NMF-Based Analysis of SPECT Brain Images for the Diagnosis of Alzheimer's Disease

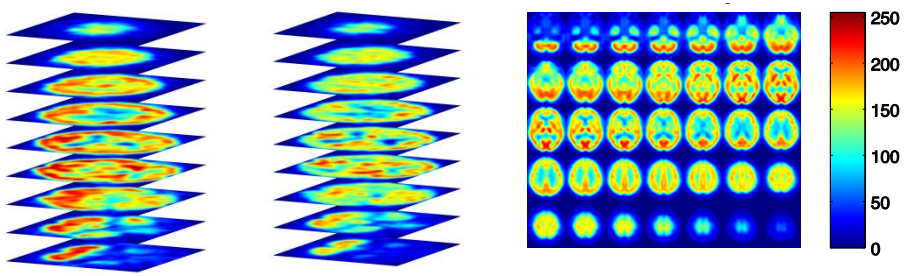
P. Padilla, J. M. Górriz, J. Ramírez,E. W. Lang, R. Chaves, F. Segovia, I. Álvarez, D. Salas-González, and M. López

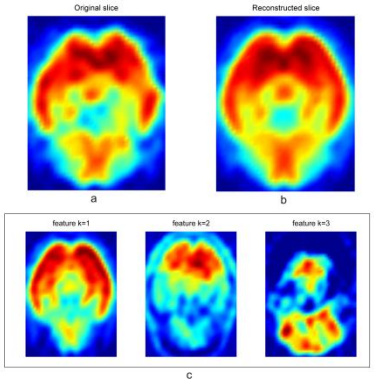
Exploratory Matrix Factorization for PET Image Analysis

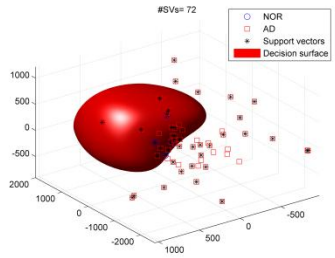
A. Kodewitz, I. R. Keck, A. M. Tomé, J. M. Górriz, and E. W. Lang

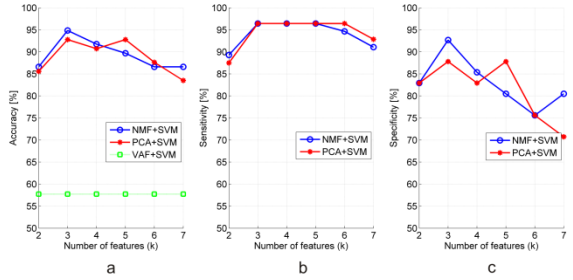
Outline

- Introduction
  - SPECT imaging
  - SPECT dataset
- Feature extraction
  - Non-negative Matrix Factorization
  - Voxel Selection: FDR
- Image classification
  - Classification task
  - Support Vector Machines
- Experiments (SPECT and PET)
  - Train and Test validation
  - CAD tool performance.
- Conclusions









Functional Brain imaging: SPECT

- SPECT images (Single photon emission computed tomography)
- Functional brain image: regional cerebral blood flow (rCBF)

SPECT Image acquisition:

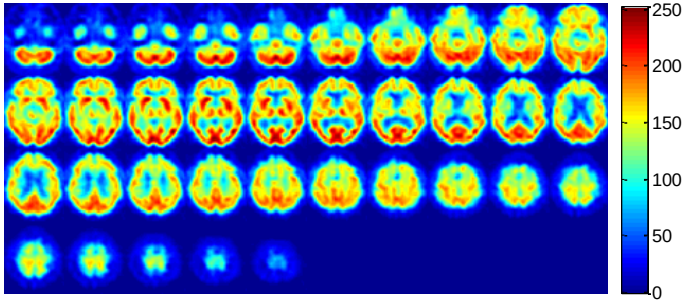
- Radionucleid Radiopharmaceutical: technetium-99m labeled ethyl cysteinate dimer (99mTc-ECD) radiopharmaceutical
- gamma camera to acquire multiple 2D images (projections): 3-head gamma camera Picker Prism 3000
- tomographic reconstruction of the multiple projections (filtered backprojection (FBP) algorithm) : 3D SPECT dataset.
- Lower resolution (4–8 mm), compared to other brain imaging techniques (fMRI, PET).

