Computationally Intelligent Methods for Mining 3D Medical Images

Despina Kontos[1], Vasileios Megalooikonomou[1], Filia Makedon[2,3]

[1]Department of Computer and Information Sciences,
Temple University, 1805 N.Broad St., Philadelphia, PA 19122, USA
{dkontos, vasilis}@temple.edu
[2]Department of Computer Science, Dartmouth College, Hanover, NH, USA
[3]University of the Aegean, Greece
makedon@cs.dartmouth.edu

Abstract. We present novel intelligent tools for mining 3D medical images. We focus on detecting discriminative Regions of Interest (ROIs) and mining associations between their spatial distribution and other clinical assessment. To identify these highly informative regions, we propose utilizing statistical tests to selectively partition the 3D space into a number of hyper-rectangles. We apply quantitative characterization techniques to extract k-dimensional signatures from the highly discriminative ROIs. Finally, we use neural networks for classification. As a case study, we analyze an fMRI dataset obtained from a study on Alzheimer's disease. We seek to discover brain activation regions that discriminate controls from patients. The overall classification based on activation patterns in these areas exceeded 90% with nearly 100% accuracy on patients, outperforming the naïve static partitioning approach. The proposed intelligent tools have great potential for revealing relationships between ROIs in medical images and other clinical variables assisting systems that support medical diagnosis.

Keywords: data mining, diagnosis, information extraction, knowledge discovery, applications.

1 Introduction

Developing intelligent tools in order to extract information that supports decision-making has been of critical importance in fields such as knowledge discovery, information retrieval, artificial intelligence, and databases. Initially, mining problems have been grouped in three categories: identifying classifications, finding sequential patterns, and discovering associations [1]. Intelligent solutions for such problems are application-dependent and different applications usually require different mining techniques. A field where artificial intelligence (AI) has the potential of introducing challenging developments is medicine [2]. Systems developed under a pure AI perspective in the early years, such as MYCIN [3], Internist-1 [4] and DXplain [5] inspired a lot of hope for leveraging diagnosis by means of technological tools.

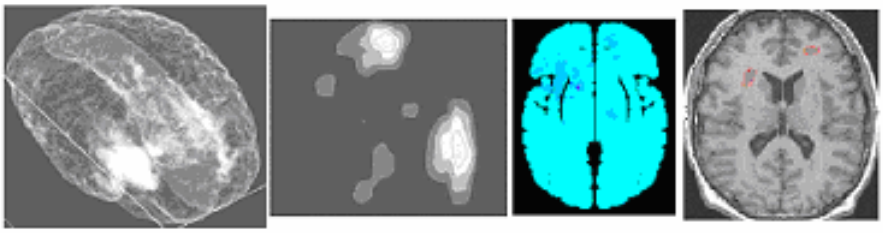


Fig. 1. Examples of Regions of Interest (ROIs) in medical images

Unfortunately, this initial esperance surrounding the deployment of intelligent diagnostic systems has been followed by the general lapse in funding for AI projects. Today, expert systems of this kind are more likely to be found in clinical laboratories and educational settings. On the other hand, subfields of AI such as data mining and machine learning have witnessed profound advancement. Tools developed under these disciplines have the ability to analyze large amounts of medical data and learn the underlying patterns, leading to the discovery of new phenomena and the extraction of medical knowledge. Looking for complex patterns within large medical data repositories and discovering previously unexpected associations can be of particular interest for understanding the development of several diseases.

2 Background

In this work we are interested in developing intelligent medical imaging tools that can support diagnosis. We focus particularly in brain imaging. We are interested in mining functional associations in the brain, focusing on highly informative Regions of Interest (ROIs). Figure 1 shows examples of such regions. Several techniques have been proposed for this particular purpose and large brain image data repositories have been developed [6], [7] that consist of 3-D images from different medical imaging modalities. These capture structural (e.g., MRI[1]) and/or functional/physiological (e.g., PET[2], fMRI[3]) information about the human brain. Techniques combining findings from several disciplines, such as AI, machine learning, pattern recognition, and data mining have been employed [8], [9] to analyze this vast amount of imaging data.

