SCI INSTITUTE
TECHNICAL REPORT

Automatic Classification of Alzheimers Disease vs. Frontotemporal Dementia: A Decision Tree Approach with FDG-PET

Neda Sadeghi[*], *Tolga Tasdizen*[*], *Norman L. Foster*[†], *Angela Y. Wang*[†], *Satoshi Minoshima*[‡], *Andrew P. Lieberman*[§].

[*]School of Computing, University of Utah, [†]Center for Alzheimers Care, Imaging and Research, University of Utah, [‡]School of Medicine, University of Washington, [§]Department of Pathology, University of Michigan

UUSCI-2007-016

Scientific Computing and Imaging Institute
University of Utah
Salt Lake City, UT 84112 USA

October 29, 2007

Abstract:

We introduce a novel approach for the automatic classification of FDG-PET scans of subjects with Alzheimers Disease (AD) and Frontotemporal dementia (FTD). Unlike previous work in the literature which focuses on principal component analysis and predefined regions of interest, we propose the use of decision tree learning combined with empirically determined regions of interest as attributes. The advantages of this approach are two-fold. First, empirically determining regions of interest for distinguishing between these two diseases is relevant for clinical medical practice. Second, we illustrate that the proposed method provides better classification accuracy compared to other methods on a group of 48 autopsy confirmed AD and FTD patients.

THE UNIVERSITY OF UTAH

Automatic Classification of Alzheimer's Disease vs. Frontotemporal Dementia: A Decision Tree Approach with FDG-PET

Neda Sadeghi[1], Tolga Tasdizen[1], Norman L. Foster[2], Angela Y. Wang[2], Satoshi Minoshima[3], Andrew P. Lieberman[4]

[1] School of Computing, University of Utah, Salt Lake City, UT 84112
[2] Center for Alzheimer's Care, Imaging and Research, University of Utah, Salt Lake city, UT 84112
[3] School of Medicine, University of Washington, Seattle, WA 98195
[4] Department of Pathology, University of Michigan, Ann Arbor, MI 48109

Abstract. We introduce a novel approach for the automatic classification of FDG-PET scans of subjects with Alzheimer's Disease (AD) and Frontotemporal dementia (FTD). Unlike previous work in the literature which focuses on principal component analysis and predefined regions of interest, we propose the use of decision tree learning combined with empirically determined regions of interest as attributes. The advantages of this approach are two-fold. First, empirically determining regions of interest for distinguishing between these two diseases is relevant for clinical medical practice. Second, we illustrate that the proposed method provides better classification accuracy compared to other methods on a group of 48 autopsy confirmed AD and FTD patients.

1 Introduction

Even with advances in clinical diagnosis, it is still hard to distinguish among different neurodegenerative diseases with confidence. Correct diagnosis is challenging for clinicians because many of these disease share the same behavioral symptoms. Positron emission tomography (PET) image analysis has great potential to aid clinical diagnosis. In PET imaging, a radioactive tracer isotope incorporated into a metabolically active molecule such as FDG, a type of sugar molecule, is injected into the subject allowing the imaging of metabolic activity for glucose during life. Since brain energy normally is completely dependent upon glucose, FDG-PET accurately reflects brain function. FDG-PET imaging by providing quantitative localization of metabolic activity promises a better accuracy rate compared to qualitative judgments required to clinically distinguish different types of dementia. Determining whether a patient has Alzheimer's disease (AD) vs. frontotemporal dementia (FTD) is a prominent example where PET imaging can help distinguish between two diseases that have very similar clinical symptoms. However, analysis of these FDG-PET with anything more than visual interpretation is time intensive and not routinely used in clinical settings. Therefore, computational methods that can expedite analysis using more precise quantitative methods are of great interest.