SPECT Dataset

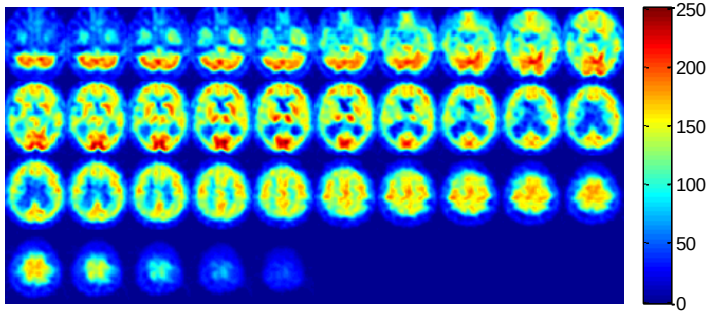
DATASET:

- 97 SPECT brain images, patients labelled by experts: 41 healthy patients without symptoms (NOR) and 56 Alzheimer's patients (AD).
- Collected from the "Virgen de las Nieves" hospital in Granada (Spain).
- 3D matrix of intensity values:
  - Initially 95x69x79: 517845 voxels.
  - Dependency on the shape of brains and orientation: Spatial normalization of the 3D matrices, in order to ensure that a given voxel in different images refers to the same brain position.
  - Intensity dependency on the dosis of radiopharmaceutical: intensities are normalized to the maximum intensity value in each 3D matrix.

NOR patient



AD patient



Non-Negative Matrix Factorization (NMF)

•Different techniques to find reduced linear representations of a dataset by means of space transformations: Principal Component Analysis (PCA), Independent Component Analysis (ICA), etc.

• **For non-negative datasets** there is one suitable technique: **Non-negative Matrix Factorization (NMF):**

Non-negative Matrix Factorization (NMF):

Given a non-negative data matrix **A**, NMF finds an approximate factorization of matrix **A**≈**WH**, by means of non-negative matrices **W** and **H**, where:

A=[A_1, A_2,…,A_M] (*N x M*)
W=[W_1,W_2,…,W_K] (*N × K*)
H=[H_1,H_2,…,H_M] (*K x M*)

$$A_{nm} = \sum_{k=1}^{K} W_{nk} H_{km}$$

M→ number of observations (profiles)
N → number of variables in each profile
K → number of features in the new space

After NMF factorization, the data contained in **H** (*K x M* elements) can be considered as a projection of the original database **A** into the reduced space spanned by the new NMF basis **W** (with *K* vectors), with lower number of features (*K*) than the original database **A**.

Non-Negative Matrix Factorization

•**Factorization rule:** Given the data matrix **A**, the optimal choice of matrices **W** and **H** are defined to be those nonnegative matrices that minimize the reconstruction error between **A** and **WH**.

•A variety of error functions (Err) have been proposed:

$$Err_1 = \frac{1}{NM}\|A - WH\|^2 = \frac{1}{NM}\sum_{nm}(A_{nm} - (WH)_{nm})^2$$ → *Frobenius Norm*

$$Err_2 = D(A||WH) = \sum_{nm}(A_{nm} log \frac{A_{nm}}{(WH)_{nm}} - A_{nm} + (WH)_{nm})$$ → *Kullback-Leibler divergence*

Different NMF algorithms: with multiplicative update rule, with additive update rule, alternating least squares, etc. In this case due to its fast convergence alternating least squares algorithm (ALS) is developed:

1. initialize W as a random positive matrix
2. solve H with previous W and original A in matrix equation $W^TWH=W^TA$
   set negative elements of H to 0
3. solve W with previous H and original A in matrix equation $HH^TW^T=HA^T$
   set negative elements of W to 0.
4. compute error function, iterate algorithm till convergence, and end factorization.

