure 4 (a-c), the second method is appropriate to detect tumor areas (yellow color), but there are some noise areas that must be removed (Figure 4 d-f). To remove undesired areas, segments with an area higher than 300 pixels and circularity greater than

0.5 are selected as tumor areas. After modification, tumor areas are detected according to Figure 4 (g-h). It seems that the second method is appropriate to detect tumor areas, since results are similar to the ground truth images. The third method for tumor detection is based on watershed segmentation. Results of the third method is presented in Figure 5.

Figure 5. The Result of Watershed Segmentation in Tumor Detection.

The image enhancement step (Figure 5 (a-c)) proves effective in separating tumors from undesired areas. Then, the erode operator can remove white areas in border of images (Figure 5 (d-f)). Afterwards, the reconstructed image is presented in Figure 5 (g-h). the dilation operator is used to expand tumor area according to Figure 5 (i-k). Final complemented image is presented in Figure 5 (l-n). As can be seen, bright areas in border of images are removed, and allowing for the identification of tumor areas. Finally, results of watershed segmentation are presented in Figure 5 (o-q). These figures suggest that the third method can successfully detect tumor areas.

Output of confusion matrix for K-means+GA method in tumor detection is presented in Figure 6.

Figure 6. The Confusion Matrix for K-means+GA Clustering for Image 1 to 3.

Figure 6 (a) shows the confusion matrix for image 1. 255990 and 5701 pixels regarding no-tumor and tumor areas were correctly detected. 368 pixels of tumor are incorrectly classified as no-tumor class. Moreover, 85 pixels of no-tumor are assigned tumor incorrectly. About 6 percent error in detection of tumor area is observed.

Figure 6 (b) and (c) are presented confusion matrix for Images 2 and 3, respectively. As can be seen, the accuracy of detection tumor area is higher than Image 1. An error of 2 to 3 percent is observed in detection of tumor area in the mentioned images.

Figure 7 shows the confusion matrix of the Otsu thresholding method for Images 1 to 3. Similar to K-means+GA method, error in image 1 in the second method is higher than images 2 and 3.

Figure 7. The Confusion Matrix for Otsu Thresholding Method in Images 1 to 3.

In Image 1, 368 pixels of tumor area were assigned as no-tumor class, which is about 6 percent error. The accuracy of tumor detection in image 2 and image 3 is more than 97 percent that is acceptable. Overall, the Otsu thresholding method is robust enough in separating tumor and no-tumor classes.

The results of third proposed method based on watershed segmentation are presented in Figure 8.

Figure 8. The Confusion Matrix for Watershed Segmentation in Images 1 to 3.

Accuracy of separating tumor and no-tumor areas in images 2 and 3 is higher than image 1. In Image 1, 510 pixels of tumor area were incorrectly classified as non-tumor area, which indicate approximately 9 percent error. Errors in no-tumor area are lower than tumor one. In image 2, 257,421 and 4,592 pixels were correctly classified as no-tumor and tumor classes, respectively.

In this section, a comprehensive comparison among the segmentation methods is presented. Based on the confusion matrix of the proposed method, the main errors are observed in the tumor area, with an error rate of approximately 3 to 7 percent. According to outcomes, the performance of the K-means+GA and Otsu threshold methods is similar to each other. It shows that the mentioned methods can detect same areas as tumor. In fact, tumor area in images 1 to 3 can be detected accurately with effective clustering and thresholding methods. Moreover, the accuracy of watershed method is lower than the other two methods. It seems that orphological operators should be employed more effectively to minimize errors.

In conclusion, in this paper, three tumor detection methods including K-means+GA, Watershed, and Otsu were employed to separate tumor from non-tumor areas. The proposed methods were applied to three images. Based on results, three employed methods yielded acceptable outcomes. The main errors were observed in detecting tumor areas, with an error rate of approximately 3 to 7 percent in images. The K-means+GA and Watershed methods showed the minimum error over the three images.

Notably, errors of methods 1 and 2 in image 1 are higher than images 2 and 3. The accuracy of method 3 is lower than methods 1 and 2. It seems that morphological operators may affect the accuracy of results. Future study, could utilize other morphological operators to improve the accuracy of results. Moreover, use of other features such as texture for image segmentation may be

beneficial. Finally, it is advised to use other images and new segmentation methods to separate tumor and non-tumor areas.

Acknowledgements

General

The author would like to thank the following link for providing the used images in this study kaggle.com/datasets/masoudnickparvar/brain-tumor-mri- dataset .

Funding Statement

Approval

Not applicable.

Not applicable.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical Declaration

Not applicable.