One of the important challenges in using PET images for automatic diagnosis in neurology is that brain images contain a very large number of 3D pixels (tens of thousands or more) compared to a much smaller number of subjects (typically a few dozen or less) from which a classifier must be learned. This discrepancy in the dimensionality of the data as compared to the sample size makes robustly learning classifiers a challenging problem. Researchers have taken various approaches to overcome this problem. One of these approaches is region of interest (ROI) analysis. This method considers regions definition based on prior knowledge and analyzes them for abnormalities [1, 2]. By using summary measures computed over regions instead of pixels as individual attributes, the dimensionality of the data is vastly reduced. The problem with this method is that strong prior knowledge is needed, and sometimes the abnormalities are not unique to a specific area, they may overlap or cross multiple predefined ROIs. Furthermore, this limits new discoveries in new regions. As a result of these limitations with using ROI, pixel-by-pixel analysis has become a focus for research [3, 4]. However, this analysis approach has its own shortcoming; this method is less immune to noise in the image, and is subject to registration errors, and variability between individuals. In an effort to limit the number of variables while still incorporating all

image data without biasing preconceptions. Linear methods, such as principal component analysis (PCA) have been used to project the image down to a few attributes. However, PCA is not good at capturing complex, non-linear relationships in high-dimensional space. Furthermore, PCA is prone to errors if portions of the image data are missing. For instance, parts of the cerebellum can be missing from PET images due to non-optimal patient placement in the scanner. In such cases, these pixels must be manually excluded before PCA can be applied.

In this paper we introduce an alternative, method for evaluating PET images using a decision tree learning algorithm [5]. The inputs to the decision tree are binary attributes based on the thresholding of z-scores of groups of adjacent pixels. Z-scores for these groups of pixels are computed using a database of normal controls. This method is appealing for a number of reasons. First, it takes into account all the pixels in the cortex, and computes z-scores for spatial neighborhoods instead of single pixel, then the decision tree evaluates which areas have the most discriminative power and recombines regions as necessary to create largest possible regions with maximum information gain. One advantage of decision tree learning over PCA is the ability to choose a binary mask defining a region of interest rather than using a weighted sum of all pixels in the image. In other words, decision trees allow us to pick certain regions rather than using all the pixels. This is important from the point of view of clinical practice where defining regions that are affected in certain disease is a question of interest. Furthermore, decision tree learning picks regions dynamically rather than based on predefined regions that are used in ROI analysis. Finally, because of the grouping of adjacent pixels, this method is less sensitive to noise or registration error. The combination of attribute selection that is proposed in this paper and the use of decision tree learning makes a good choice for analyzing brain images. In this paper, we introduce this approach as a robust method for locating the areas of brain that can be used to distinguish between AD and FTD subjects. The proposed method is completely automated and evaluating test subjects after the decision tree has been learned is of negligible computational cost. This classification task is more subtle and more difficult than distinguishing between one of these diseases and normal subjects. The rest of this paper will discuss specifics of the proposed method and discuss validation results using FDG-PET images of a group of 48 autopsy-confirmed AD and FTD patients.

2 Method

Our approach to automatic AD vs. FTD classification is to learn a decision tree classifier using the ID3 algorithm. The ID3 algorithm chooses what attributes to use at different nodes of the tree based on the information gain criterion. Starting from the root node which contains all training examples, an attribute is chosen for that node which splits the node into to child nodes. Each child node which has examples of both classes (AD and FTD) are then split by picking an attribute using the same principles and more child nodes are created. The process stops when all nodes have only one class (AD or FTD) associated with them. However, the novelty of our approach is not in the specific classification algorithm, but in the definition of the attributes that the classifier is based on. The proposed attributes are empirically determined regions of interest on the cortical surface that have maximum spatial extent and maximum information gain as described next. We describe them in detail next.

