Dominant Gray Level-based Genetic K-means Clustering Algorithm for MRI Image Segmentation

Maha Ibrahim Khaleel

Department of Computer Technologies Engineering, Alsafwa University College, Iraq
maha.ibrahim@alsafwa.edu.iq (corresponding author)

Musab Ahmed Mohammed

Department of English, University of Kirkuk, Iraq
mcs.musab@uokirkuk.edu.iq

Maryam Qays

Department of Clinical Laboratory Sciences, Alzahraa University for Women, Iraq
maryam.qays@alzahraa.edu.iq

ABSTRACT

In this paper, a method and fresh results associated with medical image segmentation of brain Magnetic Resonance Imaging (MRI) scans are presented. Gray-converted segmentation and Genetic Algorithm (GA) are utilized along with unsupervised k-means classification. The image segmentation employed indicates the tissue type or the anatomical structure of each pixel. The cluster centroid initialization is performed by GA. GA offers efficient search processes (selection, crossover, and mutation), suited to determine global optima regarding centroid problems. As a result, this research offers more accurate, reliable, and efficient image segmentation for MRI, by improving the k-means algorithm with GA. The results indicate that the accuracy obtained from the proposed method is at least 3.5% higher than the PSO algorithm in this matter.

Keywords-image processing; image segmention; k-means algorithm; clustering; genetic algorithm; MRI

I.   INTRODUCTION

The use of Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) images has significantly benefited treatment planning, medical evaluations, and clinical research. MRI scans, known for their high precision, provide detailed three-dimensional information about the soft tissues within the human body. Unlike the X-ray imaging, MR images do not necessitate exposure to radiation during the scanning process. Consequently, MRI scans have gained increasing popularity [1]. A range of diagnostic techniques are utilized to ascertain the underlying cause of a patient's symptoms. These approaches may involve performing a biopsy [2-5] or utilizing various imaging methods, and conducting either a brain MRI or a CT scan. The accurate diagnosis of brain tumors is a significant challenge for medical science, with particular emphasis given on MR) due to the predominantly grayscale nature of the MRI images. MRI imaging, in contrast to other imaging techniques, offers high contrast value [6]. The dominant gray level problem often pertains to challenges associated with analyzing and interpreting grayscale intensity variations in MR images. The specific problem related to dominant gray levels in MR images is that they often exhibit various shades of gray representing different types of tissues, such as gray matter, white matter, cerebrospinal fluid, and lesions. The challenge lies in accurately differentiating between these tissues based on their dominant gray levels, which is crucial for diagnosis and treatment planning. In this paper, we specifically selected an axial view of a brain image from an MRI scan due to the lower risk associated with MRI scans compared to CT brain scans.

Clustering entails the categorization of a set of objects into groups sharing similar traits, and it is widely deployed in diverse fields including statistics, machine learning, pattern recognition, data mining, and image processing .Clustering research is typically categorized into two main groups: hierarchical methods and partitioning methods [7]. There are several clustering techniques within these categories, including techniques like fuzzy C-means, hierarchical clustering, k-means, spectral clustering, Gaussian mixture models, etc. This study specifically introduces the k-means method to address

certain limitations [8]. Essentially, the basic objective of any clustering algorithm is to find a global or even approximate optimum. Typically, the k-means criteria may lead convergence to a suboptimal region [9].

The k-means clustering algorithm is based on the optimization of the similarity scale between each cluster with the lowest value and the highest value for the values within the clusters. In other words, the k-means algorithm attempts to reduce the similarity among different clusters and increase the similarity within a cluster. The k-means clustering technique initializes k cluster centroids randomly. Through multiple iterations of the algorithm, data are assigned to specific clusters based on a criterion function, which aims to place similar data points within the same cluster and maintain a considerable separation between clusters. The k-means algorithm is a fundamental partitioning method in cluster analysis. However, it is sensitive to the initial positioning of the cluster centroids, and different initial centroids can yield notably diverse outcomes. Isolated data points also have a substantial impact on the algorithm. In this paper, the optimal value of k is optimized using the Genetic Algorithm (GA), which is considered one of the most efficient optimization algorithms [10]. The purpose of clustering is to enhance the probability of obtaining the global optimum within the solution area (exploration), yet guaranteeing sufficient pressure to acquire even better options from the current individuals (exploitation). Additionally, in sensible multimodal domain challenges, it is advisable to evaluate various global optima or some localized optima that could be suitable alternatives for the global optima. Standard GAs perform well on single optimization problems but do not provide multiple alternatives. By combining the strength of clustering evaluation and GA search, the proposed k-GA procedures permit the assessment of multimodal features [11-13].

