ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Brain Tumor Statistical Structure Analysis And Prediction Based On Self Organised Map

A.Mary Joycy[1]

Assistant Professor, Department of CSE, RVS College of Engineering and Technology, Dindigul, Tamilnadu[1]

ABSTRACT: Magnetic resonance (MR) imaging has several advantages over other medical imaging modalities, including high contrast among different soft tissues, relatively high spatial resolution across the entire field of view and multi-spectral characteristics. Therefore, it has been widely used in quantitative brain imaging studies. This Paper analyses various clustering techniques to track tumor objects in Magnetic Resonance (MR) brain images. The input to this system is the MR image of the axial view of the human brain. Automated MRI (Magnetic Resonance Imaging) brain tumor segmentation is a difficult task due to the variance and complexity of tumors. In this paper, a statistical structure analysis based tumor segmentation scheme is presented, which focuses on the structural analysis on both tumorous and normal tissues. Firstly, two kinds of features including intensity-based, symmetry-based and texture-based are extracted from structural elements. Then a classification technique using AdaBoost that learns by selecting the most discriminative features is proposed to classify the structural elements into normal tissues and abnormal tissues.

KEYWORDS: SOM, Ada Boost, MRI

I. INTRODUCTION

The incidence of brain tumors has increased over the time and differs according to gender, age, race, and geography. In recent years, neurology and basic neuroscience have been significantly advanced by imaging tools that enable in vivo monitoring of the brain. In particular, Magnetic Resonance Imaging (MRI) has proven to be a powerful and versatile brain imaging modality that allows noninvasive longitudinal and 3D assessment of tissue morphology, metabolism, physiology, and function. In this paper we use luminosity-based segmentation method. This project analyses various clustering techniques to locate tumor objects in Magnetic Resonance (MR) brain images. The input is the MRI image of the axial view of the human brain. The Clustering algorithms used are K-means Clustering, Self Organizing Map (SOM) In this paper, a statistical structure analysis method is presented and applied to MRI brain tumor segmentation. Firstly, MR images are divided into small structure elements, and then three different kinds of features are extracted from each element, which quantify the intensity, symmetry, and texture properties of different tissues.

II. STRUCTURE ANALYSIS METHODOLOGIES

Three kinds of features are extracted block by block in one image. Secondly, AdaBoost algorithm is applied to select the most discriminative features and design a classifier to categorize the blocks into normal and tumorous groups. When a new image comes, only those selected features are extracted and the trained classifier is used to categorize the tumor in the image. The training and detection process flow of the proposed method is shown in figure 1. It should be noticed that the input images are preprocessed beforehand, including skull stripping which eliminates the skull from the brain image and scale normalization to adjust the intensity scale of the input images.

III. FEATURE EXTRACTIONS

Intensity-based features

Intensity-based statistical features are extracted from each block, including the mean intensity, maximum intensity, minimum intensity, range (maximum intensity minus minimum intensity), central pixel's intensity, variance, standard variance, median intensity, skewness, and kurtosis. The intensity values directly reflect the physical characteristics of

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

tissues in MRI, however, different tissues may have overlapping of intensity values. In order to achieve good segmentation performance, other information such as anatomic knowledge should also be considered.

Symmetry-based features

A remarkable characteristic of normal brain MR images is the symmetry of two cerebral hemispheres. The brain image with tumor will turn asymmetric because tumor usually occurs in one cerebral hemisphere and holds the normal structure's place. The simplest way to detect the asymmetry is subtracting one hemisphere from the other pixel by pixel. However, the human brain is not exactly symmetric, and there are always some slight variances. Thus in this paper, an asymmetry map *S* is calculated based on the original MR image *I*.

$$S(i,j) = \min_{(k,l)\in N(i',j')} |I(i,j) - I(k,l)| \quad (1)$$

(*i'*, *j'*) is the symmetric pixel of (*i*, *j*); *N* (*i'*, *j'*) is a small neighborhood of pixel (*i'*, *j'*), defined by equation (2); δ is the radius of *N*, which is a small value selected empirically.

$$N(i',j') = \{(k,l) \mid \|(k,l),(i',j')\| \le \delta\} \quad (2)$$

The symmetry-based feature is defined as the asymmetry map *S* value of the central pixel in each block.

