



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.488**

 9940 572 462

 6381 907 438

 [ijircce@gmail.com](mailto:ijircce@gmail.com)

 [www.ijircce.com](http://www.ijircce.com)

# Hybrid Clustering Technique Based Segmentation of Brain Tumor

**K.R.Barani, S.Kavitha, Msc.,M.Phil.,Ph.d.,**

Student, Dept. of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram, TamilNadu, India

Assistant Professor, Dept. of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram, TamilNadu, India

**ABSTRACT:** Inference of tumor and edema areas from brain magnetic resonance imaging (MRI) data remains challenging owing to the complex structure of brain tumors, blurred boundaries, and external factors such as noise. To alleviate noise sensitivity and improve the stability of segmentation, an effective hybrid clustering algorithm combined with morphological operations is proposed for segmenting brain tumors in this paper. The main contributions of the paper are as follows: firstly, adaptive Wiener filtering is utilized for denoising, and morphological operations are used for removing nonbrain tissue, effectively reducing the method's sensitivity to noise. Secondly, K-means++ clustering is combined with the Gaussian kernel-based fuzzy C-means algorithm to segment images. This clustering not only improves the algorithm's stability, but also reduces the sensitivity of clustering parameters. Finally, the extracted tumor images are postprocessed using morphological operations and median filtering to obtain accurate representations of brain tumors. In addition, the proposed algorithm was compared with other current segmentation algorithms. The results show that the proposed algorithm performs better in terms of accuracy, sensitivity, specificity, and recall.

## I. INTRODUCTION

Magnetic Resonance Images (MRI) acquisition provides different sequences like T1- Weighted Images (T1WI), T2-Weighted (T2WI), Proton Density Images (PDI), FluidAttenuated Inversion Recovery (FLAIR) etc. Massive information on tissue structure and pathology can be extracted from these sequences, but each sequence differs in available information content [1]. Slice by slice examination and extraction of small details and abnormalities from these large numbers of sequences is a tedious job in clinical applications. Multispectral data analysis combines the slices of the same brain portion from each sequence to form a single suite so that it helps to analyze the corresponding pixel information as a pixel signature [1]. For example, see the sample slices of T1WI, T2WI and Diffusion Weighted Image (DWI) shown in Figure 1. It is observed that details present in an image vary from slice to slice. In Figure 1, T1WI shows White Matter (WM) information clearly, whereas T2WI contains Gray Matter (GM) and Cerebro Spinal Fluid (CSF) information. DWI fails to distinguish the brain tissues, but pathological information is clearly visible in it.

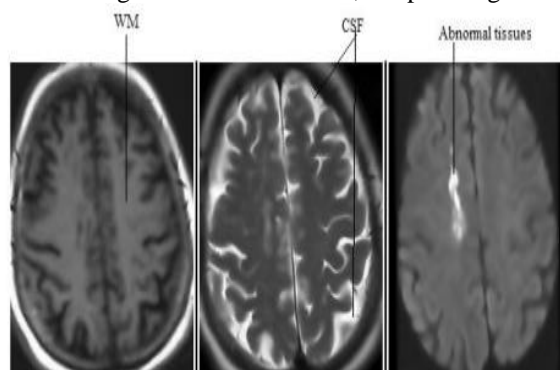


Figure 1. Input Slices of T1WI, T2WI and DWI (from left to right)

Researchers in MRI analysis have been intensively working for last few decades to improve the performance of existing data mining techniques using multispectral approaches. But it remains as a challenge because classification accuracy highly depends upon the input data characteristics and feature analysis methods. Pre-processing, feature extraction and classification are the main steps involved in a typical multispectral analysis system. Preprocessing techniques like image registration, denoising and contrast improvement can contribute much to select the best features

for further analysis. Classification methods in multispectral analysis can be effectively classified into two categories, unsupervised and supervised learning. Unsupervised methods like k-means, Fuzzy C-Means (FCM) and Expectation Maximization (EM) can give satisfactory results for MR image analysis. But radiologists often rely on feedback from previous data and diagnosis to reach at a correct opinion for each case. Supervised learning techniques follow similar strategy, and widely used in computer aided categorization of MRI data. Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are the two widely accepted techniques in supervised MRI classification.