Non-Negative Matrix Factorization

SPECT Dataset: 3D matrices (**S**) for each patient (*m*): **S**(*m, x, y, z*). Change of dimensions needed for NMF Transformation: unfold 3D matrices into 1D vectors for each patient.

$\mathbf{S}(m, x, y, z) \rightarrow \mathbf{S}_{unfold}(m, n)$ $\quad n=[1,..,N],\ N=X \cdot Y \cdot Z$

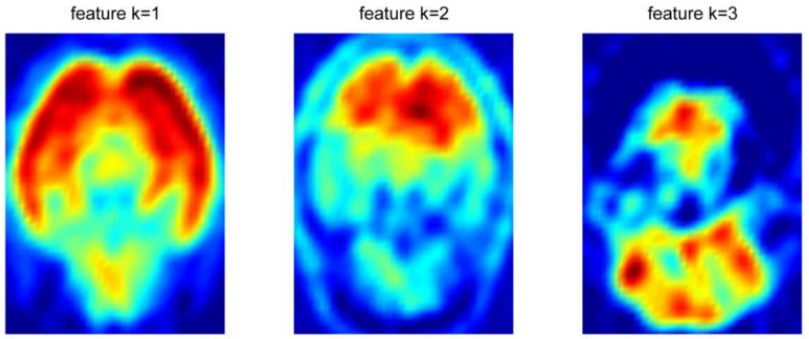
$m=[1,..,M]$, $M=97$, number of observations (patients)

$\mathbf{A}\ (N \times M) = [\mathbf{S}_{unfold}(1,1{:}n)^T, \ldots, \mathbf{S}_{unfold}(M,1{:}n)^T] \rightarrow \mathbf{A} \approx \mathbf{WH}$

- $\mathbf{H}^T\ (M \times K)$, 97 observations with *K features.*
- $\mathbf{W}^T\ (K \times N)$, *K eigenvectors of NMF basis.*

The *k eigenvectors of NMF basis can be reorganized in 3D matrices* (*'eigenfaces'*).

$\mathbf{W}^T\ (K \times N) \rightarrow \mathbf{W}_{3D}(K \times X \times Y \times Z)$



NMF 'eigenfaces' for one *z* slice

Fisher Discriminant Ratio

Each SPECT image prepared for feature extraction and classification contains *48x40x35 = 67200* voxels (downsampled by a factor of 0.5). This implies lower accuracy and but lower computational cost for the NMF algorithm.

Two reasons for reducing the number of voxels for further NMF transformation:

- Selection of the most discriminant voxels in terms of class separability: **Fisher Discriminant Ratio (FDR).**
- Lower number of voxels for subsequent transformation tasks.

Fisher Discriminant Ratio (FDR):

The FDR criterium is a T-test for the selection of the most representative variables of a set of observations (in this case voxels of the different patient's set of SPECT images). FDR ranks the 'voxel importance' according to the mean and variance of each class (AD or NOR).

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

μ_i → mean of class i.
σ_i → standard deviation of class i.

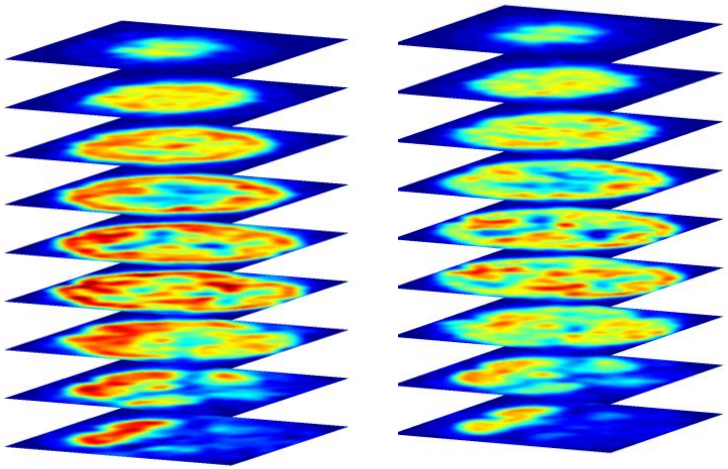
Classification task

- Two different classes: healthy patients (NOR) and Alzheimer's dementia patients (AD).