This paper aims to utilize GA to harness its adaptive search capabilities in order to identify the cluster centroids of k-means. Additionally, it seeks to improve the k-means algorithm by incorporating the method proposed in [14] to address the impact of isolated data points. A GA is an adaptable global optimization search method that emulates the principle of survival of the fittest in a natural setting. Its core components include gene coding, fitness evaluations, population initialization, and the execution of evolutionary operations, namely selection, crossover, and mutation [15]. The effective use of a systematic chromosome selection method in the initial population enables the identification and capture of clusters with diverse shapes and sizes [16].

Image segmentation is a fundamental process that involves categorizing pixels into meaningful image sections corresponding to the objects within the image. Various criteria, including gray value, texture, mobility, and shape, can be utilized to classify these regions as homogeneous entities [17]. Image segmentation is a crucial preliminary phase in image processing and has diverse applications in fields, such as medical imaging, satellite object detection, face recognition, traffic control systems, and machine vision [18]. Image segmentation methods can be broadly classified into two categories: region-based and edge-based. While edge-based

techniques primarily focus on identifying edges, which may sometimes be fragmented or incomplete, region-based approaches rely on groups of pixels with similar characteristics, such as gray-level intensity information or color image components [19]. Segmentation involves dividing a single image into multiple sections, creating distinct sets of pixels from the same image. This process is essential for further analysis and the extraction of relevant information from an image. It is often referred to as a labeling procedure that assigns similar characteristics to each pixel in an image. This operation results in the creation of pixel clusters that share common properties. In the context of medical image processing, segmentation holds significant importance [20]. The k-means clustering technique is well-suited for image segmentation and is among the most frequently utilized segmentation methods.

Head MRI employs a predominant magnetic domain, transmission waves, and a computer to generate an exceptionally detailed image of the brain, surpassing the level of clarity attained by alternative imaging techniques. In recent years, many researchers have adopted computer-assisted methods for image segmentation [8]. GAs have gained widespread application in addressing clustering challenges due to their advantages in seeking optimal solutions in data clustering tasks. In various data clustering scenarios, combining GAs with clustering techniques, such as k-means, yields more interpretable results than using a standalone clustering algorithm. The GKMC technique was initially developed to achieve a globally optimal partition in the field of pattern recognition, and it has demonstrated its ability to converge to an optimal solution [21].

The proposed solution performs MRI image segmentation and clustering applying k-means clustering with GA.

II.   THE PROPOSED METHOD

In this paper, a method for segmenting brain tissues in MR images based on k-means objective function combined with GA is introduced. Distinct gray-level areas are used to express different forms of brain tissue (white matter-WM, gray matter-GM, and cerebrospinal fluid-CSF). Gray-levels can be classified based on gray-value similarity, and gray-values classified into the same type are regarded to be the same tissue. Despite the fact that clustering algorithms have a disadvantage in accuracy due to their non-global optimal searching ability, they have been used in MR brain picture segmentation.