The usual examination workflow is performed by expert neurologists or radiologists, who are able to figure out complex anatomical patterns and subtle changes with clinical meaning. The process that an expert follows when examining a particular case involves two different kinds of tasks: those related with image perception, such as visual search or exploration paths, and others associated with cognitive skills, mainly related to diagnostic reasoning and decision making. An expert structures a diagnosis by using contextual knowledge and fusing information from different sources, a process that has been recently under study. At analyzing structural brain MR images, a main aim is to find anatomical changes, either local or global, related to functional disturbances. In particular, radiologists examine images by looking at distinctively regions and compare them by searching difference. In the computational attempt of emulating the human vision process—a synchronized collaborative work between the brain and low level visual mechanisms—the concept of visual attention has introduced a generation of techniques that are able to transform an image into a hierarchy of relevant regions, known as salient regions. Relevant regions in radiological terms may be defined as those image areas that are visually altered and are entailed with a certain degree of clinical interpretability. Nevertheless, most methods used to compare brains establish local rather than regional (salient) differences. Currently, a morphometric brainanalysis consists of a set of strategies aimed to extract and quantify anatomical differences between groups of subjects. Commonly, this analysis comprises two main processes: first, all images are warped or registered together to a common referenceframe or template, and second, a quantification of the estimated local deformation required to register is computed, producing specific measurements of interest. voxel-based morphometry (VBM) and deformation-based morphometry (DBM) are currently the most used techniques to compare populations. In VBM, local differences, found in brain tissue segmentations, are voxel-by-voxel statistically analyzed, while DBM statistically compares information coming from the deformations fields obtained after registration to the template. With these methods, one-to-one correspondences between subjects are assumed and statistics are computed for the same voxel across all subjects. However, conclusions are limited when the same structure may be partially present, or when a single anatomical region may exhibit multiple shapes across the population. On the other hand, some pathologies may affect not only unique anatomical structure A detailed review of some supervised and unsupervised classification methods in MS lesion segmentation is described.

## II. LITERATURE SURVAY

Multispectral approach to brain MRI analysis has shown great advance recently in pathology and tissue analysis. However, poor performance of the feature extraction and classification techniques involved in it discourages radiologists to use it in clinical applications. Transform based feature extraction methods like Independent Component Analysis (ICA) and its variants have contributed a lot in this research field. But these global transforms often fails in extraction of local features like small lesions from clinical cases and noisy data. Feature extraction part of the recently introduced Multiresolution Independent Component Analysis (MICA) algorithm in microarray classification is proposed in this work to resolve this issue. Effectiveness of the algorithm in MRI analysis is demonstrated by training and classification with Support Vector Machines (SVM). Both synthetic and real abnormal data from T1-weighted, T2-weighted, proton density, fluid-attenuated inversion recovery and diffusion weighted MRI sequences are considered for detailed evaluation of the method. Tanimoto index, sensitivity, specificity and accuracy of the classified results are measured and analyzed for brain abnormalities, affected white matter and gray matter tissues in all cases including noisy environment. A detailed comparative study of classification using MICA and ICA is also carried out to confirm the positive effect of the proposed method. MICA based SVM is found to yield very good results in anomaly detection, around 2.5 times improvement in classification accuracy is observed for abnormal data analysis.

- the accuracy of the automated method compared to manual segmentations performed by two cardiologists;
- the ability of the method to compute reliable characteristics of the LV (ejection fraction and left ventricular mass);
- the temporal continuity of the resulting automated segmentation;
- the time-efficiency (about 3' to segment a sequence of 25 3D-images on a low-end computer) of the proposed method; and
- The robustness of the few parameters whose setting rely mostly on physical and anatomical facts.