2.1 Pre-Processing

Warping an image into a template space, which is necessary for performing comparisons across subjects, is typically the first step in group analysis. The MNI coordinates used by the popular Statistical Parametric Mapping (SPM) package is an example of this image warping. For this study, we choose to use the stereotactic surface projection (SSP) package [6] instead of SPM. AD and FTD predominantly affect the gray matter in the cerebral cortex; therefore, we are concerned with metabolic activity of this region rather than the entire brain. After warping 3D images into the Talairach coordinate system [7], SSP uses a three-dimensional stereotactic technique to extract the peak metabolic activity in the cerebral cortex and project it to the brain surface[6]. Using this technique, 3D FDG-PET images are converted into a list of 16000 surface pixels and their associated metabolic values. Since each patient has a different overall metabolic rate, and global scaling of the image values depends on the specific scanner used, we need a way to

normalize the data in order to minimize the global differences among patients. Normalization using global average metabolism in the brain is problematic because it is affected by disease severity. The Pons located on the brain stem is known to be relatively spared by AD and FTD [8] which makes it a good candidate for normalization. In this paper, the data was normalized to the pontine metabolic value which was also extracted from the FDG-PET images by SSP.

2.2 Attribute Selection

In this section, we describe a method for locating different areas of the cerebral cortex projected to the surface of the brain that have the most discriminative power to distinguish between FTD and AD. We would like these areas to be as large as possible, without losing their discriminative power. Due to the averaging effect, measures computed from larger regions are less susceptible to problems cause by additive, independent noise. Also, comparing areas rather than comparing individual pixels reduces the effects of any registration error in warping the brain to the SSP coordinates.

In the context of classification, information gain measures how well a given attribute separates the training examples into their target classifications. A binary attribute separates a set of training examples Ω into two mutually exclusive and exhaustive subsets Ω_t and Ω_f, i.e. the subsets for which the attribute is true and false, respectively. Then information gain is defined as

$$InfoGain = H(\Omega) - \left(\frac{|\Omega_t|}{\Omega} H(\Omega_t) + \frac{|\Omega_f|}{\Omega} H(\Omega_f) \right), \quad (1)$$

where H denotes the entropy of a set. Let $p_\Omega(AD)$ and $p_\Omega(FTD)$ denote the fraction of examples with AD and FTD target classifications in the set Ω. Then entropy is calculated as

$$H(\Omega) = -\left(p_\Omega(AD) \log_2 p_\Omega(AD) + p_\Omega(FTD) \log_2 p_\Omega(FTD) \right). \quad (2)$$

If a set has examples that all belong to the same class then it has entropy 0; on the other hand, equal fractions of two classes result in a maximum entropy of 1.

Let C denote the set of locations that are extracted by SSP as cortical locations. Ideally, we want to choose a subset of C that is as large as possible while having maximum information gain in its metabolic values. However, this problem is intractable for large C due to the large number of combinatorial possibilities that arise in defining regions. However, we can easily find approximations to maximally large regions with high information gain using the following procedure. Given a location x in C, define a set of neighboring locations as other points on the cortical surface that are within distance R of the point of interest x

$$S(x,R) = \left\{ x' \in C \middle| \|x - x'\| < R \right\} \quad (3)$$

We will refer to these sets of locations as building-block neighborhoods. Next, we calculate the total metabolic activity for the building-block neighborhood at each x using

$$A_k(x) = \sum_{x' \in S(x,R)} M_k(x'), \quad (4)$$

where $M_k(x)$ is the metabolic activity normalized to pons at pixel x, for the k'th subject. While metabolic activity can be used directly as an attribute in building a classifier, approaches in the Neurology literature typically first convert metabolic scores to z-scores using a database of normals [8]. We follow the same approach here. Using the same definitions given in Equations (3) and (4), we can also calculate the total metabolic activity for normal controls. Then, given a point of interest x, we compute the mean metabolic activity and standard deviation at that location for normal controls; denote these values by m(x) and $\sigma(x)$, respectively. Then the z-score for the k'th subject's total metabolic activity for the neighborhood at x is computed as

$$Z_k(x) = \frac{m(x) - A_k(x)}{\sigma(x)} \tag{5}$$

The next step is to create a binary attribute Y.

$$Y_k(x) = \begin{cases} 1 \; if \; Z_k(\bar{x}) \geq T \\ 0 \; \text{otherwise} \end{cases}, \tag{6}$$

where T is a threshold which could be either fixed or calculated automatically by the decision tree learning using information gain as a measure of optimality. We will analyze the choice of T in more detail in Section 3.

