

3D Local Binary Pattern for PET Image Classification by SVM

Application to Early Alzheimer Disease Diagnosis

Christophe Montagne, Andreas Kodewitz, Vincent Vigneron, Virgile Giraud and Sylvie Lelandais
University of Evry, IBISC Laboratory, 40 Rue du Pelvoux, CE 1455, 91020 Evry Cedex, France

Keywords: Local Binary Pattern, Feature Extraction, Positron Emission Tomographic images, Alzheimer disease, Machine Learning.

Abstract: The early diagnostic of Alzheimer disease by non-invasive technique becomes a priority to improve the life of patient and his social environment by an adapted medical follow-up. This is a necessity facing the growing number of affected persons and the cost to our society caused by dementia. Computer based analysis of Fluorodeoxyglucose PET scans might become a possibility to make early diagnosis more efficient. Temporal and parietal lobes are the main location of medical findings. We have clues that in PET images these lobes contain more information about Alzheimer's disease. We used a texture operator, the Local Binary Pattern, to include prior information about the localization of changes in the human brain. We use a Support Vector machine (SVM) to classify Alzheimer's disease versus normal control group and to get better classification rates focusing on parietal and temporal lobes.

1 INTRODUCTION

The number of people affected by the Alzheimer Disease (AD) is growing. In 2010, this dementia affected 35.6 million people (Wimo and Prince, 2010). This disease affects patient himself and his social environment. The detection of the early states of Alzheimer disease allows to begin some treatments for the patient and slow down the progress of the disorder (Salmon, 2008). Most of works focused on temporal lobes, few of them on parietal lobes (Kodewitz et al., 2011). Some works on the SPECT-scans (Single Photon Emission Computed Tomography) are due to Ramirez with a computer-aided diagnosis based on a selection of image parameters (first and second order statistics) and Support Vector Machine (SVM) (Ramirez et al., 2009). Others methods with PET-scans (Positron Emission Tomography) are based on covariance analysis of voxels to classify AD versus Normal Control (NC) (Scarmeas et al., 2004). The main idea of this work is that we can automatically analyse the PET scans to detect AD patient versus NC patient using textural features. This classification can help doctors to make their early diagnostic and try to localize some brain areas where Alzheimer Disease is growing. In our study, we use 3D-scans extracted from the ADNI database (Alzheimer's Disease Neuroimaging Initiative) and we propose to use

a textural operator which is the Local Binary Pattern (LBP) (Pietikäinen and Ojala, 2000) extended to the 3D. Then this new feature was studied by an ANalysis Of VAriance (ANOVA). The analysis highlighted a link between LBP patterns and the type of patient or some brain areas. Based on this new feature, we train a machine learning algorithm SVM to classify AD versus NC PET-scans. For this step, in a first time we use as Region Of Interest (ROI) parietal and temporal lobes, and then, in a second time full PET-scans are considered.

In the first section we describe the materials used in this study. Secondly we introduce the 3D Local Binary Pattern that we developed. Then we explain the ANOVA method and the obtained results. Finally after a short presentation of SVM, we discuss on the results of the classification using the LBP characteristics and we conclude.

2 MATERIALS

In this study, PET-scan provides three-dimensional functional imaging data that measures the metabolism in the brain which are acquired by a non-invasive method, Figure 1 shows how the AD reduces the metabolism in the brain. This scan is obtained by 18F-

fluorodesoxyglucose (in short 18F-FDG) injection to a patient. This radioactive isotope allows to capture the distribution of brain activity. With this technique, we can detect the first abnormalities before structural alteration (Bennys et al., 2001).

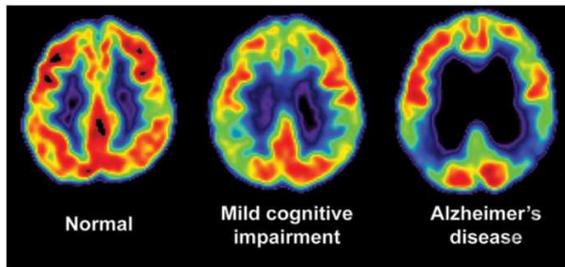


Figure 1: PET scans of AD, MCI and NC patient.