AdaBoost

As the feature extraction strategy mentioned above, 3 kinds of features are extracted. However, not all the features are equally effective. AdaBoost learns the classification by selecting only those individual features that can best discriminate among classes. Furthermore it provides a final classifier as well as the feature selection strategy. The AdaBoost algorithm takes as input a training set (*x*1, *y*1), …, (*xm*, *ym*), where each *xi* belongs to the feature space *X*, and each *label yi* is in label set *Y* = {-1,+1}. -1 represents normal structures, and +1 represents tumorous structures. AdaBoost calls a given weak classifiers repeatedly in a series of rounds *t* = 1, …, *T*. One of the main ideas of the algorithm is to maintain a distribution or a set of weights over the training set. The weight of this distribution on training example *I* on round *t* is denoted by *Dt*(*i*). Initially, all weights are set equally. On each round, the most effective weak classifier is selected based on the current distribution, then the weights of incorrectly classified examples are increased so that the weak classifier is forced to focus on the hard examples. The final classifier is created by combining the weak classifier selected on each round. The outline for AdaBoost is given as below [7]. Given 1 1 (,), ,(,) *m m x y* L *x y* where , { 1, 1} *i i x* □ *X y* □*Y* = { 1, 1}

IV. EXPERIMENTAL RESULTS

We conducted our experiments on MR images from 10 different patients with gliomas. Each patient has 3 volumes of MR images, T1, T2, and FLAIR. Each volume contains 24 slices in axial plain with 5 mm slice thickness. MR imaging was performed on 3.0T Siemens devices. The imaging conditions of different protocols are: T1 weighted (TR=1680ms, TE=13ms, TI =800ms), T2 weighted (TR=5800ms, TE=103ms) and FLAIR weighted (TR=9000ms, TE=100ms, TI=2500ms). According to section 2, firstly the images are divided into small structure elements (blocks), and then 3 kinds of features are extracted from each block. It should be noticed that all the features are extracted respectively from multi-protocol MR images, T1, T2 and FLAIR, so the dimension number should be multiplied by 3. The ground truth is the tumor contour delineated by experienced doctors. From all the training images, 40000 blocks 20000 positive and 20000 negative) are extracted to train the AdaBoost classifier. Positive means normal tissues and negative means tumorous tissues. In order to test the classifier, 40000 blocks (20000 positive and 20000 negative) are extracted from the images in test set. The training and test error curves of the AdaBoost classifier as a function of the boosting round number are shown in figure 3. It means FLAIR provides the most information for tumor segmentation, T2 provides less and T1 provides the least. This result is in accordance with the conclusion in medical imaging, that FLAIR and T2 are more sensitive in pathological discrimination than T1. The distribution of the selected features is shown in table 1.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Table 1. **The distribution of selected features**

Number of Features	T1	T2	FLAIR	Total
Intensity Features	2	5	6	13
Symmetry Features	0	0	1	1
Texture Features	7	7	12	26
Total	9	12	19	40

On each round, AdaBoost selects a weak classifier with the minimum classification error in current distribution. Each weak classifier has a weight which determines its effectiveness in the final strong classifier. 3 features with the highest weights selected by AdaBoost (denoted by *F1*, *F2*, and *F3*) together with the original MR images are illustrated in figure 1 respectively. *F1*is the texture feature extracted from FLAIR image with *F2* is the maximum intensity value extracted from T1 image. *F3* is the median intensity value extracted from T2 image.