The next step is to calculate the information gain for all the attributes, Y(x) and find the one with maximum information gain. Then using a threshold such as 90% of this maximum value or larger, we can define a larger set of attributes which have near maximum information gain. Denote the set of cortical locations of near maximum information gain attributes with $X_{optimal}$. Then we define a new attribute from the union of this set of cortical locations as

$$N(x,R) = \bigcup_{x' \in X_{optimal}} S(x',R), \tag{7}$$

This union of attributes also provides a way to define arbitrarily shaped neighborhoods on the cortical surface rather than the building-block neighborhoods that we started with. The new attribute value for the new region is then computed by replacing the original neighborhood with the new arbitrarily shaped neighborhood:

$$B_k(x) = \sum_{x' \in N(x,R)} M_k(x'), \tag{8}$$

Finally, using Equations (5) and (6) we can compute the z-score for this attribute and use it for classification in the decision tree.

3 Results

Our approach was validated with a set of 48 autopsy confirmed cases of AD and FTD. Of these cases, 34 of the subjects had AD and 14 had FTD. A set of 33 normal subjects was used for computing z-scores. In this section, we will discuss the behavior of the proposed classification method and present results of validation results on the set of 48.

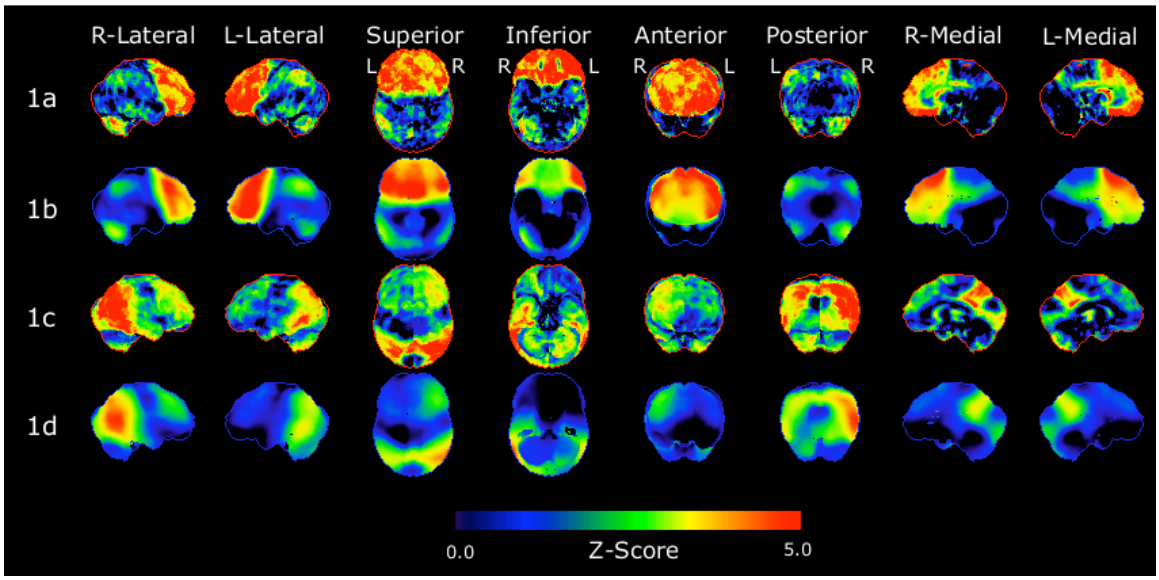

Fig. 1. [Color Image] Eight surface projections of cortical locations extracted from 3D FDG-PET images by SSP. The colormap indicates z-scores, a measure of difference from a database of normal controls. Red indicates significantly lower metabolism than normal controls while blue indicates no significant difference. a) Single pixel z-scores of a patient with FTD. b) Neighborhood z-scores of the same FTD patient. c) Single pixel z-scores of a patient with AD. d) Neighborhood z-scores of the same AD patient