### III. SYSTEM ANALYSIS

#### 3.1 PROBLEM IDENTIFICATION

- In the scheme does not permit the direct derivation of deformation parameters.
- Detection accuracy is low
- This particularly complex segmentation task, prior knowledge is required.
- Major challenges linked to this segmentation task.

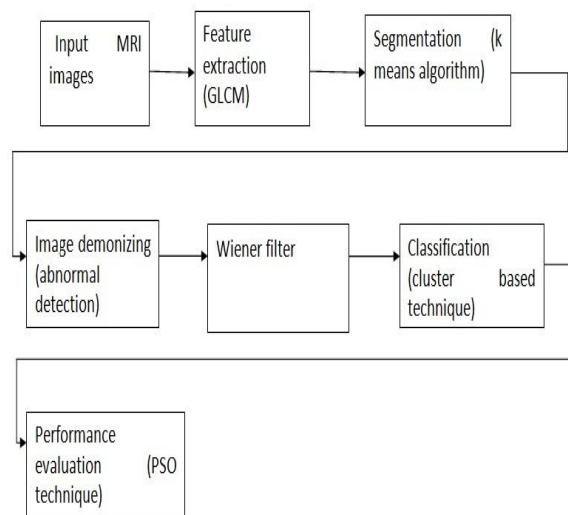
#### 3.2 EXISTING SYSTEM:

In existing studies suggest neuro imaging may become a valuable tool in the early diagnosis of neurodegenerative diseases by extracting anatomical patterns and revealing hidden relations from structural magnetic resonance (MR) images. At analyzing structural brain MR images, a main aim is to find anatomical changes, either local or global, related to functional disturbances. In particular, radiologists examine images by looking at distinctively regions and compare them by searching differences. The most popular technique has been by far the support vector machine (SVM), which has been applied to classifying individuals with several neurological disorders. In this Methods that face the problem of high-dimensionality are the ones that perform image synthesis

#### 3.3 PROPOSED METHOD:

The proposed method is based on a two-phase PSO model that combines bottom-up and top-down approaches to achieve accurate classification of brain MR images as normal controls or probable AD subjects. K-means with PSO based on clustering which is used to classify the problem. In the proposed approach, the pre-defined kernels convert the input image into individual feature saliency maps, whose pso correspond to dimensions of the kmeans space. Most brain lesion segmentation methods based on outlier detection, the proposed method is generic. It does not consider single voxels independently and makes no assumption about shape or intensity profile of the abnormality.

#### 3.4 ARCHITECTURE DIAGRAM

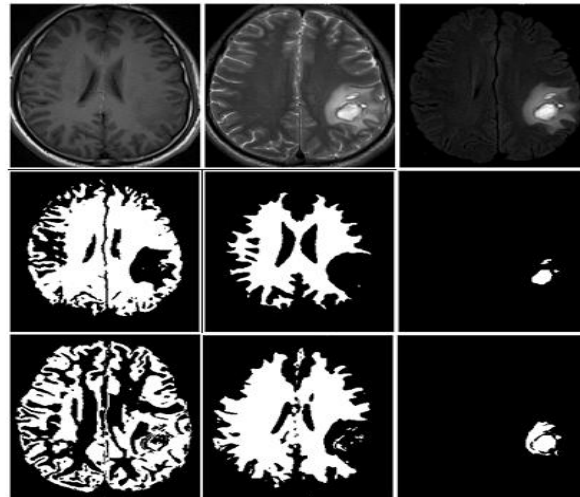


### IV. SYSTEM IMPLEMENTATION

#### 4.1 IMAGE SELECTION

MRI sequences for clinical trials are usually acquired with different size and orientation. Registration of the images to form a co-registered multispectral suite is the preliminary step in analysis process. Each pixel vector in a multispectral image forms the spectral signature corresponding to that pixel and a collection of these spectral signatures generate multi signals.