2.1 ADNI Database

The American Alzheimer's Disease Neuroimaging Initiative (ADNI, <http://www.loni.ucla.edu/ADNI/>) collects data of patients affected by AD, mild cognitive impairment (MCI) and normal control group (NC). This includes diagnosis based on mental state exams (e.g. mini mental state exam (MMSE)), biomarkers, MRI and PET scans. In this database we chose PET scans with 18F-FDG tracer from AD and NC patients. Each image contains $91 \times 109 \times 91$ voxels after normalization (the PET-scans have initially different sizes). For our purpose we extracted 166 3D-scans (one 3D-scan for one patient) which are sorted in 82 AD and 84 NC. Moreover, the AD scans are taken early in the disease to simulate a situation of early diagnosis.

2.2 Brain Atlas

Today, the definite diagnosis of AD is based on the post-mortem observation of intracellular neurofibrillary tangles (NTF): β -amyloid deposition in the form of extracellular senile plaques and blood vessel deposits, synapse dysfunction and loss. NTF deposition originates in the medial temporal lobes and then begins to cluster in the adjacent inferior temporal and posterior cingulate cortex in mild AD, and finally spreads to the parieto-temporal and prefrontal association cortices. Medical doctors use this information in their analysis of PET scans. They search for abnormal variations in the brain metabolism and do their diagnosis based on the location of these abnormalities. In order to emulate this method, we decide to use a brain atlas which gives us a brain model. With the ©Matlab toolbox WFU PickAtlas (Maldjian et al., 2003), using the Talairach daemon by Lancaster et al. (Lancaster

et al., 2000), it is even possible for nonspecialists to select a volume of interest in the brain by knowing the name of a certain area, e.g. the name of the lobe, and create an indexed mask for this volume of interest. We used these prerequisites to create a simplified map which distinguishes only three zones: temporal lobes, parietal lobes and the rest of the brain (see figure 2). Each voxel of a PET-scan matches with a voxel of this atlas because our atlas size is the same as our normalised brain PET-scans. So we have a unique model of brain for all individuals. Parietal and temporal areas are considered identical from one brain to another and adopt a location and volume average. Parietal lobes contain 26850 voxels, temporal lobes contains 33103 voxels and rest of the brain contain 119907 voxels in each PET-scan.

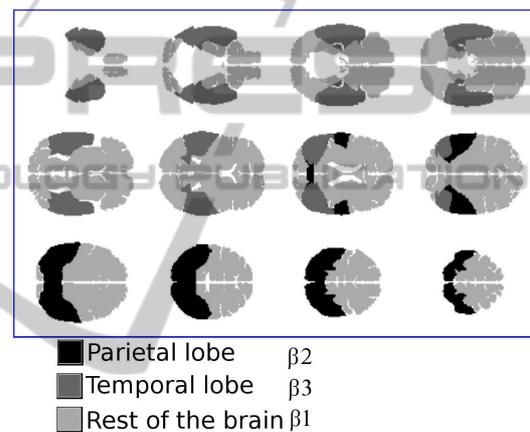


Figure 2: Brain atlas.

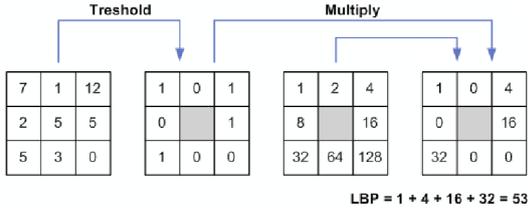
3 LOCAL BINARY PATTERN

3.1 2D Local Binary Pattern

LBP was proposed by Pietikäinen et Ojala in the 90's (Pietikäinen and Ojala, 2000). The idea is to give a pattern code to each pixel. Gray level of the central pixel g_c is compared to gray level of each neighboring pixel g_i . Then by thresholding, a value of 0 or 1 is associated to each neighboring pixel.