Since we are working with cortical locations and their associated values extracted from the 3D FDG-PET image rather than with the 3D image directly, it is useful to observe the data in the form of surface projections. The figures in this paper present the reader with eight surface projections: right/left lateral, superior, inferior, anterior, posterior, and right/left medial,. Furthermore, we present z-scores, differences from a database of 33 normal controls computed as described in Equation (5), instead of direct metabolic values from the PET images. Figure 1a depicts single pixel z-scores of a patient with Frontotemporal dementia where frontal association cortex and anterior temporal association cortex of the brain show hypometabolism. Note that red colors indicate significant hypometabolism (lower metabolism compared to normal controls) while blue colors indicate no significant difference. Figure 1b depicts building-block neighborhood (R=8) z-scores computed with Equations (7) and (8) for the same FTD patient. These pictures demonstrate the averaging effect that is introduced by neighborhood z-scores which make the approach less susceptible to noise. To discuss this point further, consider defining arbitrarily shaped regions as unions of single pixel regions rather than the disk shaped building blocks defined in Equation (3). Due to the higher amount of noise in single pixel z-scores, their information gain measures will also be more prone to noise that would result in a non-robust neighborhood selection process. In fact, the radius of the building block disk play an important role in the success of the algorithm as will be discussed later in this section.

Figure1c and d depict single pixel and building-block neighborhood (R=8) z-scores for a patient with Alzheimer disease where temporo-parietal regions show significant reduction in glucose metabolism relative to normal control. Often the distinction between AD and FTD is not as apparent as the cases illustrated in Figure 1, some patients who are affected with Alzheimer disease also show reduction of metabolism in Frontotemoral regions of the brain, making diagnosis more difficult.

Figure 2 displays the decision tree learned using all 48 cases in the training set. This tree has three decision nodes, each associated with an

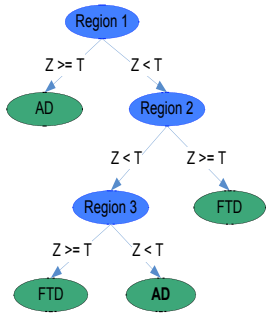

Fig. 1. Decision Tree for T=3 and R=8

arbitrarily shaped region that was found using the methods discussed in Seciton 2. These regions are shown in Figure 3. The inputs into our decision tree are binary attributes based on the thresholding of z-scores of groups of adjacent pixels. This tree was generated based on a building-block neighborhood radius of R=8 and a z-score threshold value of T=3. The significance of the selection for the R and T parameters will be discussed later in this section. In this example, at each decision node in the tree, the algorithm considers the question "Is the z-score of the given region greater than or equal to 3?" Depending on the answer the algorithm either selects a classification, or the next decision node in the tree.