Clinical Image Analysis



Two sets of clinical data are used in visual and quantitative analysis. The first dataset contains T1WI, T2WI and FLAIR images with specifications as described in total 70 multispectral slice sets were selected for analysis from 20 abnormal cases. Top row shows slices in a sample multispectral image set. T2WI and FLAIR images show lesion surrounded by edema, but that information is not visible in T1WI. Classified results from ICA+SVM and MICA+SVM are given in Figure 5 middle row and last row respectively with GM in the 1<sup>st</sup> column, WM in the 2<sup>nd</sup> column and abnormality in the 3<sup>rd</sup> column.

MICA+SVM results show the lesion and the surrounding edema (Figure last row last column) with a clear description of the separation between lesion and edema in the original image. On observing the affected portion of WM (Figure 2<sup>nd</sup> column), MICA+SVM results looks better than ICA+SVM results. However, MICA cannot reach the performance of ICA in classification of WM.

**4.2 IMAGE DENOISING**

Images will often get corrupted in transmission and acquisition due to noise. Therefore we need image restoration to remove the additive noise (mostly Additive White Gaussian Noise) from the image while keeping its original features as much as we can. In here we discuss a Method of Image Denoising using Wiener Filtering and Adaptive Median Filtering. Image Denoising has become a very essential process in Image Restoration. Today several techniques exist such as Wiener Filtering, Gaussian scalar mixture, BM3D (Image denoising, 2010) which is used for image denoising. They have been successfully used in areas such as medical imaging and astronomy.



The goal of the Wiener filter is to filter out noise that has corrupted a signal. It is based on a statistical approach. Typical filters are designed for a desired frequency response. However, the design of the Wiener filter takes a different approach. One is assumed to have knowledge of the spectral properties of the original signal and the noise,

and one seeks the LTI filter whose output would come as close to the original signal as possible. Wiener filters are characterized by the following:

1. Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation
2. Requirement: the filter must be physically realizable, i.e. causal (this requirement can be dropped, resulting in a non-causal solution)
3. Performance criterion: minimum mean-square error (MMSE) This filter is frequently used in the process of deconvolution; for this application, see Wiener deconvolution.

#### 4.3 KMEANS AND PSO ALGORITHM

The k-means clustering algorithm finds the desired number of distinct clusters and their centroids. A centroid is defined as the point whose coordinates are obtained by computing the average of each of the coordinates (i.e., feature values) of the points of the jobs assigned to the cluster [2]. Formally, the k-means clustering algorithm follows the following steps.

1. Choose a number of desired clusters,  $k$ .
2. Choose  $k$  starting points to be used as initial estimates of the cluster centroids. These are the initial starting values.
3. Examine each point (i.e., job) in the workload data set and assign it to the cluster whose centroid is nearest to it.
4. When each point is assigned to a cluster, recalculate the new  $k$  centroids.
5. Repeat steps 3 and 4 until no point changes its cluster assignment, or until a maximum number of passes through the data set is performed.

Before the clustering algorithm can be applied, actual data samples (i.e., jobs) are collected from observed workloads. The features that describe each data sample in the workload are required *a priori*. The values of these features make up a feature vector  $(F_{i1}, F_{i2}, \dots, F_{iM})$ , where  $F_{im}$  is the value of the  $m^{\text{th}}$  feature of the  $i^{\text{th}}$  job. Each job is described by its  $M$  features. For example, if job 1 requires 3MB of storage and 20 seconds of CPU time, then  $(F_{11}, F_{12}) = (3, 20)$ . The feature vector can be thought of as a point in  $M$ -dimensional space. Like other clustering algorithms, k-means requires that a distance metric between points be defined [2]. This distance metric is used in step 3 of the algorithm given above. A common distance metric is the Euclidean distance. Given two sample points,  $p_i$  and  $p_j$ , each described by their feature vectors,  $p_i = (F_{i1}, F_{i2}, \dots, F_{iM})$  and  $p_j = (F_{j1}, F_{j2}, \dots, F_{jM})$ , the distance,  $d_{ij}$ , between  $p_i$  and  $p_j$  is given by:

$$d_{ij} = \sqrt{\sum_{m=1}^M (F_{im} - F_{jm})^2} \quad (1)$$

If the different features being used in the feature vector have different relative values and ranges, the distance computation may be distorted since features with large absolute values tend to dominate the computation [2]. To mitigate this, it is common for the feature values to be first scaled in order to minimize distortion. There are several different methods that can be used to scale data. The method used in this paper is z-score scaling. Z-score scaling uses the number of standard deviations away from the mean that the data point resides [5]. The z-score equation is

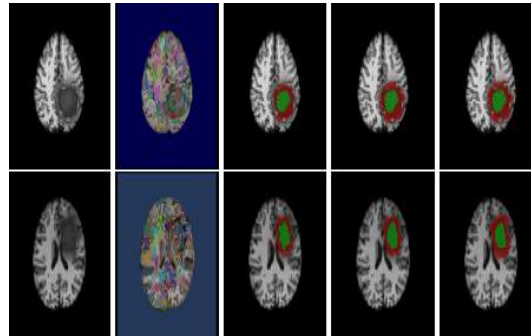
$$F_{im}^* = \frac{F_{im} - \mu_m}{\sigma_m} \quad (2)$$

where  $F_{im}$  is the value of the  $m^{\text{th}}$  feature of the  $i^{\text{th}}$  job (i.e., the data point),  $\mu_m$  is the mean value of the  $m^{\text{th}}$  feature, and  $\sigma_m$  is the standard deviation of the  $m^{\text{th}}$  feature. Thus, before the algorithm is applied, the original data set is scaled, using the z-score scaling technique, where the feature mean is subtracted from the feature value and then divided by the standard deviation of that feature (i.e.,  $F_{im}$  is replaced by its scaled value  $F_{im}^*$ ). This technique has the effect of normalizing the workload features so that no single feature dominates in the clustering algorithm.

The number of clusters to be found, along with the initial starting point values are specified as input parameters to the clustering algorithm. Given the initial starting values, the distance from each (z-scored scaled) sample data point to each initial starting value is found using equation (1). Each data point is then placed in the cluster associated with the nearest starting point. New cluster centroids are calculated after all data points have been assigned to a cluster. Suppose that  $C_{im}$  represents the centroid of the  $m^{\text{th}}$  feature of the  $i^{\text{th}}$  cluster. Then,

$$C_{im} = \frac{\sum_{j=1}^{n_i} F_{i,jm}^*}{n_i} \quad (3)$$

where  $F_{i,jm}^*$  is the  $m^{\text{th}}$  (scaled) feature value of the  $j^{\text{th}}$  job assigned to the  $i^{\text{th}}$  cluster and where  $n_i$  is the number of data points in cluster  $i$ . The new centroid value is calculated for each feature in each cluster. These new cluster centroids are then treated as the new initial starting values and steps 3-4 of the algorithm are repeated. This continues until no data point changes clusters or until a maximum number of passes through the data set is performed.



#### 4.5 PARTICLE SWARM OPTIMIZATION (PSO)

As stated before, PSO simulates the behaviors of bird flocking. Suppose the following scenario: a group of birds are randomly searching food in an area. There is only one piece of food in the area being searched. All the birds do not know where the food is. So what's the best strategy to find the food? The effective one is to follow the bird which is nearest to the food. PSO learned from the scenario and used it to solve the optimization problems. In PSO, each single solution is a "bird" in the search space. We call it "particle". All of particles have fitness values which are evaluated by the fitness function to be optimized, and have velocities which direct the flying of the particles. The particles fly through the problem space by following the current optimum particles.