Due to its relative ease of implementation and ability to scale well to very big datasets, k-means is one of the most common clustering methods, but it has certain drawbacks. It does not ensure that the global optimum clustering solution will be found. In fact, outliers and noisy data can make the algorithm quite sensitive: the quality of the final clustering can be greatly dependent on the location of the initial cluster centroids. To put it another way, k-means will frequently find a local rather than a global minimum. As a result, segmentation results are frequently locked in a local optimum. We employed an evolutionary method to enhance the accuracy and discover the ideal place for the initial cluster centroids. The GA is simple to implement, is suitable for noisy environments, and

may be utilized in distributed and parallel modes, as well as for multi-objective tasks. Figure 1 depicts the flowchart of the proposed method.
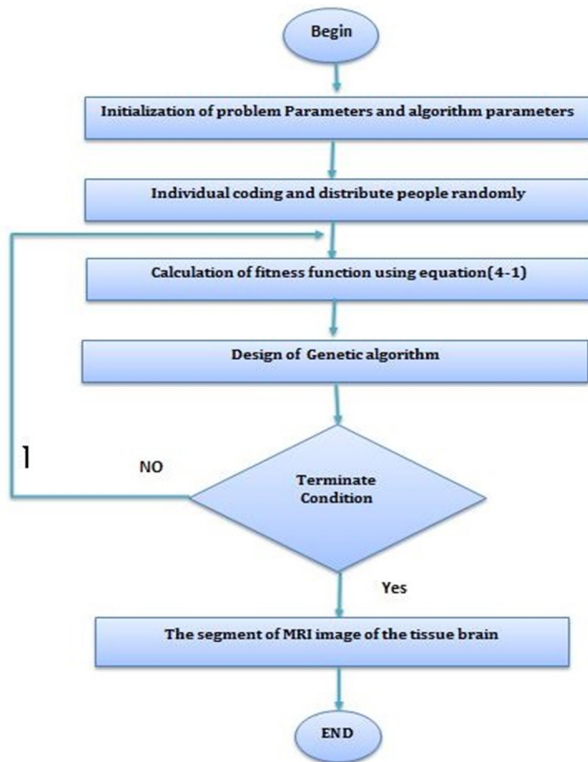

Fig. 1.   Flowchart of the proposed method.

III.   STEPS OF THE PROPOSED METHOD

The proposed method comprises six essential steps:

1) Parameter Initialization

This section entails the configuration of four critical parameters, namely 'npop,' 'K,' 'MaxIteration,' 'Pc,' and 'Pm.' 'npop' refers to the population size, representing the number of individuals within the population. As 'npop' increases, the population becomes larger, but this expansion also extends the time required for computation. Typically, values for 'npop' are chosen within the range of 30 to 100. 'K' denotes the user-defined number of clusters. 'MaxIteration' represents the maximum number of iterations allowed before terminating the process. A higher 'MaxIteration' value results in a longer computation time. 'Pc' is the probability of crossover, a factor that regulates the creation of new individuals during the initial stages of evolution. The choice of 'Pc' is crucial. If set too small, the evolution may converge rapidly, possibly getting trapped in local optima. On the other hand, if 'Pc' is excessively large, the evolution may not converge. It is common to select 'Pc' values within the range of 0.4 to 0.9. 'Pm' denotes the probability of mutation, and if it is set too high, it can lead to the disruption of many high-quality individuals, potentially preventing the evolutionary process from converging. Typically, 'Pm' values are chosen within the range of 0.001 to 0.1.

2) Individual Coding and Random Population Distribution

The length of each chromosome corresponds to the number of clusters, and each string inside a chromosome contains a sequence of real numbers that describes the set of k cluster centers for the problem. The k cluster centers are chosen at random from the dataset, and each chromosome is initialized in accordance with this to generate the initial population. This initialization process is repeated for every chromosome. It is important to note that the data in this example are categorized into three groups, which correspond to the three different types of brain tissues.

3) Calculation of the Fitness Function

In this step, the value of the objective function is calculated. To assess the adaptability of each individual during the evolutionary process, a fitness function is employed. In this context, the fitness function is derived from the k-means algorithm and is utilized for MRI segmentation. The equation representing the fitness function is:

$$E_k^2 = -\sum_{k=1}^{k} \sum_{x \in c_k} ||X - c_k||^2 \qquad (1)$$

where k represents the classification number, X represents the input data, and c_k is the center of the kth cluster.

4) Design of Genetic Operations

Selection was computed according to:

$$P_i = c(1 - c)^{i-1} \qquad (2)$$

where c represents a constant ($c = 0.4$), and i denotes the ith individual. Subsequently, the roulette wheel selection method was applied. In each iteration, a random real number within the range of 0 to 1 was generated, and this number functioned as the pointer for selecting individuals. As an illustration, if the random number happened to be 0.784, it would correspond to the selection of the 3rd individual, as demonstrated in Table I.