The goal of the classification task is to separate a set of binary labelled training data consisting of, *N*-dimensional patterns x_i and class labels y_i:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l) \in (R^N \times \{NOR, AD\})$$

This classifier will be used to classify new samples (x, y).



NOR patient AD patient

- In order to evaluate the CAD tool, the success rate (Acc), sensitivity (Sens) and specificity (Spec) are obtained, the latter defined as:

$$Sensitivity(Sens) = \frac{TP}{TP + FN}$$

TP : number of true positives (AD patients properly classified).

FN: number of false negatives (AD classified as NOR).

$$Specificity(Spec) = \frac{TN}{TN + FP}$$

TN: number of true negatives (NOR patients properly classified).

FP: number of false positives (NOR classified as AD).

Support Vector Machines

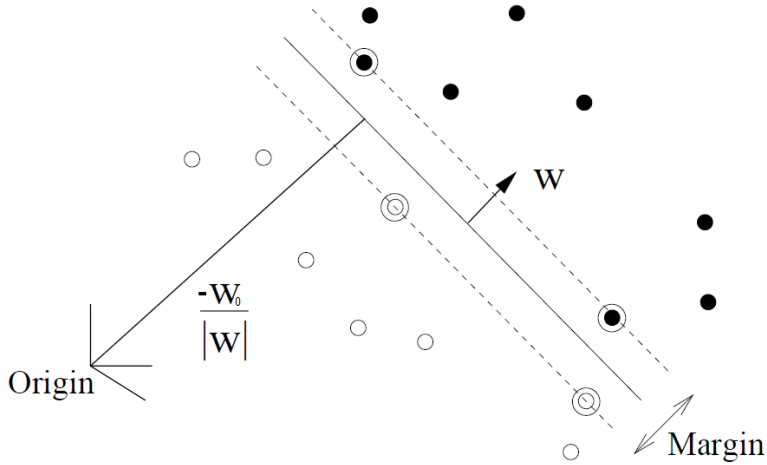
The objective is to build a function *f* with the training data, able to properly classify new unclassified data. $f: R^N \rightarrow \{\pm 1\}$

SVMs define decision hyperplanes in a multidimensional feature space. The separation hiperplane is defined by means of the so-called Support Vectors (SV).

- Linear SVMs,

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0 = 0$$

where w is the weight vector and *w0 is the threshold. w is orthogonal to the* decision hyperplane.



- Kernel SVMs,

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i \Phi(s_i) \cdot \Phi(x) + w_0 = \sum_{i=1}^{N} \alpha_i y_i K(s_i, x) + w_0$$

  - Quadratic
  - Polynomial
  - Radial basis function (RBF)
  - Sigmoid
  - …

Experiments: CAD tool

• The performance of the CAD tool is tested on the set of 97 real SPECT images (41 NOR and 56 AD), by means of a cross-validation strategy:

Cross-validation strategy: Leave-one-out train and test approach:

- Observation x_n *(for n=1,…,97)* is extracted from the database. The resulting *N-1=96* SPECT image database is used for training the classifier.
- Observation x_n is entered into the classifier for testing.

Procedure: in each *n*-evaluation of the *N*=97 train and test evaluations:

Training set:

-FDR to select the most discriminant voxels of the 96 training samples (selection of the proper threshold).

-NMF for data reduction into *k* variables (selection of proper *k* for optimum data approximation).

- SVM works on this training set.

Test patient:

-Selection of the same voxels than in the training step and NMF transformation.