Two kinds of functional associations in the human brain are of particular interest when developing intelligent brain imaging tools. The first kind refers to associations between lesioned structures and concomitant neurological or neuropsychological deficits. The second includes associations between brain activation patterns and tasks performed. For this case, experiments are designed where subjects are asked to perform a certain task and their brain activation level is measured. A current obstacle in this type of analysis is the lack of intelligent tools to assist in diagnosis and medical decision making using methods that automatically classify such patterns (i.e., activation regions) and quantitatively measure levels of their similarity.

---

[1] Magnetic Resonance Imaging: shows soft-tissue structural information.
[2] Positron Emission Tomography: shows physiological activity.
[3] Functional-Magnetic Resonance Imaging: shows physiological activity.

The tools we present in this paper focus on analyzing 3D functional Magnetic Resonance Imaging (fMRI) that shows functional activity of the human brain. Current popular techniques employed for this purpose can be computationally expensive when analyzing activation levels or they do not model activation patterns entirely. More specifically, *statistical parametric mapping* (SPM) [10] analyzes each voxel's changes independently of the others and builds a corresponding map of statistical values. The significance of each voxel is ascertained statistically by means of Student's t-test, F-test, correlation coefficient, or other univariate statistical parametric tests. The multiple comparison problem (which occurs when computing a statistic for many pairwise tests) is usually handled by estimating corrected p-values for clusters. Although approaches have been proposed that seek to overcome the multiple comparison problem [11], they are based on a linearization of the 3D domain that might fail to preserve 100% the spatial locality of the ROIs.

Another approach to detect functional associations in the human brain is to model (estimate) their underlying distributions when distinct classes are present (controls vs. patients) [12], [13], utilizing parametric, non-parametric or semi-parametric techniques. EM and k-means algorithms [14] have been employed for this purpose, and statistical distance based methods have been used to distinguish among distributions. The Mahalanobis distance [15] and the Kullback-Leibler divergence [14] are most often employed. The main problem of these techniques is that real data are not accurately modeled using a simple mixture of Gaussian components, since they correspond to highly non-uniform distributions.

We seek to develop intelligent brain imaging tools that can provide decision-making support for diagnosis. We propose a unified framework for analyzing functional activity in the human brain. Our approach consists of two basic steps. The first is based on an adaptive recursive partitioning of the 3D domain to discover discriminative areas. This technique reduces the multiple comparison problem encountered in voxel-based analysis by applying statistical tests to groups of voxels. Compared with [11] this step of analysis is performed directly on the 3D domain (hyper-rectangles) without any loss of spatial locality. For classification, to avoid problems with distribution estimation techniques that are not suitable for non-uniform real datasets, we use neural networks having as inputs measurements obtained from highly discriminative ROIs. The second step is to further characterize these highly informative ROIs by extracting k-dimensional feature vectors (signatures) the uniquely represent them. As a case study we look at a group of patient and control subjects from a study on Alzheimer's disease (AD) [16]. Our intention is to develop intelligent tools that can provide both good classification and in depth quantitative analysis of discriminative activation patterns expressed by ROIs. In the context of the proposed framework we want to support diagnosis when the fMRI image of a new subject is presented. In other words, we seek to determine the group to which it belongs, i.e., control versus patient.

3 Methodology

The tools we propose combine methodologies initially presented in the field of data mining and image processing. We focus on mining associations between fMRI activa-

tion and other non-spatial attributes (i.e. clinical assessment). Furthermore we provide an efficient characterization mechanism for representing and compacting highly informative ROIs such that classification, indexing and similarity searches are feasible under the perspective of a medical imaging repository. In the discussion that follows we present the method for a two-class problem although it can be easily extended to more than two classes.