Authors Contribution

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Sonia Rezaei. The first draft of the manuscript was written by Sonia Rezaei and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

References

- 1. Abd-Ellah MK, Awad AI, Khalaf AAM, Hamed HFA. A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned. *Magnetic Resonance Imaging*. 2019; 61DOI
- 2. Soomro TA, Zheng L, Afifi AJ, Ali A, Soomro S, Yin M, Gao J. Image Segmentation for MR Brain Tumor Detection Using Machine Learning: A Review. *IEEE reviews in biomedical engineering*. 2023; 16DOI
- 3. Abdalla HEM, Esmail MY, editors. Brain tumor detection by using artificial neural network. 2018 International conference on computer, control, electrical, and electronics engineering (ICCCEEE). 2018: IEEE.
- 4. Tseng C, Tang C. An optimized XGBoost technique for accurate brain tumor detection using

- feature selection and image segmentation. Healthcare Analytics. 2023; 4DOI
- 5. Kanchanamala P, Revathi K, Ananth MBJ. Optimization-enabled hybrid deep learning for brain tumor detection and classification from MRI. *Biomedical Signal Processing and Control*. 2023; 84:104955.
- 6. Saeedi S, Rezayi S, Keshavarz H, R Niakan Kalhori S. MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC medical informatics and decision making*. 2023; 23(1)DOI
- 7. Jabbar A, Naseem S, Mahmood Ta, Saba T, Alamri FS, Rehman A. Brain Tumor Detection and Multi-Grade Segmentation Through Hybrid Caps-VGGNet Model. *IEEE Access*. 2023; 11DOI
- 8. Solanki S, Singh UP, Chouhan SS, Jain S. Brain Tumor Detection and Classification Using Intelligence Techniques: An Overview. *IEEE Access*. 2023; 11<u>DOI</u>
- 9. Rahman T, Islam MS. MRI brain tumor detection and classification using parallel deep convolutional neural networks. *Measurement: Sensors*. 2023; 26DOI
- 10. Alam MS, Rahman MM, Hossain MA, Islam MK, Ahmed KM, Ahmed KT, Singh BC, Miah MS. Automatic Human Brain Tumor Detection in MRI Image Using Template-Based K Means and Improved Fuzzy C Means Clustering Algorithm. *Big Data and Cognitive Computing*. 2019; 3(2)DOI
- 11. Prakash RM, Bhuvaneshwari K, Divya M, Sri KJ, Begum AS, editors. Segmentation of thermal infrared breast images using K-means, FCM and EM algorithms for breast cancer detection. 2017 International conference on innovations in information, embedded and communication systems (ICIIECS); 2017: IEEE.
- 12. Ahmmed R, Hossain MF, editors. Tumor detection in brain MRI image using template based K-means and Fuzzy C-means clustering algorithm. 2016 International Conference on Computer Communication and Informatics (ICCCI); 2016: IEEE.
- 13. Babu KR, Singal A, Sahiti K. Performance analysis of brain tumor detection using optimization based FCM technique on MRI images. *Int J Sci Technol Res.* 2020; 8(11):1717-1722.
- 14. Macqueen J, editor Some methods for classification and analysis of multivariate observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability/University of California Press. 1967.
- 15. Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*. 2010; 31(8)DOI
- 16. Golberg DE. Genetic algorithms in search, optimization, and machine learning. Addion wesley. 1989;1989(102):36..
- 17. Laddha RR, Ladhake S. A review on brain tumor detection using segmentation and threshold operations. International Journal of Computer Science and Information Technologies. 2014;5(1):607-11..
- 18. Ilhan U, Ilhan A. Brain tumor segmentation based on a new threshold approach. *Procedia Computer Science*. 2017; 120DOI
- 19. Tarhini GM, Shbib R. Detection of Brain Tumor in MRI Images Using Watershed and Threshold-Based Segmentation. *International Journal of Signal Processing Systems*. 2020; 8(1):19-25. DOI
- 20. Verma A, Ansari M, Tripathi P, Mehrotra R, Shadab SA. Brain tumor detection through MRI using image thresholding, k-means, and watershed segmentation. *Computational Intelligence in Healthcare Applications: Elsevier*. 2022;267-83.
- 21. Nasor M, Obaid W, editors. MRI tumor detection and localization by multiple threshold object counting technique. 2018 International Conference on Computer and Applications (ICCA). 2018: IEEE.
- 22. Sudharani K, Prasad TCS, Satya K. Histogram Related Threshold Technique for Region based Automatic Brain Tumor Detection. *Indian Journal of Science and Technology*. 2016; 9(48)DOI
- 23. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics.* 1979; 9(1)DOI
- 24. Sudharani K, Sarma T, Prasad KS. Advanced Morphological Technique for Automatic Brain