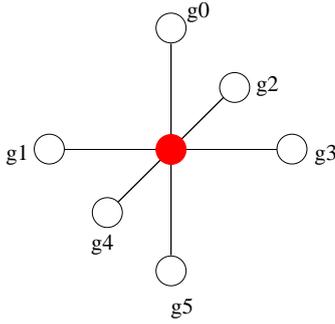
$$LBP_{P,R}^{2d} = \sum_{i=0}^{P-1} s(g_i - g_c) 2^i, \text{ with } s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else.} \end{cases} \quad (1)$$

where P is the number of neighbor pixels separated from R pixel(s) to the central pixel c . Figure 3 illustrates this feature.


 Figure 3: Example of 2D-LBP with $P = 8$ and $R = 1$.

3.2 3D Local Binary Pattern

The main idea of this work is an adaptation of the LBP to capture the 3D structure of texture. We need LBP patterns which have a rotation invariance and qualify the AD PET-scans. Some works have been performed on the extend of the Local Binary Pattern method for characterisation of 3D textures by L. Paulhac (Paulhac et al., 2008). His vision of 3D-LBP consists on a superposition of circles that represent a spherical neighborhood. Each circle can be encoded like a 2D-LBP. Our proposition consists to select 6 nearest voxels and to order them for encoding pattern (see figure 4).


 Figure 4: 3D-LBP with 6 neighbors $LBP_{6,1}^{3d}$.

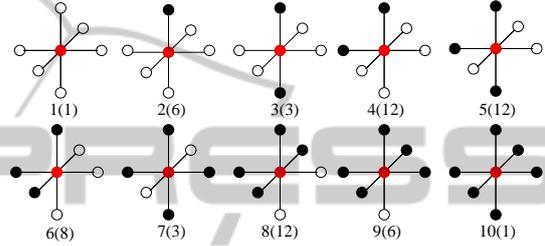
By this way, we have $2^6 = 64$ possible patterns that we merge in 10 groups according to geometrical similarities (see figure 5). Each group is filled with patterns which have the same numbers of neighbor voxels with a gray level higher than the central voxel c . The idea was to keep a rotation invariance in each group. Table 1 defines these groups with:

$$\text{card}(c) = \sum_{i=0}^{P-1} s(g_i - g_c) \quad (2)$$

where $P = 6$ is the number of voxels in neighborhood and $R = 1$ or $R = 2$ the distance between the central voxel and its neighbors. We use two values of R to capture micro and macro-structure of the texture with our 3D-LBP. In equation 2, $\text{card}(c)$ gives the number of neighbors with an higher gray level than the central voxel.

Table 1: Definition of the 10 groups of patterns rotationally invariant.

$LBP_{P,R}^{3d}$	$\text{card}(c)$	condition
1	0	
2	1	
3	2	opposite voxels
4	2	bend voxels
5	3	voxels on the same plane
6	3	voxels on different planes
7	4	voxels on the same plane
8	4	voxels on different planes
9	5	
10	6	


 Figure 5: Merging of $2^6 = 64$ patterns in 10 groups. The integer in brackets indicates the number of different patterns for each group.

In our experimental protocol, the count of these patterns in ROI of PET-scans or in full PET-scans will be the input variable of our machine learning algorithm.

3.3 ANOVA

The 3D-LBP features may be an important character for the classification of Alzheimer PET-scan. We perform an analysis of variance to find the link between a variable of interest (the count of 3D-LBP features) and some response variables (type of patient, brain area, type of pattern) which are also called *factors* and its relevance.

Suppose that the variable to explain Y follows a normal distribution in each population. y_{ij} was a realization of $Y \sim \mathcal{N}(\mu_i, \sigma)$. There are two hypothesis: H_0 -Null hypothesis, every data follows a normal distribution; H_1 -At least one sample follows an other normal distribution.

There are 3 factors:

- α_u : type of patient (AD/NC),
- β_v : brain area (temporal lobe β_3 , parietal lobe β_2 , rest of brain β_1 , full brain β_0) (see figure 2),
- γ_w : type of pattern (1... 10) (see figure 5).