TABLE I.   THE ROULETTE WHEEL SELECTION

I	1	2	3
P	0.4	0.24	0.144
$\sum P$	0.4	0.64	0.784

Crossover is a crucial process in GA, contributing to the creation of improved individuals and enhancing the diversity within the population. It plays a vital role in preventing the algorithm from quickly converging to local optima, thereby promoting the exploration process. In uniform crossover, two parent individuals are chosen based on a selection function, and their genetic information is combined to generate an offspring with a mix of their traits. Generating a random vector within the interval [0,1], denoted as R_1 and R_2, where $1-R_1$ is also considered. Chromosomes x_1 and x_2 are chosen, and the new offspring, Y_1 and Y_2 are produced:

$$Y_1 = (R_1 * x_1) + (R_2 * x_2) \qquad (3a)$$

$$Y_2 = (R_2 * x_1) + (R_1 * x_2) \qquad (3b)$$

where the crossover rate is 0.8.

Mutation involves the random alteration of one or more gene bits within an individual, leading to increased variety

among individuals. Every individual has the chance to experience mutation, with a random probability of mutation affecting a gene bit, resulting in the creation of a new individual. In this case, the mutation rate is set to 0.2.

5) Termination Condition

Steps 3 and 4 were repeatedly executed until the optimal positions of the cluster centroids' were reached. This entailed monitoring the difference between the centroid matrices from the last two iterations, ensuring that it diminished to a value below a specific constant.

6) Result

The result of this approach is a segmented MRI image.

IV. DATASET

MRI brain scans from the Montreal Neurology Institute at McGill University [22] were used to gather data for the investigation.

V. EVALUATION

To assess the efficiency and precision of the proposed method in segmenting the MRI dataset, accuracy was employed as an evaluation metric. The proposed method was compared to that of the PSO algorithm [21], which has demonstrated satisfactory performance in this domain. Accuracy is computed by:

$$\text{Accuracy} = \frac{\left(TP + TN\right)}{\left(TP + TN + FP + FN\right)} \quad (4)$$

where TP = True Positive, FP = False Positive, TN = True Negative, and FN = False Negative.

VI. SIMULATION RESULTS

The acquired outcomes of the proposed approach are presented below. The proposed technique was executed independently 20 times and the mean accuracy was considered. In the current study, 40 individuals and a maximum of 350 iterations were taken into account. The image dimensions were 250×250 pixels. In case an image was in RGB format, it was converted to grayscale. The current study opted for 4 centroids for the k-means clustering process.

A. Sample 1

In the first sample, the accuracy of the proposed algorithm is 80%, while the PSO algorithm reached 82%. Figure 2 depicts the original and the segmented image, while Figure 3 portrays the GA's fitness value versus the generation number.

B. Sample 2

In the second example, the accuracy of the proposed algorithm is 93%, surpassing the PSO that reached 85.76%. Figure 4 exhibits the original and the segmented image, whereas Figure 5 displays the GA's fitness value versus the generation number.
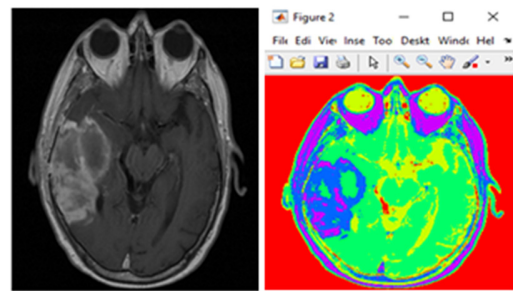

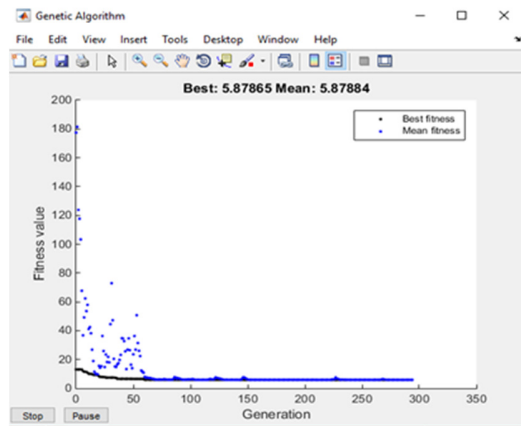Fig. 2.   Original and segmented image 1.

