IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SEGMENTATION OF A BRAIN TUMOR USING A HYBRID CLUSTERING TECHNIQUE

[1]Mrs. M. Angel Shalini, [2]Mr.E. Vishnu Prasath,[3]Mr. S. Dineshkumar

[1]Assistant Professor, [2]Scholar, [3]Scholar

[1]PG And Research Department of Computer Applications,

[1,2,3]Hindusthan College of Arts And Science, Coimbatore, Tamilnadu, India.

ABSTRACT

Because of the complicated structure of brain tumours, indistinct boundaries, and external influences such as noise, inferring tumour and edoema areas from brain magnetic resonance imaging (MRI) data remains difficult. This work proposes an effective hybrid clustering technique combined with morphological procedures for segmenting brain tumours to reduce noise sensitivity and increase segmentation stability. The following are the paper's main contributions: To begin, adaptive Wiener filtering is employed for denoising, then morphological processes are used to remove non-brain tissue, effectively lowering the method's noise sensitivity. Second, to segment images, K-means++ clustering is merged with the Gaussian kernel-based fuzzy C-means method. This clustering enhances the algorithm's stability while simultaneously lowering the sensitivity of the clustering settings. Finally, to acquire realistic representations of brain tumours, the retrieved tumour images are post processed utilising morphological procedures and median filtering. The suggested technique was also compared to other existing segmentation algorithms. The suggested algorithm outperforms the competition in terms of accuracy, sensitivity, specificity, and recall.

Keywords – MRI, Kmeans, cluster, segmentation, morphological

I. INTRODUCTION

T1-Weighted Images (T1WI), T2-Weighted Images (T2WI), Proton Density Images (PDI), FluidAttenuated Inversion Recovery (FLAIR), and more sequences are available for MRI capture. These sequences can yield a lot of information about tissue shape and disease, but the amount of information available varies [1]. In clinical applications, slice by slice analysis and extraction of minute features and anomalies from huge numbers of sequences is a time-consuming task. Multispectral data analysis combines slices of the same brain area from each sequence into a single suite, allowing the pixel information to be analysed 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.

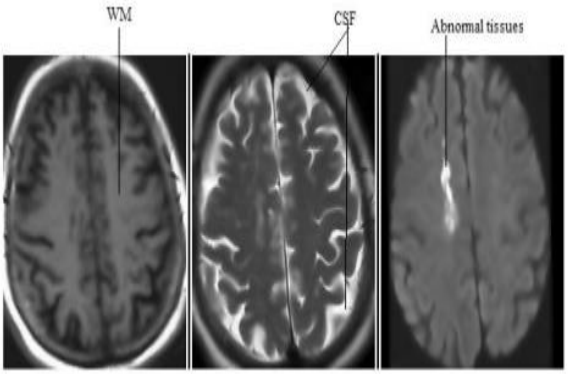


Figure 1. Input Slices of T1WI, T2WI and DWI (from left to right)

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. Expert neurologists or radiologists undertake the standard examination workflow, which includes identifying complicated anatomical patterns and minor alterations with clinical significance. When an expert examines a case, they perform two types of tasks: those linked to picture perception, such as visual search or exploratory paths, and those connected to cognitive skills, such as diagnostic reasoning and decision making. An expert constructs a diagnosis by combining information from several sources and employing contextual knowledge, a process that has recently been studied.

One of the main goals of studying structural brain MR images is to look for anatomical alterations, either local or global, that are linked to functional disorders. 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 deforma-tion-based morphometry (DBM) are currently the most used techniques to compare populations. Local differences in brain tissue segmentations are statistically assessed voxel-by-voxel in VBM, whereas DBM statistically compares information from deformation fields produced after registration to the template.

One-to-one correspondences between subjects are assumed in these methods, and statistics for the same voxel are computed across all subjects. When the same structure is partially present, or when a single anatomical location exhibits multiple shapes across the population, conclusions are constrained. However, some diseases can damage more than just one anatomical structure. In MS lesion segmentation, a full overview of some supervised and unsupervised classification approaches is presented.

II. LITERATURE SURVAY

Pathology and tissue analysis have lately made significant progress using a multispectral approach to brain MRI analysis. However, radiologists are hesitant to employ it in clinical settings due to the poor performance of the feature extraction and classification algorithms involved. Methods based on transforms, like as Independent Component Analysis (ICA) and its derivatives, have made significant contributions to this field of study. However, in clinical cases and noisy data, these global transforms frequently fail to retrieve local features such as tiny lesions. This paper proposes that the feature extraction element of

the recently published Multiresolution Independent Component Analysis (MICA) algorithm in microarray classification be used to solve this problem. Training and classification with Support Vector Machines highlight the algorithm's effectiveness in MRI analysis (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

- Direct derivation of deformation parameters is not allowed in the system.
- The detection precision is poor.
- Prior information is necessary for this exceptionally difficult segmentation challenge.
- There are several significant obstacles associated with this segmentation process.

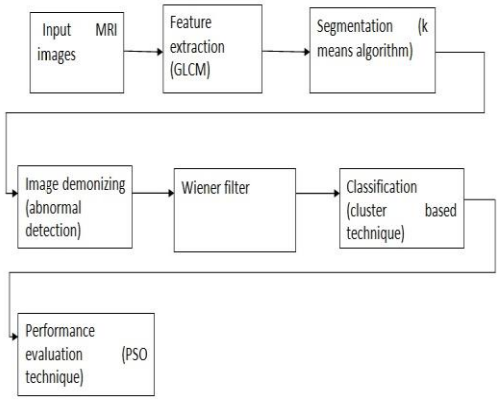
3.2 EXISTING SYSTEM:

Existing research suggests that by identifying anatomical patterns and exposing hidden relationships from structural magnetic resonance (MR) images, neuro imaging might become a significant tool in the early identification of neurodegenerative illnesses. One of the major goals of studying structural brain MR images is to look for anatomical alterations, either local or global, that are linked to functional disorders. Radiologists, in particular, study pictures by focusing on specific regions and comparing them to find differences. The support vector machine (SVM), which has been used to identify people with a variety of neurological illnesses, is by far the most used technology. Image synthesis methods are the ones that deal with the challenge of high-dimensionality.

3.3 PROPOSED METHOD:

To obtain reliable categorization of brain MR images as normal controls or likely AD individuals, the suggested method is based on a two-phase PSO model that combines bottom-up and top-down methods. The problem is classified using K-means with PSO based on clustering. The pre-defined kernels in the proposed method turn the input picture into individual feature saliency maps, whose pso match to kmeans space dimensions. The suggested technique is general, unlike other brain lesion segmentation algorithms that rely on outlier identification. It does not evaluate individual voxels and makes no assumptions about the abnormality's shape or intensity profile.

3.4 ARHITECTURE DIAGRAM

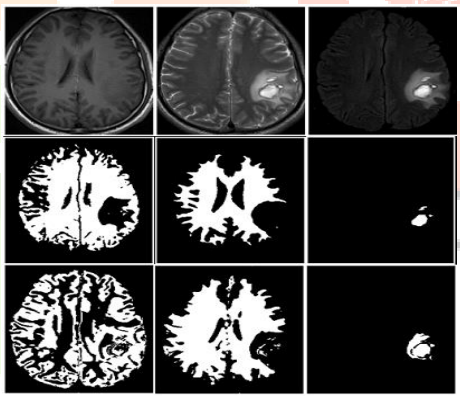


IV. SYSTEM IMPLEMENTATION

4.1 IMAGE SELECTION

Clinical trial MRI sequences are often collected with varying sizes and orientations. The first stage in the analysis procedure is to register the pictures to create a co-registered multispectral suite. A spectral signature is formed by each pixel vector in a multispectral image, and a collection of these spectral signatures generates multi signals.

Clinical Image Analysis



In visual and quantitative analysis, two types of clinical data are employed. The first dataset consists of T1WI, T2WI, and FLAIR pictures with the following parameters. From 20 anomalous instances, 70 multispectral slice sets were chosen for examination. The top row depicts slices from a multispectral picture series. The lesion is surrounded by edoema in T2WI and FLAIR images, although this information is not evident in T1WI. Figure 5 middle row and final row show the classified findings from ICA+SVM and MICA+SVM, with GM in the first column, WM in the second column, and abnormality in the third 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 2nd 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

Noise causes image corruption during transmission and capture. As a result, image restoration is required to remove the additive noise (mainly Additive White Gaussian Noise) from the image while preserving as many of the image's original qualities as possible. Using Wiener Filtering and Adaptive Median Filtering, we provide an image denoising method. Denoising has become a critical step in the image restoration process. Today, image denoising algorithms such as Wiener Filtering, Gaussian scalar mixture, and BM3D (Image denoising, 2010) are available. They've been employed in fields like medical imaging and astronomy with great success.