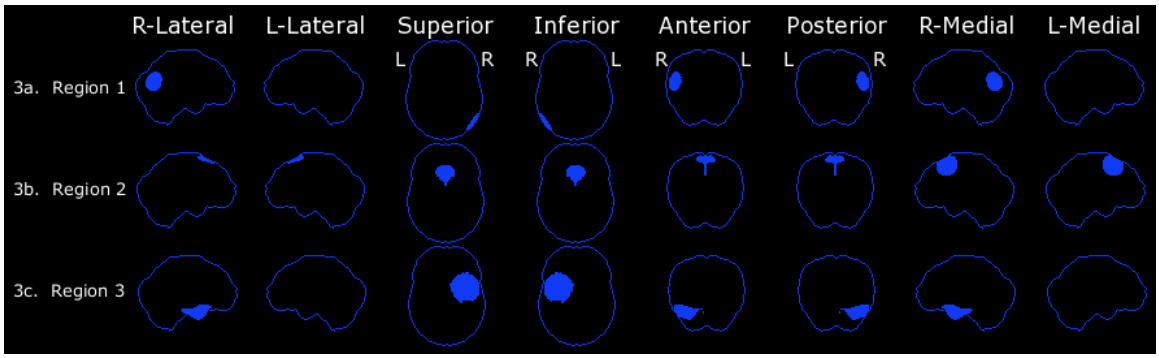

Fig. 2. Regions 1-3 of decision tree of Figure 2 shown as surface projections. a) Region 1 in Figure 2. This area corresponds to posterior temporo-parietal cortex which is greatly affected in AD. b) Region 2 in Figure 2. This corresponds to anterior temporal cortex, typically affected in FTD and not in AD. c) Region 3 in Figure 2. This area is the right anterior temporal cortex, typically affected in FTD and not in AD.

Figure 3 shows the regions that correspond to the different nodes of the decision tree using the same surface projection technique used in Figure 1. Figure 3aa corresponds to Region 1 of the decision tree. If the z-score of the this region is greater than 3, the algorithm classifies the subject as having AD, if the value is less than 3; it goes to the next node. This area corresponds to posterior temporo-parietal cortex which is greatly affected in AD. Hence, its selection by the proposed method and decision tree learning algorithm conforms to our expectations from a neurological point of view. Figure 3b displays Region 2 of the decision tree. If the z-score is greater than 3, the algorithm classifies the subject as having FTD, but if it's not, it goes to the next node. This selection is also in accordance with neurological views of FTD and AD; this region corresponds to anterior temporal, typically affected in FTD and not in AD. Finally, Region 3 of the decision tree is shown in Figure 3c. If the z-score of this selected area is more than 3, the subject is classified as having FTD, and if the z-score is less than 3, the subject is being classified as having AD. This area is the right anterior temporal, typically affected in FTD and not in AD. However , it was not immediately clear why the algorithm showed a preference for right vs. left anterior temporal regions. One explanation is the small sample size that is left after two previous branchings. This small sample size might not be able to support a robust decision in choosing the region with highest information gain resulting in over-fitting; hence a neurologically unexplained preference for right vs. left anterior temporal regions.

We validate the proposed approach using a set of 48 subjects (33AD, 14 FTD) with ground truth determined by autopsy results. We experimented with different radii for defining building-block neighborhoods and different threshold values to compare against z-scores for determining the value of binary attributes that were the input of the decision tree. The accuracy of the algorithm was assessed with a leave-one-out-cross validation approach: one of the 48 subjects is left out of the training set and the resulting classifier is tested on this left-out subject, the experiment is repeated 48 times leaving each subject out once. Table 1 summarizes our findings. We conclude that radius of eight (R=8) and threshold value of three (T=3) offer the best results in terms of accuracy. Notice that R=0 corresponds to starting from single pixels and as expected performs poorly. On the other hand, using very large R for the building blocks, such as R=16, can diminish the discriminative power of the resulting regions as seen in Table 1. Using T=2 was not selective enough in terms of difference from normal controls (very few binary attributes with value 1) while T=4 was overly selective (too many binary attributes with value 1). The threshold T=3 provided the

best trade-off. We also experimented with automatically choosing optimal T values for each node of the decision tree based on maximum information gain criteria as is typically done when using decision trees with continuous-valued attributes. However, the results obtained with this approach were worse than fixing the threshold at T=3.