We search the influence of a factor set ($\alpha_u, \beta_v, \gamma_w$) on the variable of interest Y . This method can be ap-

plied on each factor or on all together. For example, Eqs. 3 and 4 show a case with a single factor.

$$y_{ik} = \mu + m_{\alpha_i} + \varepsilon_{ik} \quad (3)$$

where μ is an offset, m_{α_i} the theoretical mean of the sample i and ε_{ik} are Gaussian i.i.d error terms ($\mathcal{N}(0, \sigma^2)$).

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} m_{\alpha_1} \\ m_{\alpha_2} \\ m_{\alpha_3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \end{pmatrix} \quad (4)$$

where m is the theoretical mean of the sample. This equation shows the variance decomposition:

$$\sum_{i=1}^3 \sum_{k=1}^{n_i} (X_{ik} - m)^2 = \sum_{i=1}^3 n_i (m_i - m)^2 + \sum_{i=1}^3 \sum_{k=1}^{n_i} \varepsilon_{ik}^2 \quad (5)$$

$$SC_T^2 = SC_A^2 + S_R^2$$

m_{α_i} and m are estimated by: $\bar{y}_i = \sum_{k=1}^{n_i} y_{ik}$ and $\bar{y} = \sum_{i=1}^3 \sum_{k=1}^{n_i} y_{ik}$. Under H_0 , we have:

$$\frac{SC_T}{\sigma^2} \sim \chi_{n-1}^2, \quad \frac{SC_A}{\sigma^2} \sim \chi_{3-1}^2, \quad \frac{S_R}{\sigma^2} \sim \chi_{n-3}^2 \quad (6)$$

and:

$$T = \frac{SC_A^2 / (3-1)}{S_R^2 / (n-3)} \sim F_{p-1, n-p}. \quad (7)$$

Under H_0 , the statistics T follows a Fisher distribution with $(3-1, n-3)$ degrees of freedom. The critical value is $T > F_{p-1, n-p, 1-\alpha}$, where α settles the confidence level with which one can reject the hypothesis H_0 (by convention often chosen to be 95%).

Table 2 exposes a part of the results. The p -value of each group of pattern on each area of brain is displayed.

The p -value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the H_0 hypothesis is true. The lower is the p -value, the higher is the probability of H_1 to be true. We choose a p -value threshold to 0.005: we accept H_1 is true with 99.5% or more (bold p -values in the table). We can see p -values corresponding to the full brain (β_0) or rest of the brain (β_1) are higher than our threshold: the count of the 3D-LBP patterns is not informative in these areas. In the parietal and temporal lobes (β_2, β_3), we notice the

Table 2: p -values for the different brain areas.

LBP Group number	Pr(> F)			
	β_0	β_1	β_2	β_3
1	2.13e-1	1.42e-1	5.83e-5	2.63e-2
2	5.47e-1	3.14e-1	1.66e-5	1.86e-4
3	2.71e-1	6.38e-1	3.14e-4	2.51e-7
4	7.89e-1	1.60e-1	1.76e-3	1.07e-2
5	9.77e-1	2.24e-1	8.60e-5	4.69e-3
6	7.61e-1	4.20e-1	1.56e-3	5.72e-3
7	9.22e-1	3.38e-1	7.40e-3	1.39e-1
8	7.62e-1	4.13e-1	2.50e-3	1.79e-3
9	6.00e-1	2.12e-2	3.22e-3	3.69e-1
10	3.49e-1	2.54e-2	1.86e-1	6.43e-2

power of some LBP patterns for discriminating the type of patient α_i .

These observations allow to use machine learning to create a classifier with the 3D LBP features extracted from the PET-scan.

4 CLASSIFICATION BY SUPPORT VECTORS MACHINE

The SVM was introduced by Vapnik (Vapnik, 1998). We used a description of the SVM which is based on Schölkopf and Smola (Schölkopf and Smola, 2002) works. Consider a training data set consisting of two separable classes in a n -dimensional feature space. This means each class forms its own cluster and those clusters do not intersect. From this follows that there exists at least one hyperplane able to separate the two classes with one class on each side of the hyperplane.