The Wiener filter's purpose is to remove noise that has damaged a signal. It employs a statistical methodology. Filters are often intended to have a specific frequency response. The Wiener filter, on the other hand, is designed differently. The spectral parameters of the original signal and the noise are assumed to be known, and the LTI filter whose output is as near to the original signal as feasible is sought. The following are characteristics of Wiener filters:

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}, … , F_{iM}), where F_{im} is the value of the m^{th} feature of the i^{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}, … , F_{iM}) and p_j = (F_{j1}, F_{j2}, … , F_{jM}), the distance, d_{ij}, between p_i and p_j is given by:

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

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

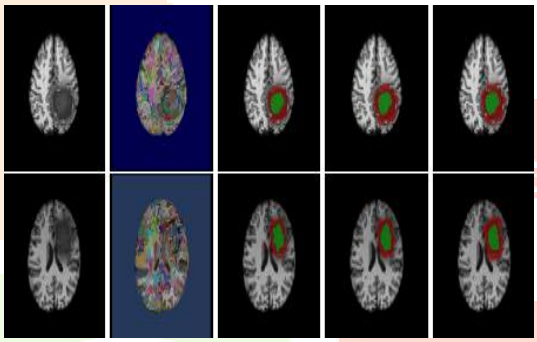
$$F^*_{im} = \frac{F_{im} - \mu_m}{\sigma_m} \qquad (2)$$

where F_{im} is the value of the m^{th} feature of the i^{th} job (i.e., the data point), μ_m is the mean value of the m^{th} feature, and σ_m is the standard deviation of the m^{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^{th} feature of the i^{th} cluster. Then,

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

where $F^*_{i,jm}$ is the m^{th} (scaled) feature value of the j^{th} job assigned to the i^{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)

PSO, as previously said, replicates bird flocking tendencies. Consider the following scenario: a flock of birds is seeking for food in an area at random. In the area being searched, there is just one piece of food. The birds are all confused about where the food is. So, how do you go about finding the food? The most successful method is to follow the bird closest to the meal. PSO took what it had learnt from the scenario and applied it to the optimization challenges. Each solution in PSO is a "bird" in the search space. It's known as "particle." All particles have fitness values that are assessed by the fitness function in order to be optimised, as well as velocities that direct the particles' flight.

4.6 IMAGE FEATURE EXTRACTION AND IMAGE FEATURE SELECTION

The purpose of feature extraction is to find locations in a scene that are next to each other. By borders, we mean groups of pixels that either divide objects or express changes in an object's surface shape. Step edges and crease edges are the two fundamental forms of boundaries considered in this project. In a single object, step edges indicate depth discontinuities, whereas crease edges represent a crease, or a discontinuity in the surface normal. The gradient magnitude may be used to identify steps edges in the range data, and the gradient magnitude of the 3D surface normal can be used to detect crease edges. To decrease additive noise, use an ator on both the range and amplitude data.

4.7 CLASSIFICATION

We demonstrated in the previous section that the DLD issue may be seen as a binary classification problem with just one conditional class probability. We now demonstrate that this view has algorithmic ramifications. Assume that we assign the label 1 to each sample of our training set T = (x1,...,xn) obtained from Q. Furthermore, we create a second training set T 0 = (x 0 1,...,x 0 n 0) from and label each sample with 1. A new training set is created by combining these labelled sample sets, which may subsequently be used by a binary classification method.

V. CONCLUSION AND FUTURE ENHANCEMENT

5.1 CONCLUSION

In recent years, the multispectral technique has greatly aided MRI analysts in reducing the time and accuracy of clinical trial analysis. However, due to the inefficiency of present approaches, extracting highly crucial information such as tiny lesions is a significant difficulty in pathology analysis. The proposed multi resolution analysis combined with ICA is Shown to be an effective method for resolving this problem. The method's performance in abnormality analysis is investigated and evaluated using SVM classification. The suggested technique outperforms ICA-based classifications in lesion/tumor identification, according to experimental results utilising synthetic and clinical data. Experiments on noisy synthetic pictures back up these findings, with MICA achieving satisfactory results. Refinements of MICA are under consideration to give equal priority to normal and abnormal tissue classification in future works.

5.2 FUTURE ENHANCEMENT

On an MR picture, we detect the afflicted portion in this project. We suggested block-adaptive windows, which considerably increase the quality of picture energy distribution predictions. Denoising performance is improved using the doubly local Wiener filtering approach with block-adaptive windows. Also, to figure out how the probabilistic label fusion model and the recently developed kmeans & pso segmentation approach are related. Another addition is that picture registration includes label information to increase registration accuracy. Registration refinement enhances segmentation accuracy, according to test data. The approach generates dependable clinical indicators that accord well with manual assessments. It may be beneficial in the diagnosis of cancer illness for physicians.

REFERENCES

[1] M. Jolly, “Fully automatic left ventricle segmentation in cardiacCine 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 cardiacMRimages usinga probabilistic atlas and the EM algorithm,”

[3] M. Lynch, O. Ghita, and P. F. Whelan, “Segmentation of the left ventricleOf the heart in 3- MRI data using an optimized NonrigidTemporal model,”

[4]C. Petitjean and J. N. Dacher, “A review of segmentation methodsIn short axis cardiac MR images,”

[5]M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing.NewYork: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.