For the first step of the analysis we employ *Adaptive Recursive Partitioning* (ARP) that has been so far applied mainly to realistic and synthetic 3D region datasets of discrete (binary) voxel values [17]. Some initial results from attempts to apply the technique on real fMRI datasets have been presented in [18]. The main idea of this technique is to treat the initial 3D volume as a hyper rectangle and search for informative regions by partitioning the space into sub-regions. The intelligence of the tool lies in the selectivity of partitioning the hyper rectangles in an adaptive way. Only hyper rectangles that do not exhibit statistically significant discriminative power are selected to be partitioned recursively. More specifically, for each sample, we use the mean V_{mean} of all voxel values belonging to the volume (hyper-rectangle) under consideration as a measurement of activation/deactivation level. The adaptive partitioning of the 3D space continues in the following way: A hyper-rectangle is partitioned only if the corresponding attribute V_{mean} does not have a sufficient discriminative power to determine the class of samples. To decide this, we can apply statistical parametric (e.g. t-test [19]) or non-parametric tests (e.g. Wilcoxon rank sum [20]). The procedure progresses recursively until all remaining sub-regions are discriminative or a sub-region becomes so small that cannot be further partitioned. For this purpose, we define the maximum number of partitioning steps (depth) that the partitioning can go through. If the splitting criterion is satisfied, the spatial sub-domain (or hyper-rectangle) corresponding to the node of the oct-tree is partitioned into 8 smaller sub-domains. The corresponding tree node becomes the parent of eight children nodes, each corresponding to a subdomain and the new measurements V_{mean} corresponding to the region data in the sub-domains become new candidate attributes. Observe that the proposed method effectively reduces the multiple comparison problem encountered when using voxel-based analysis. The number of times a statistical test is applied is significantly reduced since we selectively deal with groups of voxels (hyper rectangles).

After detecting ROIs of highly discriminative activation we propose a second step of detailed quantitative characterization of these regions, aiming to extract unique signatures. We apply a method that efficiently extracts a k-dimensional feature vector using concentric spheres in 3D (or circles in 2D) radiating out of the ROI's center of mass, initially presented in [21] and applied on artificially generated data. Here we demonstrate the potential of the technique to be utilized for characterizing real ROIs. The proposed technique extends the original idea of Sholl's analysis [22] (i.e. the use of concentric circles radiating out of the root of the tree to partition a tree-like structure) to non-tree like structures. The process is described by the following steps: (i) estimate the center of mass, *m,* of the region (for non-homogeneous regions this is calculated using a weighted contribution based on each voxel's value), (ii) construct a series of 1,…,k concentric spheres in 3D (or circles in 2D) radiating out of *m*, using regular increments of radius, and (iii) construct the feature vectors f_s and f_r of size k measuring respectively at each increment the fraction of the sphere (or circle) occupied by the region and the fraction of the region occupied by the sphere (circle).

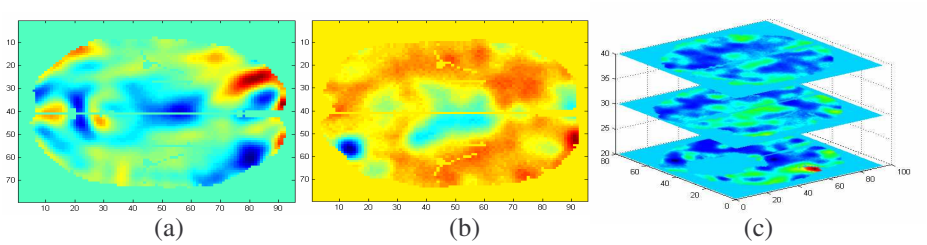
Fig. 2. Intersecting concentric circles with the ROI being characterized

The feature vectors obtained are of the form (a) $f_s = [f_{1s}, f_{2s},..., f_{ks}]$ or (b) $f_r = [f_{1r}, f_{2r},..., f_{kr}]$ respectively. The features f_{is} or f_{ir} (where i=1,..,k), obtained at each increment of radius, express the sum of voxels belonging to the intersection of the sphere (or circles in 2D) with the ROI, divided by (a) the total number of voxels belonging to the sphere or (b) the total number of voxels belonging to the ROI. The sum of voxels for the non-homogeneous ROIs are calculated by a weighted contribution of each voxel, based on its value. Figure 2 illustrates a snapshot of the characterization process for a ROI in 2D. This technique has been shown to be two orders of magnitude faster than mathematical morphology (namely the "pattern spectrum") although it achieves comparable to or even better characterization results [21].