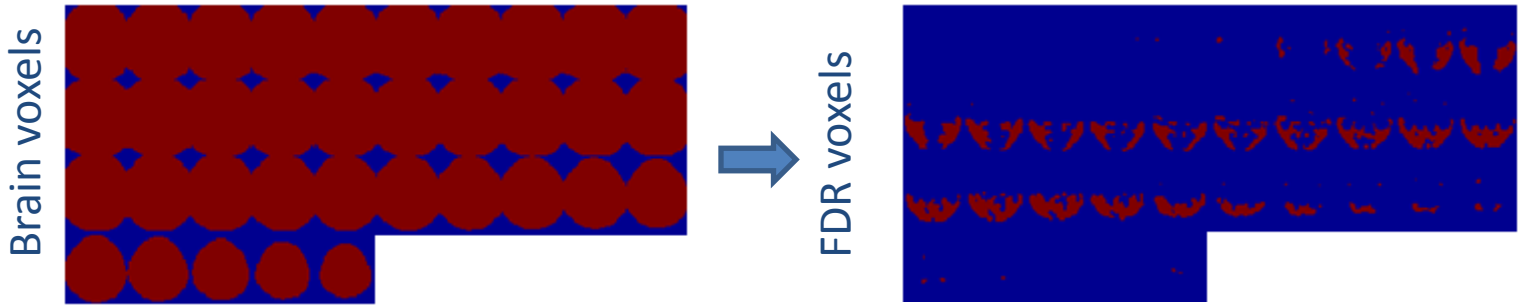
-Classification with the trained SVM.

Experiments: Training step (I)

For each Train and Test group

1-FDR to select the most discriminant voxels:

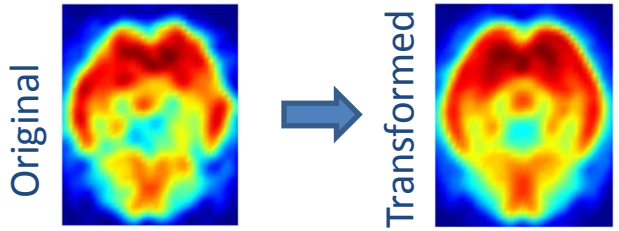
- Study and selection of proper threshold: ε=0.2*max(FDR)
- Reduction of the number of voxels: *67200* voxels to 5978 voxels

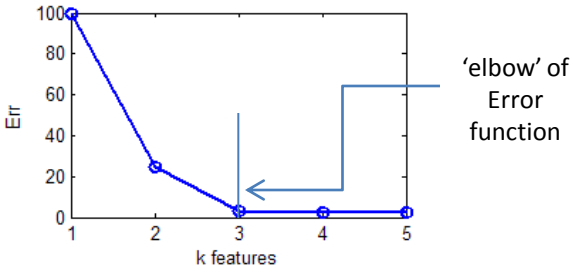


2- NMF for data reduction:

- Study and selection of proper *k* value ('elbow' of error function): *k*=3
- Reduction of the number of variables: 5978 voxels to *k*=3 NMF variables.

Example of NMF slice transformation:



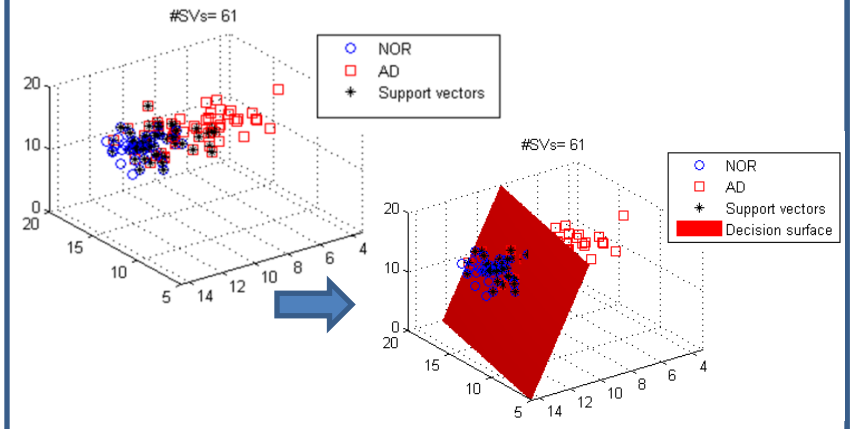


Experiments: Training step (II)