Fig. 3.   Diagram obtained from the implementation of GA to improve k-means centroids for image 1.
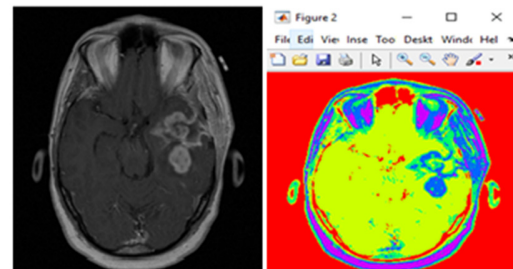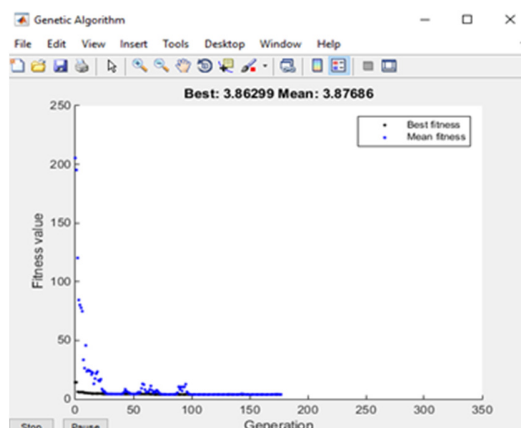

Fig. 4.   Original and segmented image 2.

Fig. 5.   Diagram obtained from the implementation of GA to improve k-means centroids for image 2.

C. Sample 3

In the third example, the accuracy of the proposed algorithm is 45%, whereas the PSO algorithm reached 42.3%. Figure 6 illustrates the original and the segmented images, whereas Figure 7 shows the GA's fitness value versus the generation number.
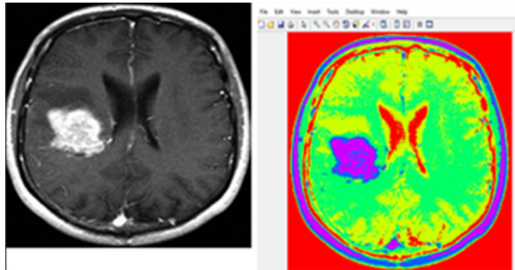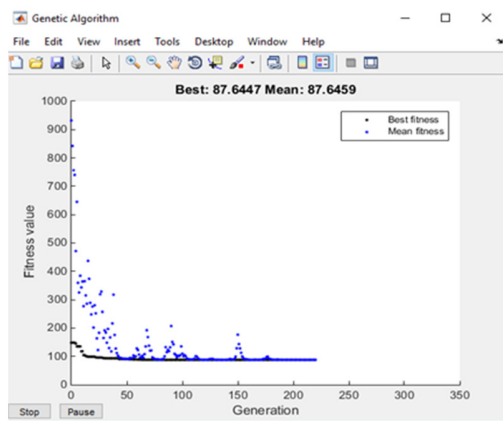

Fig. 6.   Original and segmented image 3.

Fig. 7.   Diagram obtained from the implementation of GA to improve k-means centroids for image 3.