We use a Gaussian kernel in SVM, $\gamma \geq 0$, $k(x, x') = e^{(-\|x-x'\|^2/\gamma^2)}$. In figure 6 we can see the contour plot of the error landscape resulting from a grid search on hyperparameter cost C and gamma γ . C is a cost constant for the Lagrangian. This grid is obtained with the function *tune* in R. The red circle shows the area with the best parameters.

PET-scans used in this study come from the ADNI database. Each scan contains $91 \times 109 \times 91$ voxels. We selected 82 patients from AD group and 84 from NC group (one patient = one PET-scan). We try to classify AD versus NC. Each voxel in the image is encoded by $LBP_{6,1}^{3d}$ and $LBP_{6,2}^{3d}$. We have 10, 20 or 30 input variables if we use 1, 2 or 3 areas of brain (parietal β_2 , temporal β_3 , rest of the brain β_1) in the classifier. With full brain (β_0) we have 10 variables. And when we use the $LBP_{6,1}^{3d}$ and $LBP_{6,2}^{3d}$, the number of variables is doubled.

These variables constitute our input vector for SVM. In table 3 we see the classification rate for the 3 brain areas or for parietal and temporal lobes. Other

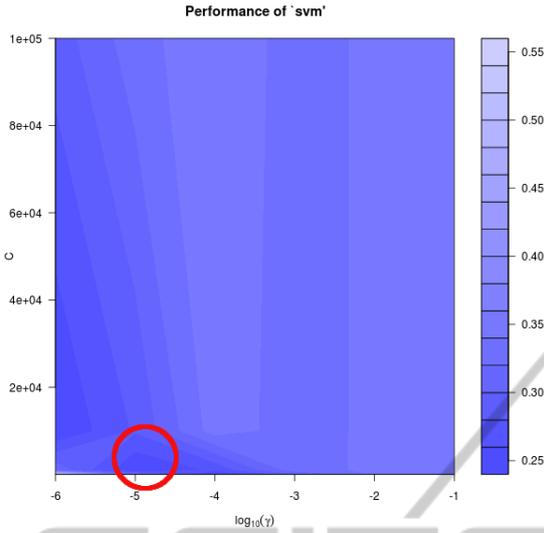


Figure 6: Visualisation of the tuning parameters C and γ .

results with only one brain area (β_1 , β_2 or β_3) or full brain (β_0) don't appear in this table because of lower results. The cross validation consists in a number of runs equal to *size of the base* of a SVM. For each run we divide the main database with a part used for training and the other for testing. The mean of all these runs give us the rate of cross validation.

Table 3: Precision, recall and cross validation (CV) for SVM classification - ①: $LBP_{6,1}^{3d}$, ②: $LBP_{6,2}^{3d}$. $\beta_{1,2,3}$ is put for the use of all brain areas; $\beta_{2,3}$ is for temporal and parietal lobes.

LBP type	Brain areas	Precision		Recall		CV
		AD	NC	AD	NC	
①	β_1	68%	70%	70%	68%	60%
	β_2	67%	70%	70%	67%	68%
	β_3	75%	71%	67%	79%	64%
	$\beta_{1,2,3}$	89%	81%	79%	90%	75%
	$\beta_{2,3}$	85%	79%	76%	86%	66%
②	β_1	64%	67%	68%	63%	56%
	β_2	73%	70%	66%	76%	70%
	β_3	65%	71%	68%	76%	66%
	$\beta_{1,2,3}$	87%	79%	76%	89%	75%
	$\beta_{2,3}$	82%	82%	81%	83%	68%
① & ②	β_1	74%	77%	77%	74%	63%
	β_2	74%	70%	67%	77%	69%
	β_3	77%	72%	67%	81%	67%
	$\beta_{1,2,3}$	87%	79%	76%	89%	77%
	$\beta_{2,3}$	82%	82%	81%	83%	69%

Table 3 shows results obtained with different variables of $LBP_{6,1}^{3d}$ and $LBP_{6,2}^{3d}$. We can notice that all the results are very close together. But when we use the 3 brain areas, the SVM classifier reaches better rates than just using parietal and temporal lobes. The best

CV rate is 77% with $LBP_{6,1}^{3d}$ & $LBP_{6,2}^{3d}$. We can assume that this value is relatively low because the scans correspond to the beginning of the disease when the diagnosis is more difficult than for a patient severely affected.