Table 1. Classification accuracy results for different radii (R) and threshold values (T)

Radius Size - R	T= 2, Z>T	T=3, Z > T	T=4, Z> T
2	75%	81%	88%
4	71%	90%	92%
8	78%	94%	73%
12	67%	88%	78%
16	75%	80%	59%

These results can also be compared to previously reported experiments in the literature with the same dataset [8]. The visual rating of six neurologists using the SSP z-score images was reported to be 89%. Using principal component analysis (PCA) to project the data to a lower-dimensional space followed by learning a linear classifier in this space gave results in the range 80-85% depending on the dimensionality used in PCA. Slightly better results, as high as 90%, were obtained using Partial Least Squares, which takes into account target classifications, instead of PCA [8].

4 Conclusions

We have discussed a method for the automatic classification of brain images of AD and FTD subjects. The dynamic region selection of decision tree based on the information gain is a powerful tool, especially when compared to ROI analysis which requires prior knowledge. This method also makes it possible to locate areas of abnormalities for specific type of dementia, something that is not easily seen in PCA or in other classifiers such as neural networks. Furthermore, due to the grouping of adjacent pixels, this method is less sensitive to noise or any registration error. The results obtained using the proposed method are very encouraging in classifying FTD and AD subjects. We are planning to further confirm our results on other databases of AD and FTD subjects. In future research directions, we are planning to apply the same image-based classification paradigm to other diseases that present similar clinical symptoms. This method could be modified to correlate clinical symptoms with brain regions to determine which areas of the brain are associated with particular symptoms, and may be useful in determining a particular approach to take when treating a particular patient.

Acknowledgements. This study was partially supported by an anonymous donation to the Center for Alzheimer's Care, Imaging and Research at the University of Utah, NIH grant RO1AG22394, a pilot cooperative project grant from the National Alzheimer¹s Coordinating Center (NIH Grant AG16976) and by the Michigan Alzheimer¹s Disease Research Center (NIH Grant AG08671). Co-author Satoshi Minoshima developed and owns the copyright to the Neurostat Neurological Statistical Imaging Software used in this project. We thank Andrew P. Lieberman for reviewing the neuropathology, and David E. Kuhl, Sid Gilman, Gus Buchtel, David Knesper, R. Scott Turner, and Kirk Frey for making images from their research available for this study.

References

1. Hooper H.R., McEwan A.J., Lentle B.C., Kotchon T.L., Hooper P.M.: Interactive three-dimensional region of interest analysis of HMPAO SPECT Brain Studies, *J Nucl Med* 1990; 31:2046-2051
2. Charpentier P., Lavenu I., Defebvre L., Duhamel A., Lecouffe P., Pasquier P., Steinlin M.: Alzheimer's disease and frontotemporal dementia are differentiated by discriminant analysis applied to 99mTc HmPAO SPECT data, *J Neurol. Neurosurg. Psychiatry* 2000; 69;661-663
3. Frisoni, Testa C., Zorzan A., Sabattoli F., Beltramello A., Soininen H., Laakso M.P.: Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry, *J. Neurol. Neurosurg. Psychiatry* 2002; 73; 657-664
4. Ishii K., Kawachi T., Sasaki H., Kono A.K., Fukuda T., Kojima Y., and Mori E.: Voxel-Based Morphometric Comparison between early- and late-onset mild Alzheimer's disease and assessment of diagnostic performance of Z Score images, *J Neuroradiol* 2005; 26;333-340
5. Mitchell, T.M., Machine Learning, McGraw Hill, New York, NY, 1997
6. Minoshima S., Kirk A.F., Koeppe R.A., Foster N.L., Kuhl D.E.: A diagnostic approach in Alzheimer's disease using three-dimensional strereotactic surface projections of fluorine-18-FDG PET, J *Nucl Med* 1995;36;1238-48
7. Talairach J., Tournoux P.: Co-planar Stereotaxic Atlas of the Human Brain, Thieme Medical Publishers, New York, NY, 1988
8. Higdon et al. A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging, *Statistics in Medicine, Statist.* Med. 2004; 23:315-326