The purpose of extending these two approaches to be applicable on real data and combining them in the context of a unified approach is to create an intelligent brain informatics tool. This can be useful for mining associations between spatial patterns and clinical assessment as well as providing compact characterization for interesting ROIs, overall assisting diagnosis with classification and similarity searches. One of the computational advantages of the proposed tool is that it operates on groups of voxels (hyper-rectangles) significantly reducing the multiple comparison problem encountered when applying statistical tests on a voxel wise basis (SPM). Finally, the selectivity that the system exhibits when partitioning the space in an adaptive recursive manner guides the analysis to focus only on highly informative ROIs, avoiding unnecessary processing.

4 Experimental Evaluation

Our dataset consisted of 3D activation contrast maps of 9 controls and 9 Alzheimer's disease (AD) patients. The task was designed to probe semantic knowledge of categorical congruence between word pairs, exploring neuroanatomical correlates in AD [16]. Figure 3 shows sample views of these contrast activation maps. Preprocessing of the data included spatial normalization, i.e. registration to a standard template. Each subject's task-related activation was analyzed individually versus the subject's rest condition, resulting in individual contrast maps giving a measurement of fMRI signal change at each voxel. Background noise was removed by subtracting the signal value measured in representative background voxels from all the voxels of the 3D volume. Finally, we masked the data using a binary mask extracted from the T1 canonical atlas that was used as the template for the registration. Only signal within the binary mask was included in the analysis.



(a) (b) (c)

Fig. 3. Sample views of the contrast activation maps in our dataset. 2D slices of (a) a control and (b) a patient sample. 3D view of 2D slices for a sample fMRI activation volume (c)

4.1 Mining Informative Patterns

As a first step of mining informative patterns and associations we applied ARP using as splitting criterion the t-test with threshold levels for the p-value 0.05 and 0.01. A p-value reflects the probability of observing a test statistic as extreme as or more extreme than the observed value, assuming that the null hypothesis is true. In our case, the null hypothesis is that the two groups do not differ significantly with respect to activation levels. The values of 0.05 and 0.01 are the typical values used in the literature for such statistical tests. The maximum allowed tree depth was set to either 3 or 4. ARP uses these parameters to refrain from further partitioning a 3D hyper-rectangle. The above values for the tree depth were determined based on the resolution of the original images and a trade-off between the size of the discovered regions and the number of tests performed. Due to space limitations, in Figure 4, we present the indicated ROIs for a significance threshold of 0.05 and a maximum tree depth of 3, overlaid on the T1 canonical atlas template. The significance of each region is annotated using a color coding (colorbar). The majority of significant regions determined by the proposed approach that could discriminate Alzheimer patients from controls were within the medial temporal lobe. The findings of multiple distributed regions in this area that differentiate patients and controls, as detected by ARP, is consistent with atrophy observed in widespread cortical and subcortical areas in AD [23] and may be consistent with a distributed reorganization of networks subserving the semantic memory task [16].

To further verify the validity of these results we include the following classification experiments that can be viewed as building a model for assisting in diagnosis. More specifically, for the classification model we used Neural Networks. To avoid overfitting due to a small training dataset we applied one-layer perceptron networks trained by the Pocket algorithm [24]. As inputs to the classifier we used the attributes V_{mean} of the discovered regions (after being standardized to have zero mean and unit standard deviation), and a binary class label indicating the class of the samples (control vs. patient). The leave-one-out approach was employed to evaluate out of sample classification performance [14],[15]. More specifically, the training set consisted of patients and controls with indices 1,2,3,…,i-1,i+1,…9, and the method was tested on patient and control with index i, where i=1,…,9. Taking into account the stochastic nature of the Pocket algorithm, we repeated the process of training and testing the model in each of the leave-one-out loops for 5 times and averaged the percentage of the