5 CONCLUSIONS

We propose a new method to identify patients affected by the Alzheimer Disease from their PET-scan. With the ANOVA we have shown LBP are important features to class AD *versus* NC PET-scans. Then we train and test a classifier to create an automatic computer-aided diagnosis for the Alzheimer Disease using the LBP characteristics extracted from the PET-scan. We reach to a score of 77% by using SVM method with $LBP_{6,1}^{3d}$ and $LBP_{6,2}^{3d}$ patterns.

Currently we improve this work by slicing in small cubes, in each of these block we count the LBP patterns. Then these counts are used as variables for a Random Forest classification. RF is adapted to compute and to have a good classification rate with this high number of features.

An other way to improve the results can be to develop some new LBP types to capture more accurately the texture of brain in the PET-scans. We can notice that some regions that are outside temporal and parietal lobe contain also a certain amount of discriminating information which is captured with our approach. In this case we can extend the analysed brain areas to find other ROI. A brain mapping of each PET-scan can be used to define more cleverly position of the temporal and parietal lobes, or some other subparts. The idea is to customize an atlas for each patient rather than our standard atlas for all the database. This way is already used on IRM-scans. With this, we have to segment brain to find the edges of the different areas. In the same way, merge IRM and PET data will be an interesting search.

ACKNOWLEDGEMENTS

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf).

REFERENCES

- Bennys, K., Rondouin, G., Vergnes, C., and Touchon, J. (2001). Diagnostic value of quantitative EEG in alzheimer's disease. *Clinical Neurophysiology*, 31(3):153–160.
- Kodewitz, A., V., V., Montagne, C., and Lelandais, S. (2011). Where to search for alzheimer's disease related changes in pet scans? RITS, Rennes, France.
- Lancaster, L., Woldorff, M. G., and Parsons, L. M. (2000). Automated talairach atlas labels for functional brain mapping. *Human Brain Mapping*, 10:120–131.
- Maldjian, J. A., Laurienti, P. J., Burdette, J. H., and Kraft, R. A. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets. *NeuroImage*, 19:1233–1239.
- Paulhac, L., Makris, P., and Ramel, J. (2008). Comparison between 2d and 3d local binary pattern methods for characterisation of three-dimensional textures. In Series, B., editor, *Lecture Notes in Computer Science*, volume 5112/2008, pages 670–679, Póvoa de Varzim, Portugal. ICIAR.
- Pietikäinen, M. and Ojala, T. (2000). rotation-invariant texture classification using feature distributions. *Pattern Recognition*, 33:43–52.
- Ramírez, J., Górriz, J., and et al., D. S.-G. (2009). Computer-aided diagnosis of alzheimer's type dementia combining support vector machines and discriminant set of features. *Information Sciences*.
- Salmon, E. (2008). Différentes facettes de la maladie de type azheimer. *Rev Med Liege*, 63(5-6):299–302.
- Scarmeas, N., Habeck, C. G., and et al., E. Z. (2004). Covariance pet patterns in early alzheimer's disease and subjects with cognitive impairment but no dementia: utility in group discrimination and correlations with functional performance. *NeuroImage*, 23(1):35 – 45.
- Schölkopf, B. and Smola, A. (2002). *Learning with Kernels - Support Vector Machines, Regularization, and Beyond*. The MIT Press.
- Vapnik, V. (1998). *Statistical learning theory*. Wiley.
- Wimo, A. and Prince, M. (2010). World alzheimer report 2010. <http://www.alz.co.uk/research/world-report>.