- Tumor Detection and Evaluation of Statistical Parameters. *Procedia Technology*. 2016; 24:1374-1387. DOI
- 25. Devkota B, Alsadoon A, Prasad P, Singh A, Elchouemi A. Image segmentation for early stage brain tumor detection using mathematical morphological reconstruction. *Procedia Computer Science*. 2018; 125:115-123. <u>DOI</u>
- 26. Oo SZ, Khaing AS. Brain tumor detection and segmentation using watershed segmentation and morphological operation. *International Journal of Research in Engineering and Technology.* 2014; 3(3):376-374. DOI
- 27. Moussaoui H, El Akkad N, Benslimane M. A Brain Tumor Segmentation and Detection Technique Based on Birch and Marker Watershed. *SN Comput. Sci.*. 2023; 4(4)DOI
- 28. Vincent L., Soille P.. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1991; 13(6)DOI
- 29. Meyer F. Topographic distance and watershed lines. Signal Processing. 1994; 38(1)DOI
- 30. Sinaga KP, Yang M. Unsupervised K-Means Clustering Algorithm. IEEE Access. 2020; 8DOI
- 31. Na S, Xumin L, Yong G, editors. Research on k-means clustering algorithm: An improved k-means clustering algorithm. 2010 Third International Symposium on intelligent information technology and security informatics. *IEEE*. 2010.
- 32. Ahmadyfard A, Modares H, editors. Combining PSO and k-means to enhance data clustering. 2008 international symposium on telecommunications. *IEEE*. 2008.
- 33. Niknam T, Amiri B. An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. *Applied Soft Computing*. 2010.
- 34. Kapil S, Chawla M, Ansari MD, editors. On K-means data clustering algorithm with genetic algorithm. 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC). *IEEE*. 2016.
- 35. Khotimah BK, Irhamni F, Sundarwati T. A Genetic algorithm for optimized initial centers K-means clustering in SMEs. *Journal of Theoretical and Applied Information Technology*. 2016; 90(1):23.
- 36. Yang P, Song W, Zhao X, Zheng R, Qingge L. An improved Otsu threshold segmentation algorithm. *International Journal of Computational Science and Engineering*. 2020; 22(1):146-53. DOI
- 37. Srinivas C, C, Prasad M, M, Sirisha M. M. Remote sensing image segmentation using OTSU algorithm. *International Journal of Computer Applications*. 2019; 975:8887.
- 38. Sharma A, Kumar S, Singh SN. Brain tumor segmentation using DE embedded OTSU method and neural network. *Multidimensional Systems and Signal Processing*. 2019; 30(3)DOI
- 39. Mittal K, Shekhar A, Singh P, Kumar M. Brain Tumour Extraction using Otsu Based Threshold Segmentation. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2017; 7(4)DOI
- 40. Husham S, Mustapha A, Mostafa SA, Al-Obaidi MK, Mohammed MA, Abdulmaged AI, George ST. Comparative Analysis between Active Contour and Otsu Thresholding Segmentation Algorithms in Segmenting Brain Tumor Magnetic Resonance Imaging. *Journal of Information Technology Management*. 2020; 12(Special Issue: Deep Learning for Visual Information Analytics and Management.)DOI
- 41. Annavarapu A, Borra S. An adaptive watershed segmentation based medical image denoising using deep convolutional neural networks. *Biomedical Signal Processing and Control*. 2024; 93DOI
- 42. Liu Y, Chen D, Fu S, Mathiopoulos PT, Sui M, Na J, Peethambaran J. Segmentation of Individual Tree Points by Combining Marker-Controlled Watershed Segmentation and Spectral Clustering Optimization. *Remote Sensing*. 2024; 16(4):610. DOI
- 43. Jayanna KK, Shan PB. Fusion-Based Morphological Watershed Segmentation Algorithm for Medical Images. *Research Advances in Intelligent Computing: CRC Press.* 2023;173-184.
- 44. Kavitha K, Shan PB. Medical image watermarking based on novel encoding for EHR and fusion based morphological watershed segmentation algorithm for medical images. *Multimedia Tools and Applications*. 2024; 83(9):25163-25190. DOI

- 45. Heydarian M, Doyle TE, Samavi R. MLCM: Multi-Label Confusion Matrix. *IEEE Access*. 2024; 10:19083-19095. DOI
- 46. Freitas CO, De Carvalho JM, Oliveira J, Aires SB, Sabourin R, editors. Confusion matrix disagreement for multiple classifiers. Progress in Pattern Recognition, Image Analysis and Applications: 12th Iberoamericann Congress on Pattern Recognition, CIARP 2007, Valparaiso, Chile, November 13-16, 2007 Proceedings 12; 2007: Springer.
- 47. Theissler A, Thomas M, Burch M, Gerschner F. ConfusionVis: Comparative evaluation and selection of multi-class classifiers based on confusion matrices. *Knowledge-Based Systems*. 2022; 247:108651. DOI
- 48. Tehrani NA, Farhanj F, Janalipour M. Introducing a novel dust source identification method based on edge points and paths extracted from integration of time-series MODIS products. *Remote Sensing Applications: Society and Environment.* 2023; 32:101054.