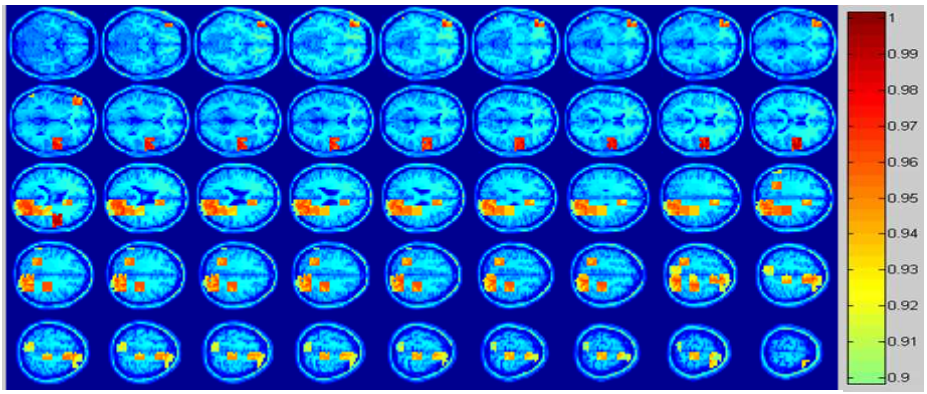


Fig. 4. Transaxial view of the T1 canonical atlas showing the areas discovered by ARP when applied with parameters: significance threshold = 0.05, maximum tree depth = 3

correct predictions to obtain the reported accuracy. Table 1 shows the overall classification accuracies as well as those obtained separately for controls and patients. These results support the argument that the regions discovered by ARP in the specific study are indeed associated with AD, thus providing significant discriminative information.

To provide a comparison basis for the proposed tools we implemented as well a static partitioning approach. This approach is naïve (as compared to the adaptive partitioning of the space) and simply partitions the space into equal length hyper-rectangles. Each dimension is actually split in l equal length bins, resulting in a total partitioning of the space of $l \times l \times l$ hyper-rectangles for the 3D domain. Again the V_{mean} of each sub-region is used as a representative attribute and the same classification model is employed. Table 2 demonstrates the classification accuracies for this scenario. It is clear that the adaptive approach outperforms the static partitioning approach, being able to indicate as well specific patterns (ROIs) where discriminative activation is observed.

Table 1. Classification accuracy based on the discriminative regions detected by ARP

Criterion	Threshold	Depth	Accuracy		
			Controls	Patients	Total
t-test	0.05	3	89%	100%	**94%**
	0.05	4	84%	100%	92%
	0.01	4	87%	100%	93%

Table 2. Classification accuracy based on the static partitioning

l	Accuracy		
	Controls	Patients	Total
2	58.89%	71.11%	65%
3	57.78%	78.89%	68.33%
4	100%	0%	50%

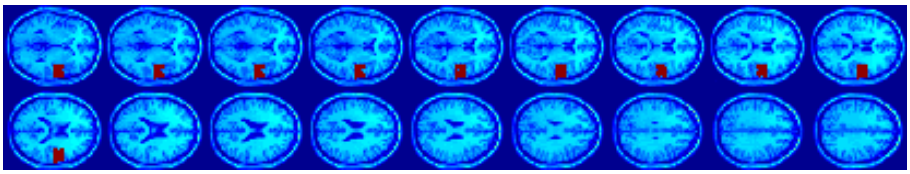


Fig. 5. The ROI used for applying the proposed feature selection technique, shown in consecutive 2D slices after being overlaid on the T1canonical brain atlas

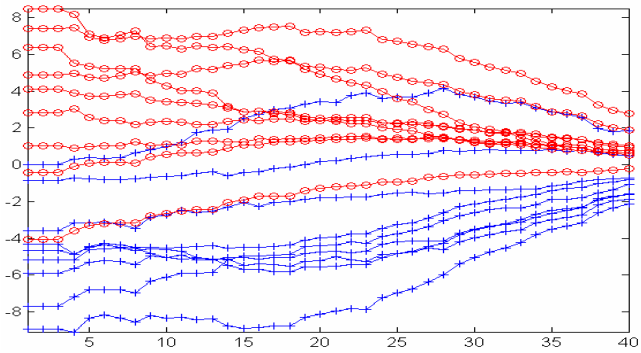


Fig. 6. The obtained f_s characterization signatures from the highly discriminative ROI. Signatures with '+' correspond to controls (blue), 'o' to patients (red)