D. Sample 4

In the fourth example, the accuracy of the proposed algorithm is 93%, whereas the PSO algorithm reached 87%. Figure 8 demonstrates the original and the segmented image, whereas Figure 9 showcases the GA's fitness value versus the generation number.
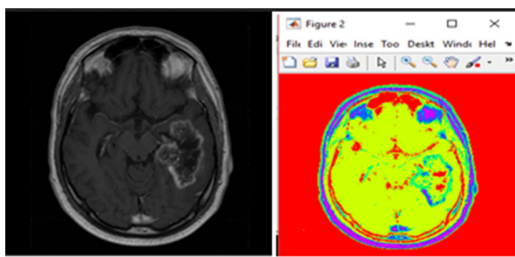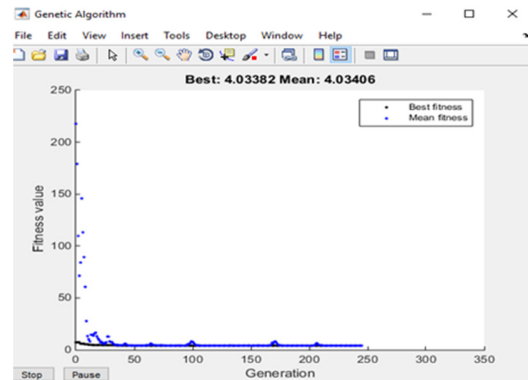

Fig. 8.   Original and segmented image 4.

Fig. 9.   Diagram obtained from the implementation of GA to improve k-means centroids for image 4.

VII.   CONCLUSION

Clustering is an important unsupervised learning problem. A cluster is a collection of similar objects that are dissimilar to the objects belonging to other clusters. A good clustering method will produce high quality clusters with high intra-class similarity and low inter-class similarity.

In this paper, a method was proposed for the segmentation of MRI brain tissue images. GA was utilized to find the best position of the cluster centroids, because the k-means approach to cluster the process is very sensitive to the position of the initial cluster centroids. The recommended method improved segmentation accuracy and speed. The results were validated and compared with the ones of PSO deploying the Accuracy metric. Segmenting accuracy and speed could be further improved. Therefore suggested future work could involve the use of adaptive Pc, Pm, and nPop values during the execution of the genetic algorithm and the utilization of feature selection for getting the best properties of the considered data set as a preprocessing step, since considering all the features in a dataset is not salutary, especially when the feature number is high.

REFERENCES

[1]   I. O. Tehrani, "Spatial Fuzzy C-Mean Sobel Algorithm with Grey Wolf Optimizer for MRI Brain Image Segmentation," Ph.D. dissertation, Universiti Teknologi Malaysia, Johor, Malaysia, 2017.

[2]   P. J. Rohini, C. S. Singh, and M. Manikandan, "Brain Tumor MRI Image Segmentation and Detection in Image Processing," *International Journal of Research in Engineering and Technology*, vol. 3, no. 13, pp. 1–5, Mar. 2014, https://doi.org/10.15623/ijret.2014.0313001.

[3]   R. Ahmmed and Md. F. Hossain, "Tumor detection in brain MRI image using template based K-means and Fuzzy C-means clustering algorithm," in *2016 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, Jan. 2016, https://doi.org/10.1109/ICCCI.2016.7479972.

[4]   S. N. Sulaiman and N. A. Mat Isa, "Adaptive fuzzy-K-means clustering algorithm for image segmentation," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2661–2668, Nov. 2010, https://doi.org/10.1109/TCE.2010.5681154.

[5]   D. Mustafi, A. Mustafi, and G. Sahoo, "A novel approach to text clustering using genetic algorithm based on the nearest neighbour heuristic," *International Journal of Computers and Applications*, vol. 44, no. 3, pp. 291–303, Mar. 2022, https://doi.org/10.1080/1206212X.2020.1735035.

[6] D. R. F. R. Anandhi and S. Sathiamoorthy, "Enhanced Sea Horse Optimization with Deep Learning-based Multimodal Fusion Technique for Rice Plant Disease Segmentation and Classification," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11959–11964, Oct. 2023, https://doi.org/10.48084/etasr.6324.

[7] I. Tebbal and A. F. Hamida, "Effects of Crossover Operators on Genetic Algorithms for the Extraction of Solar Cell Parameters from Noisy Data," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10630–10637, Jun. 2023, https://doi.org/10.48084/etasr.5417.

[8] J. Liu and L. Guo, "A New Brain MRI Image Segmentation Strategy Based on K-means Clustering and SVM," in *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Hangzhou, China, Dec. 2015, vol. 2, pp. 270–273, https://doi.org/10.1109/IHMSC.2015.182.