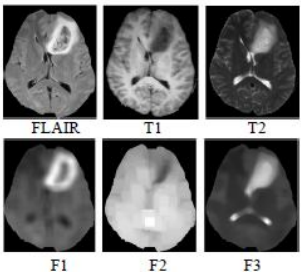

Figure 1. Original MR images(upper line) and 3 most effective features selected by AdaBoost *F1*, *F2* and *F3*(lower line)

Using the 40 selected features, the block classification accuracy on the test set by our algorithm is 98.74%. We compared this result with kNN (k nearest neighbors) algorithm and SVM (Support Vector Machine), which are widely used in medical image analysis. On the same training and test set, classification accuracies achieved by AdaBoost, kNN and SVM are shown in table 2. While using the 40 selected features, the accuracy on the test set by AdaBoost is 98.74%, which is better than 98.48% by kNN (choose k=7 which achieves the best result while ranging from 1~15) and 98.69% by SVM. The differences among three methods are not remarkable, but AdaBoost performs slightly better.

Table 2. **The classification accuracy of AdaBoost, kNN and SVM**

Classification accuracy	With feature selection	Without feature selection
AdaBoost	98.74%	98.55%
kNN	98.48%	95.47%
SVM	98.69%	96.46%

Some tumor segmentation results by the method presented in this paper are shown in figure 5. It can be observed that the results are very close to the delineations by doctors, which means our method is effective in MRI brain tumor segmentation. The correct rate, false positive rate (*FP*) and false negative rate (*FN*) of tumor segmentation are defined as below:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

$$FP = \frac{false\ positive\ pixel's\ number}{tumor\ size} \quad (5)$$

$$FN = \frac{false\ negative\ pixel's\ number}{tumor\ size} \quad (6)$$

$$correct\ rate = FP + FN \quad (7)$$

The average correct rate by the presented method is 96.82%, with *FP* of 1.3% and *FN* of 3.69%. The main factor influencing the accuracy is the presence of edema, which leads to high *FN* value, because the edema usually occurs beside the tumor and has similar appearance to the white matter. Both tumor and edema are abnormal tissues, so doctors are inclined to include the edema when delineating the tumor contours.

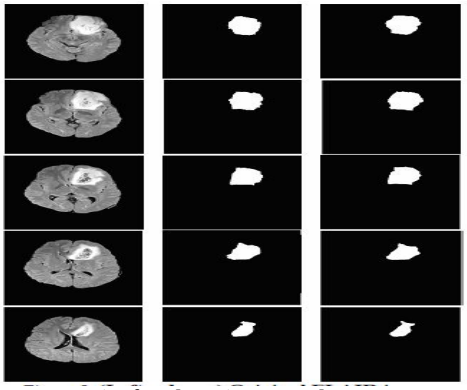

Figure 5. (Left column) Original FLAIR image; (Middle column) Tumor segmentation result by our method; (Right column) Ground truth

Fuzzy connectedness based segmentation method firstly calculate the fuzzy connected component of each pixel to the seed point using both intensity and space information, and then segment the fuzzy connected component image by region growing or threshold. In this paper, the seed point and the optimal segmenting threshold are manually selected. In table 3, it can be observed that our method performs better than both ACM and fuzzy connectedness based methods in tumor segmentation. The *FN* of our method is much lower than the other two methods. Because ACM and fuzzy connectedness methods both rely on some edge information between tumor and normal tissues, but the presence of edema obscure the edge between tumorous and normal tissues..

Table 3. **The segmentation accuracy of our method, ACM and fuzzy connectedness based method**

Segmentation Accuracy	*FP*	FN	Correct rate
Our Method	1.3%	3.69%	96.82%
ACM	1.84%	7.51%	90.65%
Fuzzy Connectedness	2.95%	5.02%	92.04%

V. SOM METHODOLOGIES

SOM followed by multiple range tests within clusters

The SOM were run using all the chemical and physical environmental variables and habitat metrics. A number of optimum clusters was then found. Subsequently, the distributions among clusters of the available indices of biotic integrity (fish for Minnesota and Ohio and benthic for Maryland) were plotted and a multiple range test among clusters was performed to determine if the differences within the clusters were statistically significant. A 95% confidence interval was used. The different statistically significant homogeneous group's distribution was obtained.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