4.2 Characterizing Highly Informative Regions

Here, we demonstrate an example of applying the proposed quantitative characterization technique described in Section 3 in order to extract unique signatures from the highly informative regions. The ROI that we focus on was constructed by two neighboring sub-regions within the medial temporal lobe of the human brain. These sub-regions have p-values of 0.0012 and 0.0025 respectively when using a t-test to determine the significance of their association with Alzheimer's disease in the experiments of section 4.1. Figure 5 illustrates the selected ROI after being overlaid on the T1 canonical atlas. We experimented with a radius increment of 0.02 extracting feature vectors of length 40. Figure 6 shows the obtained f_s feature vectors. As we can observe, signatures of subjects of the same class tend to cluster following similar behavior and the two classes barely overlap. The curvature of the signatures conveys information about the activation patterns of the original data. As demonstrated initially in [21] with synthetic data, using morphological operators for such an analysis is two orders of magnitude slower than the approach employed here.

As illustrated, patient samples exhibit positive activation in the specific ROI, whereas the control subjects have lower negative activation (deactivation) levels. This information is highly discriminative and the proposed characterization technique has the ability to represent the initial ROI in a compact form. These signatures provide both quantitative and qualitative information and can be utilized for indexing and similarity searches in the framework of a medical imaging data repository that can assist clinical decision-making and diagnosis.

5 Conclusions

We proposed a framework for constructing computationally intelligent medical informatics tools. These tools combine data mining and image processing techniques, extending them to be applied on real fMRI data. The focus is to mine associations between spatial patterns and other non-spatial clinical assessment, employing an adaptive partitioning of the space guided with statistical tests. At the same time we seek to characterize highly informative regions, providing compact signatures that uniquely identify the ROIs. These can be utilized for indexing and similarity searches in the context of a medical imaging data repository. As a case study, we analyzed an fMRI dataset obtained from a study that explores neuroanatomical correlates of semantic processing in Alzheimer's disease. We evaluated the validity of our findings providing classification experiments with neural networks. The overall classification based on activation patterns in these areas exceeded 90% with nearly 100% accuracy on patients outperforming the naïve static partitioning approach. The proposed intelligent tools have great potential for elucidating relationships between ROIs in medical images and other clinical variables assisting in medical decision-making.

Acknowledgement

The authors would like to thank A. Saykin for providing the fMRI data set and clinical expertise and J. Ford for performing some of the preprocessing of this data set. This work was supported in part by the National Science Foundation (NSF) under grants IIS-0237921 and IIS-0083423. NSF specifically disclaims responsibility for any analyses, interpretations and conclusions.

References

1. Agrawal, R., Imielinski, T., and Swami, A.: Database Mining: A performance Perspective. IEEE Transactions on Knowledge and data Engineering,5(6) (1993) 914-925
2. Coiera, E., Editorial: Artificial Intelligence in Medicine - The Challenges Ahead, Journal American Medical Informatics Association, 3,6 (1996) 363-366
3. Buchanan, B. G. and Shortliffe, E. H., editors. Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Reading, MA, (1984)
4. Miller, R.A., Pople, H.E., Myers, J.D.: INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. New Eng. J. Med. (1982) 307:468-476
5. Barnett, O., Cimino, J.J, Hupp, J. A. and Hoffer, E.P.: DXplain: An evolving diagnosis decision-support system. Journal of the American Medical Association, (1987) 258:67-74
6. Arya, M., Cody, W., Faloutsos, C., Richardson, J., and Toga, A.: A 3D Medical Image Database Management System. Int. Journal of Computerized Medical Imaging and Graphics, Special issue on Medical Image Databases, 20(4) (1996) 269-284
7. Letovsky, S., Whitehead, S., Paik, C., Miller, G., Gerber, J., Herskovits, E., Fulton, T., and. Bryan, R.: A brain-image database for structure-function analysis. American Journal of Neuroradiology, 19(10) (1998) 1869-1877