#### 4.6 IMAGE FEATURE EXTRACTION AND IMAGE FEATURE SELECTION

The goal of feature extraction is to locate points in the scene that lie along boundaries. By boundaries we mean sets of pixels that either separate objects from one another or represent changes in the surface geometry of an object. This work considers two basic types of boundaries; step edges and crease edges. Step edges are those edges that represent depth discontinuities, and crease edges are those that represent a crease, or a discontinuity in the surface normal, in a single object. From the range data, steps edges can be identified by the gradient magnitude and crease edges can be located by computing the gradient magnitude of the 3D surface normal. From the amplitude data, object boundaries coincide with high gradient magnitude when objects in color or texture. Crease edges can be located from the gradient magnitude of the amplitude image because of differing angles of incidence of the laser beam. A short overview of our feature extraction methodology is as follows. We begin with some preprocessing to reduce the effects of noise. Given a registered pair of range and amplitude images of a scene from a single pose, we median filter the data to reduce replacement noise (outliers or misfires from the laser) and implement an edge preserving, spatially variant smoothing operator on both the range and amplitude data to reduce additive noise.

#### 4.7 CLASSIFICATION

In the previous section we have shown that the DLD problem can be interpreted as a binary classification problem in which one conditional class probability is known. We now show that this interpretation has far reaching algorithmic consequences. To this end let us assume that we give each sample of our training set  $T = (x_1, \dots, x_n)$  drawn from  $Q$  the label 1. Additionally we generate a second training set  $T_0 = (x_0^1, \dots, x_0^n)$  from  $\mu$  and label each sample of it with  $-1$ . Merging these labeled sample sets gives a new training set which then can be used by a binary classification algorithm.

### V. CONCLUSION AND FUTURE ENHANCEMENT

#### 5.1 CONCLUSION

In recent years, multispectral approach helped MRI analysts a lot to improve the analysis time and accuracy of the clinical trials. However, extraction of very critical features like small lesions is a great challenge in pathology analysis due to lack of efficiency of the existing methods. The proposed multi resolution analysis coupled with ICA is demonstrated as a good choice to resolve this issue. SVM classification is used to investigate and evaluate the performance of the method in abnormality analysis. Experimental results using synthetic and clinical data confirm that

the proposed method performs better than ICA based classifications in lesion/tumor detection. Experiments conducted for noisy synthetic images also support these findings with acceptable results in favor of MICA. Refinements of MICA are under consideration to give equal priority to normal and abnormal tissue classification in future works.

## 5.2 FUTURE ENHANCEMENT

In this project we detect the affected part on a MR image. we proposed the block-adaptive windows, by which the quality of estimations of image energy distributions is significantly improved. The doubly local Wiener filtering method with block-adaptive windows obtains better denoising performance. And to find the connection between the probabilistic label fusion model and the recently proposed kmeans & pso segmentation method. Another contribution is that label information is incorporated into image registration to improve registration accuracy. Experimental results show that registration refinement improves segmentation accuracy. The method produces reliable clinical indexes which are in good agreement with the manual measurements. It can provide useful information for clinicians in cancer disease diagnosis.

## REFERENCES

- [1] M. Jolly, "Fully automatic left ventricle segmentation in cardiac Cine MR Images using registration and minimum surfaces," MIDAS.
- [2] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, And D. Rueckert, "Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm,"
- [3] M. Lynch, O. Ghita, and P. F. Whelan, "Segmentation of the left ventricle of the heart in 3- MRI data using an optimized Nonrigid Temporal model,"
- [4] C. Petitjean and J. N. Dacher, "A review of segmentation methods in short axis cardiac MR images,"
- [5] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. New York: Springer, 2010.
- [6] N. Doan, B. van Lew, B. Lelieveldt, M. van Buchem, J. Reiber, and J. Milles, "Deformation texture-based features for classification in Alzheimer's disease," SPIE Med. Imag., 2013.





**INNO SPACE**  
SJIF Scientific Journal Impact Factor

Impact Factor:  
7.488

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  [ijircce@gmail.com](mailto:ijircce@gmail.com)



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details