[9] W. Min and Y. Siqing, "Improved K-means clustering based on genetic algorithm," in *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, Oct. 2010, vol. 6, pp. V6-636-V6-639, https://doi.org/10.1109/ICCASM.2010.5620383.

[10] M. A. El-Shorbagy, A. Y. Ayoub, I. M. El-Desoky, and A. A. Mousa, "A Novel Genetic Algorithm Based k-means Algorithm for Cluster Analysis," in *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018)*, 2018, pp. 92–101, https://doi.org/10.1007/978-3-319-74690-6_10.

[11] D. Zeebaree, H. Haron, A. Mohsin Abdulazeez, and S. Zeebaree, "Combination of K-means clustering with Genetic Algorithm: A review," *International Journal of Applied Engineering Research*, vol. 12, no. 24, pp. 14238–14245, Dec. 2017.

[12] Z. Khan, J. Ni, X. Fan, and P. Shi, "An improved K-means clustering algorithm based on an adaptive initial parameter estimation procedure for image segmentation," *International Journal of Innovative Computing, Information and Control*, vol. 13, no. 5, pp. 1509–1526, Oct. 2017.

[13] P. Panwar, G. Gopal, and R. Kumar, "Image Segmentation using K-means clustering and Thresholding," *International Research Journal of Engineering and Technology*, vol. 3, no. 5, pp. 1787–1793, May 2016.

[14] I. M. Kazi, S. S. Chowhan, and U. V. Kulkarni, "MRI Brain Image segmentation using Adaptive Thresholding and K-means Algorithm," *International Journal of Computer Applications*, vol. 167, no. 8, pp. 11–15, Jun. 2017, https://doi.org/10.5120/ijca2017914330.

[15] L.-H. Juang and M.-N. Wu, "MRI brain lesion image detection based on color-converted K-means clustering segmentation," *Measurement*, vol. 43, no. 7, pp. 941–949, Aug. 2010, https://doi.org/10.1016/j.measurement.2010.03.013.

[16] L. Zeinalkhani, A. A. Jamaat, and K. Rostami, "Diagnosis of Brain Tumor Using Combination of K-Means Clustering and Genetic Algorithm," *Frontiers in Health Informatics*, vol. 7, Nov. 2018, Art. no. 6.

[17] I. Mehidi, D. E. C. Belkhiat, and D. Jabri, "A Fast K-means Clustering Algorithm for Separation of Brain Tissues in MRI," in *2020 2nd International Conference on Mathematics and Information Technology (ICMIT)*, Adrar, Algeria, Oct. 2020, pp. 132–137, https://doi.org/10.1109/ICMIT47780.2020.9046971.

[18] S. Madhukumar and N. Santhiyakumari, "Evaluation of k-Means and fuzzy C-means segmentation on MR images of brain," *The Egyptian Journal of Radiology and Nuclear Medicine*, vol. 46, no. 2, pp. 475–479, Jun. 2015, https://doi.org/10.1016/j.ejrnm.2015.02.008.

[19] R. Ghezelbash, A. Maghsoudi, and E. J. M. Carranza, "Optimization of geochemical anomaly detection using a novel genetic *K*-means clustering (GKMC) algorithm," *Computers & Geosciences*, vol. 134, Jan. 2020, Art. no. 104335, https://doi.org/10.1016/j.cageo.2019.104335.

[20] G. R. Nitta, T. Sravani, S. Nitta, and B. Muthu, "Dominant gray level based K-means algorithm for MRI images," *Health and Technology*, vol. 10, no. 1, pp. 281–287, Jan. 2020, https://doi.org/10.1007/s12553-018-00293-1.

[21] A. Mano, "A Novel Approach based on PSO Optimized K-Means in MRI Brain Image Segmentation." techrxiv, Jul. 02, 2020, https://doi.org/10.36227/techrxiv.12310100.v2.

[22] "BrainWeb: Simulated Brain Database." https://brainweb.bic.mni.mcgill.ca/brainweb/.