8. Megalooikonomou, V., Ford, J., Shen, L., Makedon, F., Saykin, F.: Data mining in brain imaging, Statistical Methods in Medical Research, vol. 9, no. 4 (2000) 359-394
9. Megalooikonomou, V., Davatzikos, C., Herskovits, E.: Mining lesion-deficit associations in a brain image database, in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, (1999) 347-351
10. Friston, KJ., Holmes, AP., Worsley, KJ., Poline, JP., Frith, CD., Frackowiak, RSJ.: Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping, (1995) 189–210
11. Kontos, D., Megalooikonomou, V., Ghubade, N., Faloutsos, C., Detecting discriminative functional MRI activation patterns using space filling curves, in Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Cancun, Mexico, (2003) 963-967
12. Lazarevic, A., Pokrajac, D., Megalooikonomou, V., Obradovic, Z.: Distinguishing Among 3-D Distributions for Brain Image Data Classification, in Proceedings of the 4th International Conference on Neural Networks and Expert Systems in Medicine and Healthcare, Milos Island, Greece, (2001) 389-396
13. Pokrajac, D., Lazarevic, A., Megalooikonomou, V., Obradovic, Z.: Classification of brain image data using meaasures of distributional distance, 7th Annual Meeting of the Organization for Human Brain Mapping (OHBM01), Brighton, UK, (2001)
14. Duda, R., Hart, P., Stork, D.: Pattern Classification, John Wiley and Sons, NY, (2000)
15. Fukunaga, K.: Introduction to Statistical Pattern Recognition, Academic Press, San Diego, (1990)
16. Saykin, A.J., Flashman, L.A., Frutiger, S.A., Johnson, S.C., Mamourian, A.C., Moritz, C.H., O'Jile, J.R., Riordan, H.J., Santulli, R.B., Smith, C.A., Weaver, J.B.: Neuroanatomic substrates of semantic memory impairment in Alzheimer's disease: Patterns of functional MRI activation, Journal of the International Neuropsychological Society, 5 (1999) 377-392
17. Megalooikonomou, V., Pokrajac, D., Lazarevic, A., V., Obradovic, Z.:, Effective classification of 3-D image data using partitioning methods, in Proceedings of the SPIE 14th Annual Symposium in Electronic Imaging: Conference on Visualization and Data Analysis San Jose, CA, Jan. (2002)
18. Megalooikonomou, V., Kontos, D., Pokrajac, D., Lazarevic, A., Obradovic, Z., Boyko, O., Saykin, A., Ford, J., Makedon, F.: Classification and Mining of Brain Image Data Using Adaptive Recursive Partitioning Methods: Application to Alzheimer Disease and Brain Activation Patterns", presented at the *Human Brain Mapping Conference (OHBM'03)*, New York, NY, Jun. (2003)
19. Devore, J.L.:Probability and Statistics for Engineering and the Sciences, 5th edn., International Thomson Publishing Company, Belmont, (2000)
20. Conover, W.J.: Practical Nonparametric Statistics, Wiley, New York, (1999)
21. Megalooikonomou, V., Dutta, H., Kontos, D.: Fast and Effective Characterization of 3D Region Data, In Proc. of the IEEE International Conference on Image Processing (ICIP), Rochester, NY, (2002) 421-424
22. Sholl, D.: Dendritic Organization in the Neurons of the Visual and Motor Cortices of the Cat," Journal of Anatomy., vol. 87, (1953) 387-406
23. Flashman, L.A., Wishart, H.A., Saykin, A.J.: Boundaries Between Normal Aging and Dementia: Perspectives from Neuropsychological and Neuroimaging Investigations, in: Emory VOB and Oxman TE, editors. Dementia: Presentations, Differential Diagnosis and Nosology. Baltimore: Johns Hopkins University Press, (2003) 3-30
24. Gallant, S.I.: Perceptron-Based Learning Algorithms, in IEEE Transactions on Neural Networks, vol. 1, no. 2 (1990) 179-191