3-SVM classifier:

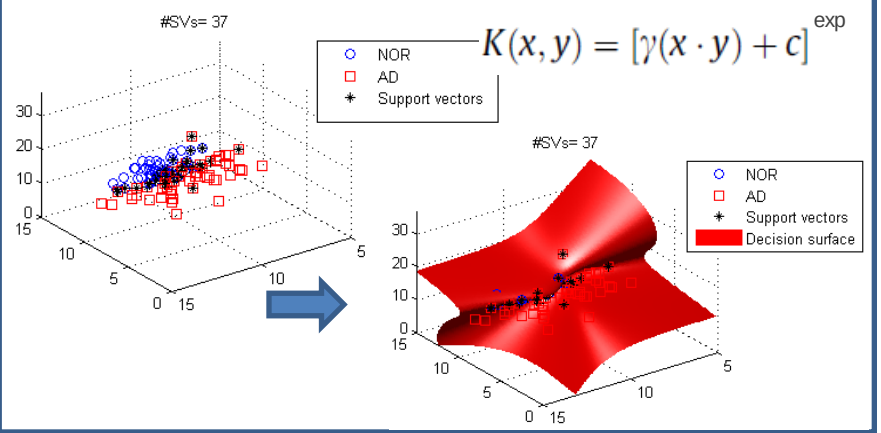
For each Train and Test group

a) linear:



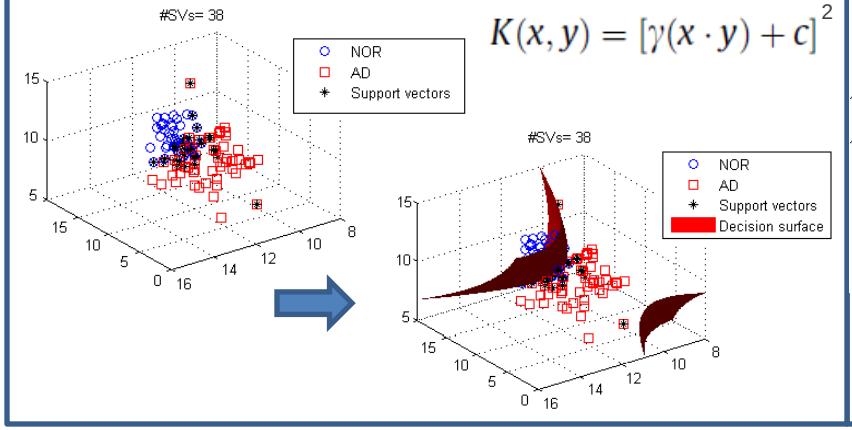
c) polynomial kernel (exp=3):

$$K(x,y) = [\gamma(x \cdot y) + c]^{exp}$$



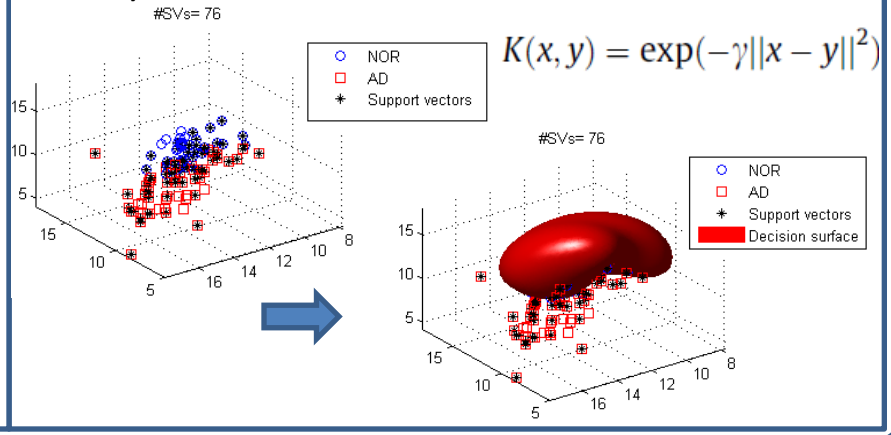
b) quadratic kernel

$$K(x,y) = [\gamma(x \cdot y) + c]^{2}$$



d) RBF kernel:

$$K(x,y) = \exp(-\gamma||x-y||^2)$$



Experiments: Test step

Test step:

1) Feature extraction and NMF transformation of the *n* test sample

2) SVM classification:

Accuracy (Acc), Sensitivity (Sens) and Specificity (Spec) computed from the 97 folds of the Leave-one-out cross validation.

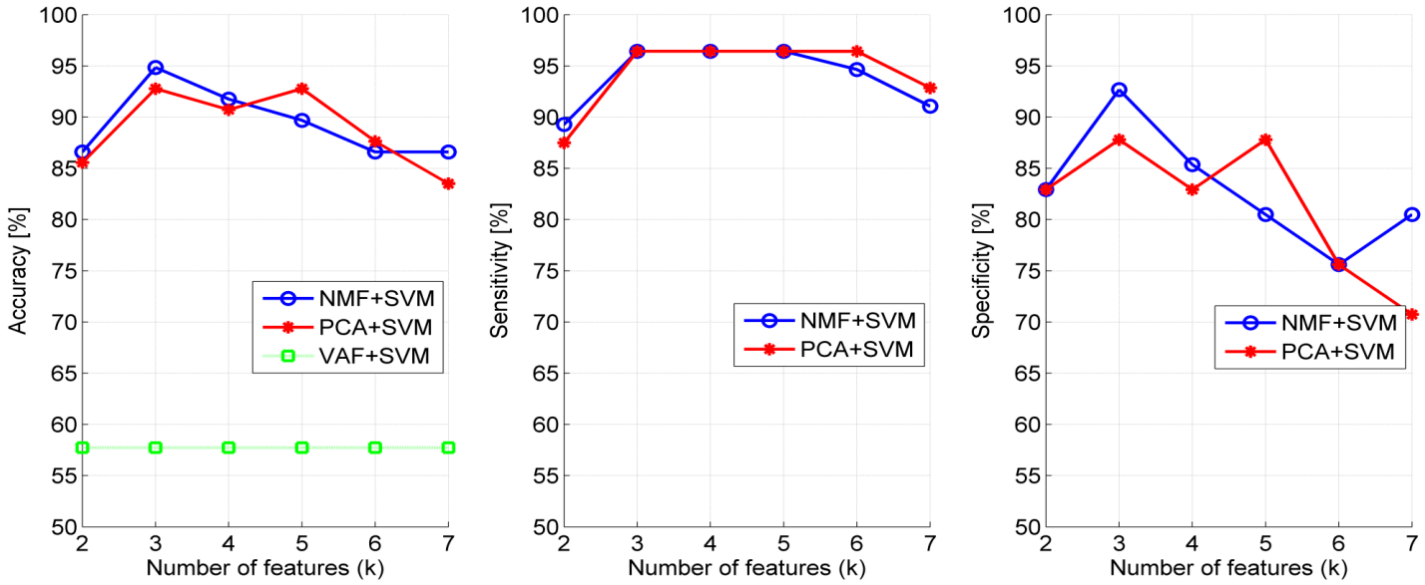
k		2	3	4	5	6	7
Linear (%)	Acc	86,60	90,72	86,60	79,38	85,57	83,51
	Sens	89,29	94,64	89,29	78,57	87,50	85,71
	Spec	82,93	85,37	82,93	80,49	82,93	80,49
Quadratic (%)	Acc	85,57	84,54	83,51	83,51	89,69	80,41
	Sens	85,71	87,50	85,71	85,71	87,50	80,36
	Spec	85,37	80,49	80,49	80,49	92,68	80,49
RBF (%)	Acc	86,60	94,85	91,75	89,69	86,60	86,60
	Sens	89,29	96,43	96,43	96,43	94,65	91,07
	Spec	82,93	92,68	85,37	80,49	75,61	80,49
Polynomial (%)	Acc	82,47	85,57	86,60	86,60	80,41	81,44
	Sens	83,93	85,71	85,71	83,93	73,21	75,00
	Spec	80,49	85,37	87,80	90,24	90,24	90,24