SOM neuron-analysis: in this case we considered the neurons as the minimal, most homogeneous group of environmental values. In a SOM, one neuron groups a few sites with very similar characteristics. The values of each environmental variable and the biotic index in each neuron were averaged. The neuron-based environmental variables were then regressed against the neuron-based biotic index. Those variables with highest correlation were considered the most important for biotic integrity. Subsequently, we analyzed the relationships among different environmental variables, especially the relationships between off-stream and in-stream habitat parameters as well as the relationships between physical variables and chemical quality values. This was done by a simple neuron-based regression among the different variables.

SOM AS CLUSTERING TOOL

SOM were an interesting tool for us because they are able to represent highly dimensional environmental vectors in a 2D plot with a meaningful order. SOM are composed of multiple units called cells or neurons in which each environmental vector corresponding to each sampled different site is placed after a weighting algorithm. SOM were first developed by Kohonen in 1984. They are considered a type of unsupervised Artificial Neural Network

SOM

Where Xi(t) represents the environmental vectors and Wij(t) the neuron weights. Once this initialization layout is obtained, the algorithm constantly updates the weights by comparing the values among neighboring cells to further reduce the distances among neurons until convergence is reached (Kohonen, 2001). These weights are usually known as codebook vectors. The training is usually performed in two phases: relatively large initial learning rates and neighborhood radius are used in the first phase to initiate the SOM. In the second phase, both learning rates and neighborhood radius are then initially small to achieve further fine tuning of the SOM. In our case, the first tuning had 100 epochs and the fine tuning 20.

RESULTS I: SOM AND MULTIPLE RANGE TESTS

The environmental vectors available in the databases were used to find sets with similar characteristics. The clustering procedure was performed using all chemical and physical environmental variables. Subsequently, the biotic integrity indices and the environmental variables distribution within the clusters were plotted. A comparison between the distributions of the metrics and the biotic indices was performed in order to distinguish the most important metrics affecting biotic integrity.

OHIO

Clustering the database A total of 429 sites had values for each field. For this case the optimum number of clusters determined by the Davies- Even though the absolute minimum was obtained for seven clusters, we decided to choose three because it was easier for the sake of data interpretation and understanding. The SOM used in this case had a total of sixty neurons or cells.

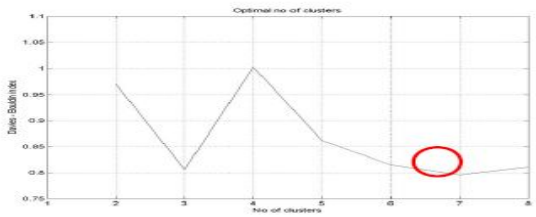

Figure 0-1. Optimum number of clusters. Ohio, all environmental variables
Habitat and biotic indices cluster distribution and analysis

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

The distribution of the habitat and biological indices using all the variables are as follows (in box plots, top line means 75th percentile, red line is 50th percentile and bottom line is 25thpercentile).

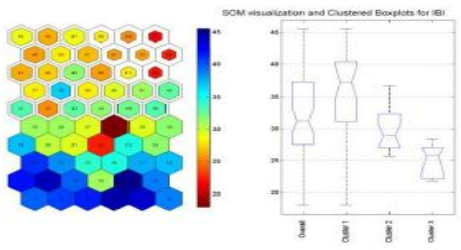

Figure 0-1. Fish IBI distribution among clusters

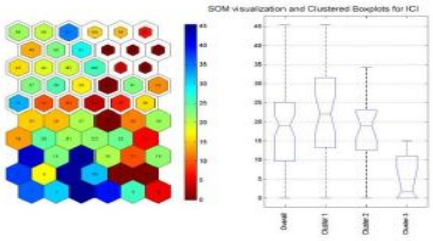

Figure 0-2. ICI distribution among clusters

The MRT tests were run to determine if the means' differences within the three clusters were statistically significant. Three homogeneous groups were found corresponding to each cluster. The MRT tests homogeneous groups are shown as follows:

Environmental variable cluster distribution and analysis

Land use and riparian area cluster distribution

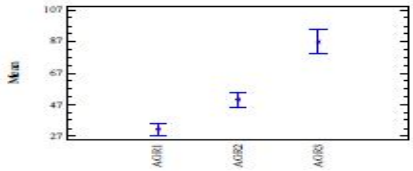

Figure 0-2. Riparian score means distribution

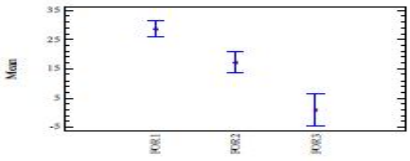

Figure 0-3. Agriculture LU means distribution

Other metrics showed statistically significant differences in only one of the clusters.

VI. CONCLUSIONS

Automated MRI brain tumor segmentation is a useful technique for diagnosis. In this paper, a statistical structure analysis method and its application to MRI brain tumor segmentation is presented. The method mainly includes 3 steps: structure elements subdivision, feature extraction, feature selection and classification. Experimental results demonstrate the features selected by our method can contribute effective and complementary information to discriminating tumor and normal tissues. SOM were an extremely useful tool in identifying sites with similar environmental stressors and were successful in revealing some of the very convoluted relationships among physical and chemical stressors and biotic integrity or among the physical and chemical stressors themselves. The clustering performed by the SOM followed by an analysis of the significant differences among clusters using Multiple Range Tests, and the subsequent comparison between biological and stressors' distributions, proved to be highly effective and successfully identified the variables that play a key role in biotic integrity, as proved in the SOM-neuron analysis. In all three states, either with the SOM+MRT analysis, the SOM-neuron analysis, or both, it was found that substrate and channel morphologic features are the two in-stream habitat parameters that have a deeper impact on biotic integrity.

REFERENCES

[1] Albert K. W. Law, F. K. Lam, and Francis H. Y. Chan, "A fast deformable region model for brain tumor boundary extraction", Proceedings of the Second Joint EMBS/BMES Conference, 2002, pp. 1055-1056.
[2] Jianguo. Liu, J. K. Udupa, et al., "A system for brain tumor volume estimation via MR imaging and fuzzy connectedness", Computerized Medical Imaging and Graphics, 2005(29), pp: 21–34.
[3] Marcel Prastawa, Elizabeth Bullitt, Nathan Moon, et al., "Automatic brain tumor segmentation by subject specific modification of atlas priors", Medical Image Computing,Academic Radiology, 2003(10), pp. 1341-1348.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

[4] P Y Lau, Frank C. T. Voon, and Shinji Ozawa, "The detection and visualization of brain tumors on T2-weighted MRI images using multiparameter feature blocks," Proceedings of the IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005, pp. 5104-5107.
[5] Angelo Zizzari, Udo Seiffert, and Bernd Michaelis. "Detection of tumor in digital images of the brain", Poceedings of the IASTED International Conference, 2001, pp. 132-137.
[6] K M Iftekharuddin, MA Islam, et al., "Automatic brain tumor detection in MRI: methodology and statistical validation," Proc. of SPIE Vol. 5747, 2005, pp. 2012~2022.
[7] Yoav Freund and Fobert E. Schapire, "A Short Introduction to Boosting", Journal of Japanese Society for Artificial Intelligence, 1999, 14(5), pp. 771-780.
[8] Xin Yuan and Shi Peng Cheng. "Microcalcification Detection Based On Localized Texture Comparison", International Conference on Image Processing (ICIP), 2004, pp. 2953~2956
[9] Oriol Pujol, Misael Rosales, Petia Radeva, and Eduard Nofrerias-Fernandez. "Intravascular Ultrasound Images Vessel Characterization Using AdaBoost", I. Magnin et al. (Eds.): Functional Imaging and Modelling of the Heart: