Q2.
Journal :
Automatic Human Brain Tumor Detection in MRI Image Using Template-Based K Means and Improved Fuzzy C Means Clustering Algorithm

Keywords: Magnetic Resonance Imaging; K-means clustering; Fuzzy C-means clustering; Template based K-means and Modified Fuzzy C-means (TKFCM); gray level intensity, Digital Image Processing, True Negative, True Positive, False Negative, False Positive.

As we know, Brain is the kernel of our body. Brain Tumor can grow in our brain and go unnoticed for long periods. Therefore, detecting it at the first try, or right stages is extremely important and is now one of the challenging issues in medical science. Tumors can be malignant or benign. Even though Benign tumors are non-cancerous, it is equally important to detect the kind of tumors at early stages. The journal has proposed a new model that uses the algorithm TKFCM – Template based K-means and improved Fuzzy C-means algorithm, for detecting the tumors from images taken from MRI (Magnetic Resonance Imaging). This algorithm uses K-means Algorithm and Fuzzy C-means algorithm while using features of the image as the data to detecting the tumors with lower error rate and higher time efficiency.

The dataset for this model is images produced from Digital Image Processing (DIP), which has been one of the promising techniques in biological science today. DIP has been used for classification, tumor detection, cancer detection, testing and examining body for a while now/ The current models used are Computer-aided Systems (CAD), Magnetic Resonance Image segmentation of medical images. These are effective but very time taking. The US statistics as per the journal says there are 34 deaths due to brain tumors every day, thus such time taking segmentation processes can not be relied on fully. The proposed algorithm claims to take just a few seconds while giving high accuracy in detection of tumors.

## Dataset

For this model, the dataset is the MRI images of the Brain. The features of the images are the attributes for each data. There are 6 features extracted for the classifiers from the images :

1. Energy: The rate of change in the brightness or color or magnitude of the pixels over local areas of image is known as the Energy. For G = gray level co-occurrence matrix and $p(x_i,y_j)$ is the image matrix,

$$ENG = \sum_{i=1}^{G} \sum_{j=1}^{G} |P(x_i, y_j)|^2$$

2. Contrast: The measurement of the intensity of color or grayscale differentiation among the pixel and its adjacent neighbors.

$$CON = \sum_{n=1}^{G} n^2 \sum_{i=1}^{G} \sum_{j=1}^{G} |P(x_i, y_j)| \quad and \quad |i-j| = n$$

3. Homogeneity: As the name suggests, the measurement of similarity in an image.

$$HOM = \frac{\sum_{i=1}^{G} \sum_{j=1}^{G} |P(x_i, y_j)|}{1 + |i+j|}$$

4. Entropy: The randomness in the texture of the image

$$ENT = \sum_{i=1}^{G} \sum_{j=1}^{G} |P(x_i, y_j)| \frac{1}{\log |P(x_i, y_j)|}$$

5. Dissimilarity: Dissimilarity in the texture of the image. Calculated by taking in account the alignment of the image with metric as angle

$$DSM = \sum_{i=1}^{G} \sum_{j=1}^{G} |P(x_i, y_j)| \quad |i-j|$$

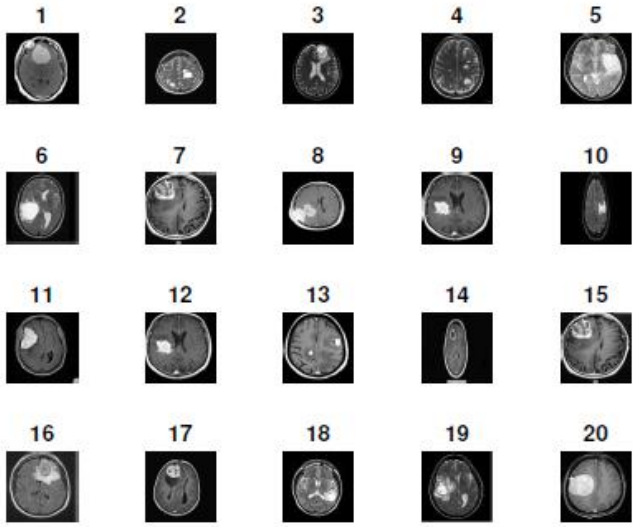
6. Correlation: The spatial dependencies between the pixels

$$COR = \frac{\sum_{i=1}^{G} \sum_{j=1}^{G} (x_i, y_i) P(x_i, y_j) - M_x M_y}{\sigma_x \sigma_y}$$

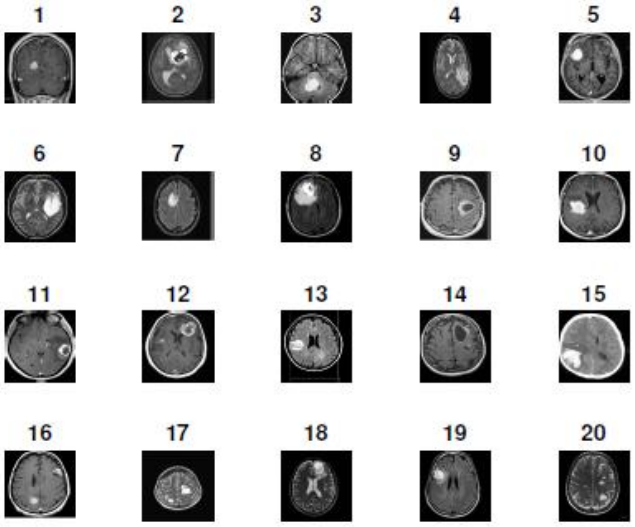
$M_x$ and $\sigma_x$ = mean and standard deviation in horizontal spatial domain ; $M_y$ and $\sigma_y$ = mean and standard deviation in vertical spatial domain

All these are the features/attributes of the dataset. By using the characteristics of the image, it is easier for analysing since the features like gray level intensity and texture will vary with the clusters in the brain, for example the areas with tumors will vary from the rest of the areas in the image.

For example, here is a dataset from the journal of 40 brain tumor images. The journal tests their proposed model on the Data 10 of the database 2. It is interesting to see how the model is actually able to determine the tumor regions with low error rate.




(a) Database 1




(b) Database 2

## Algorithm

### K-means Algorithm

This is a clustering algorithm, that divides the dataset into a number of clusters and then performs using distance as an important attribute. Distance can be calculated using different methods, but for this case, the Euclidean Distance is used since it is fast, robust and easier to understand.

- ⇨ Each pixel is a data point and the set of these data points X = $x_1$, $x_2$,…$x_n$, n= number of data items
- ⇨ V = $v_1$, $v_2$,…$v_c$ is the set of centers
- ⇨ Number of clusters = K
- ⇨ Cluster centers are defined randomly = c
- ⇨ The Euclidean distance between each data point and each cluster center is calculated. Then the particular data point is assigned to the cluster with the minimum value of Euclidean Distance. For instance, for data point $x_1$, the distance from $c_2$ is the minimum as compared to the other centers, therefore, $x_1$ is then assigned to Cluster 2 and the new cluster center is now recalculated as :

$$V_i = \frac{1}{c_i} \sum_{j=1}^{c_i} x_i$$

  $c_i$ = number of data points in the i-th cluster
- ⇨ The distance is recalculated for every data point and new cluster centers. In programming/algorithm language, the loop carries on until no data point is reassigned.

Therefore the distance is measured between each pixel/data point to each cluster center individually and the shortest distance is the cluster the point is assigned to. After reassigning, the recalculation is done which confirms that all the pixels have been compared again to all centroids. The pros of this algorithm is it ensures computational efficiency and can support multi dimensional vectors.

### Fuzzy C-means Algorithm

FCM is another clustering algorithm that is similar to K-means Algorithm, the difference is that in FCM a single data point can belong to more than one cluster. It's priority towards the i-th cluster is called the *membership* which in our case is the distance of the data point from the cluster center. The closer it is, the higher the membership with that cluster.

$$J_m(U,V) = \sum_{j=1}^{N} \sum_{i=1}^{C} (u_{ij})^m (||x_j - v_i||)^2; \quad 1 \leq m \leq \infty$$

- ⇨ Fuzzy – degree of truth – in this case we could degree of closeness (distance)
- ⇨ Each pixel is a data point and the set of these data points X = $x_1$, $x_2$,…$x_n$, n= number of data items
- ⇨ V = $v_1$, $v_2$,…$v_c$ is the set of centers
- ⇨ The number of clusters are fixed = c, where 2<= c < = n.
- ⇨ m is the hyper parameter that controls how fuzzy the cluster is. The higher the degree of fuzziness for a cluster, the more there is similarity between the data points in the cluster.

1<m<∞. m is greater than 1, because as we see the formula for FCM algorithm uses m in the denominator as (m-1).

- ⇨ Initialize the fuzzy c partition *U*
- ⇨ Fuzzy membership function $U_{ij}$ is calculated using the equation

$$U_{ij} = \sum_{k=1}^{c} \left( \frac{(||x_j - v_i||)}{(||x_j - v_k||)} \right)^{\left(\frac{-2}{m-1}\right)}$$

- ⇨ Fuzzy centers $V_i$ are calculated

$$V_i = \frac{\sum_{j=1}^{N} (U_{ij})^m \times J}{\sum_{j=1}^{N} U_{ij}{}^m}$$

- ⇨ The steps are repeated until the minimum 'j' value is achieved

## TKFCM – Template Based K-means and Modified Fuzzy C-means

This proposed algorithm uses both K-means and FCM, while using the temper of the image to tis favor – the 6 attributes of data (image) talked above. The temper gray level intensity is used with the K-means algorithm for segmentation/clustering. FCM, Euclidean distance and membership function is modified based on the image features/attributes and used for further segmentation.

The formula for TKFCM is :

$$J = \sum_{i=i+1}^{M} \sum_{j=j+1}^{N} B(x_i, y_j) \times \sum_{i=1}^{K} \sum_{j=1}^{C} P_{ij} ||x_i - c_j||^2 \times \sum_{j=1}^{R} \sum_{i=1}^{C} (U_{ij})^m d^2(x_j, v_i)$$

$P_{ij}$ = binary image matrix; M & N are the rows and columns respectively
R = centroid of cluster
K = number of data points in clusters
C = number of clusters

The last part of formula is the FCM using Euclidean distance and image features.

The middle part is the K-means algorithm.

The first part is 'coarse image', the part of equation taking into account the temper calculation of image.

$$B(x_i, y_j) = \sum_{i=i+1}^{M} \sum_{j=j+1}^{N} P(x_i, y_j) \times T_{mn}$$

$$T_{mn} = \sum_{i=1}^{M} \sum_{j=1}^{N} P(x_i, y_j) * \sum_{k=1}^{G} \sum_{l=1}^{S} P(x_k, y_l); \quad k \epsilon M, l \epsilon N$$

$T_{mn}$ = temper-based image matrix
G = number of gray level intensity

S = number of bins
$P(x_i, y_j)$ = temper of image

## Entire Process

The final process looks like this :

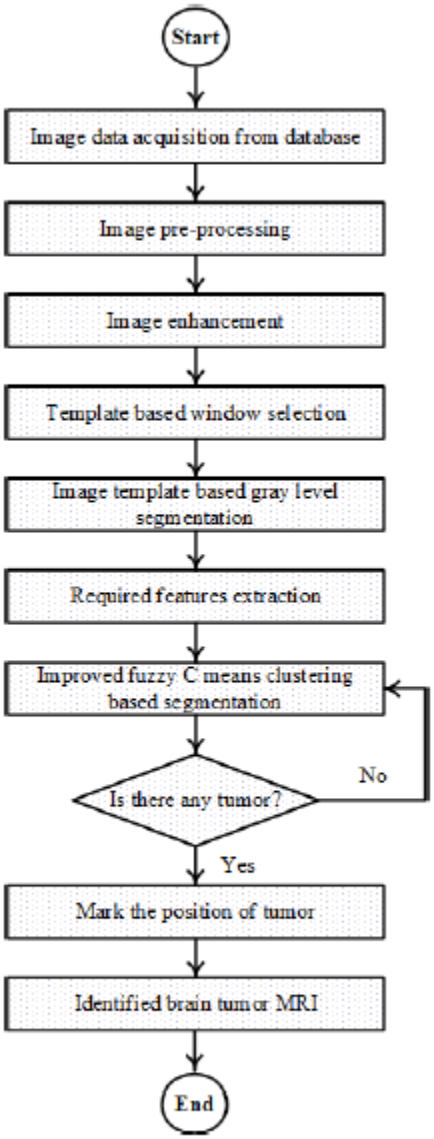



**Figure 1.** Flowchart of proposed TKFCM algorithm.

First the MRI is done, the image taken is first pre-processed and enhanced. The gray levels and image matrix is defined. The template window gets selected, and the output of the window gets segmented using the K-means Algorithm. Once that is done, the required features are extracted and finally tumor detection is done by the FCM with the updates membership – image features.

The data has high accuracy and low error rate, that means there is still some error. This error rate is measured by the values of the True Negative (TN), True Positive (TP), False Negative (FN), False Positive (FP).

TP = test result is positive and there is existing abnormality = detection done correctly
TN = test result is negative and there is no existing abnormality = detection done correctly
FP = test result is positive and there is no existing abnormality = detection not done correctly
FN = test result is negative and there is existing abnormality = detection not done correctly

The sensitivity for the model is defined as the measurement of correctly identification of presence of tumor

$$\alpha = \frac{TP}{TP + TN} \times 100$$

The specificity for the model is defined as the measurement of correctly identification of absence of tumor

$$\beta = \frac{TN}{TN + FP} \times 100$$

The accuracy for the model is defined as the measurement of actual classification.

$$\eta = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

AMA Style

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):27. https://doi.org/10.3390/bdcc3020027

KarlKarl 5, TatarizeTatarize 9, leonbloyleonbloy 66.6k2020 gold badges130130 silver badges180180 bronze badges, et al. What is "energy" in image processing? Stack Overflow. https://stackoverflow.com/questions/4562801/what-is-energy-in-image-processing#:~:text=The%20energy%20is%20a%20measure,the%20pixels%20over%20local%20areas. Published April 1, 1959. Accessed October 7, 2021.

Brightness and Contrast in Digital Images. (n.d.). Retrieved from https://www.olympus-lifescience.com/en/microscope-resource/primer/java/olympusmicd/digitalimaging/contrast/#:~:text=The term contrast refers to,than those of lower contrast