RBF kernel based SVM offers the best results in classification.

Experiments: CAD tool performance

Comparison with other feature extraction techniques (with RBF kernel SVM):

- Principal Component Analysis (PCA).
- baseline Voxel as Features (VAF) approach.



Best results in classification for *k*=3, peak values of of Acc= 94.9%, Sens= 96.4% and Spec= 92.8%

Experiments: PET images

- ADNI database
  - 18F-FDG PET data from 219 Alzheimer disease neuroimaging initiative (ADNI) participants, acquired with Siemens, General Electric and Philips PET scanners.

Table 1. Results of a slice-wise NMF feature extraction and subsequent *left*: SVM or *right*: RF three class feature classification.

actual	predicted by SVM			predicted by RF		
	NOR	MCI	DA	NOR	MCI	DA
NOR	74.4	20.3	5.3	74.1	19.4	6.5
MCI	16.7	64.4	18.9	19.5	60.1	20.4
DA	3.3	22.2	74.5	5.9	23.5	70.6
ACR	$(71 \pm 6)\%$			$(68 \pm 6)\%$		

The number of extracted features has been 12 for both, binary and three class classification. As expected, the classification rate is considerably higher when classifying NOR against DA than when classifying MCI against DA or Nor against MCI.

Experiments: PET images

- ADNI database
  - 18F-FDG PET data from 219 Alzheimer disease neuroimaging initiative (ADNI) participants, acquired with Siemens, General Electric and Philips PET scanners.

Table 2. Results of a slice-wise NMF feature extraction and subsequent *top*: SVM or *bottom*: RF two class feature classification. Average classification rates (ACR) are given for all two class classification with either a support vector machine (SVM) or a random forest (RF) classifier.

actual	predicted by SVM									predicted by RF								
	NOR	MCI	DA	NOR	MCI	DA	NOR	MCI	DA	NOR	MCI	DA	NOR	MCI	DA	NOR	MCI	DA
NOR	91.3	-	8.7	74.7	25.3	-	-	-	-	89.4	-	10.6	73.2	26.8	-	-	-	-
MCI	-	-	-	18.6	81.4	-	-	76.3	23.7	-	-	-	27.3	78.4	-	-	76.7	23.3
DA	6.2	-	93.8	-	-	-	-	20.2	81.4	13.8	-	86.2	-	-	-	-	27.3	72.7
ACR	$(92 \pm 4)\%$			$(78 \pm 7)\%$			$(78 \pm 5)\%$			$(88 \pm 5)\%$			$75 \pm 6\%$			$75 \pm 6\%$		

All classification results are based on training sets of 60 images per class and testing sets of 25 images per class, drawn from all available whole brain scans.

Conclusions

• A computer aided diagnosis (CAD) tool for the diagnosis of Alzheimer's disease (AD) has been presented.

• NMF transformation was used to transform the *space of voxels* into a new reduced space of *k* features. In order to reduce the initial number of voxels the Fisher Discriminant Ratio (FDR) is applied.

• SVM is selected for classification task, with different kernel functions.

• CAD tool validated with a SPECT dataset. Leave-one-out cross validation scheme is used for the evaluation of the tool performance.

• Quite promising outcomes of the CAD tool in terms of Success rate, Sensibility and Specificity. Peak values of Acc= 94.9%, Sens= 96.4% and Spec= 92.8%

Thank you very much for your attention