Voxel-MARS and CMARS: Methods for Early Detection of Alzheimer's Disease by Classification of Structural Brain MRI

EURO2018

Gerhard-Wilhelm Weber,
Poznan University of Technology
Alper Çevik,
Middle East Technical University

PRESENTATION OUTLINE

Introduction
- Motivation
- Objectives

Background
- Biomarkers
- MARS/CMARS

Methodology
- Data
- Methods

Performance
- Results
- Discussion

Conclusions
- Conclusions
- Future Outlook

MOTIVATION
- Definitions

- **Alzheimer's Disease (AD)**

- Most common **neurodegenerative** disease
- Most common cause of **dementia**
- Causes problems with **memory**, **thinking** and **behavior**, interferes with **daily tasks,** eventually leads to **death**
- Prevalence gets higher due to **increasing life expectancy**

- **Mild Cognitive Impairment (MCI)**

- Cognitive impairments **beyond the expected ones** due to age and education
- Increasing risk of **converting to AD** (at a rate of approximately 10% to 15% per year according to Petersen et al., 2004)

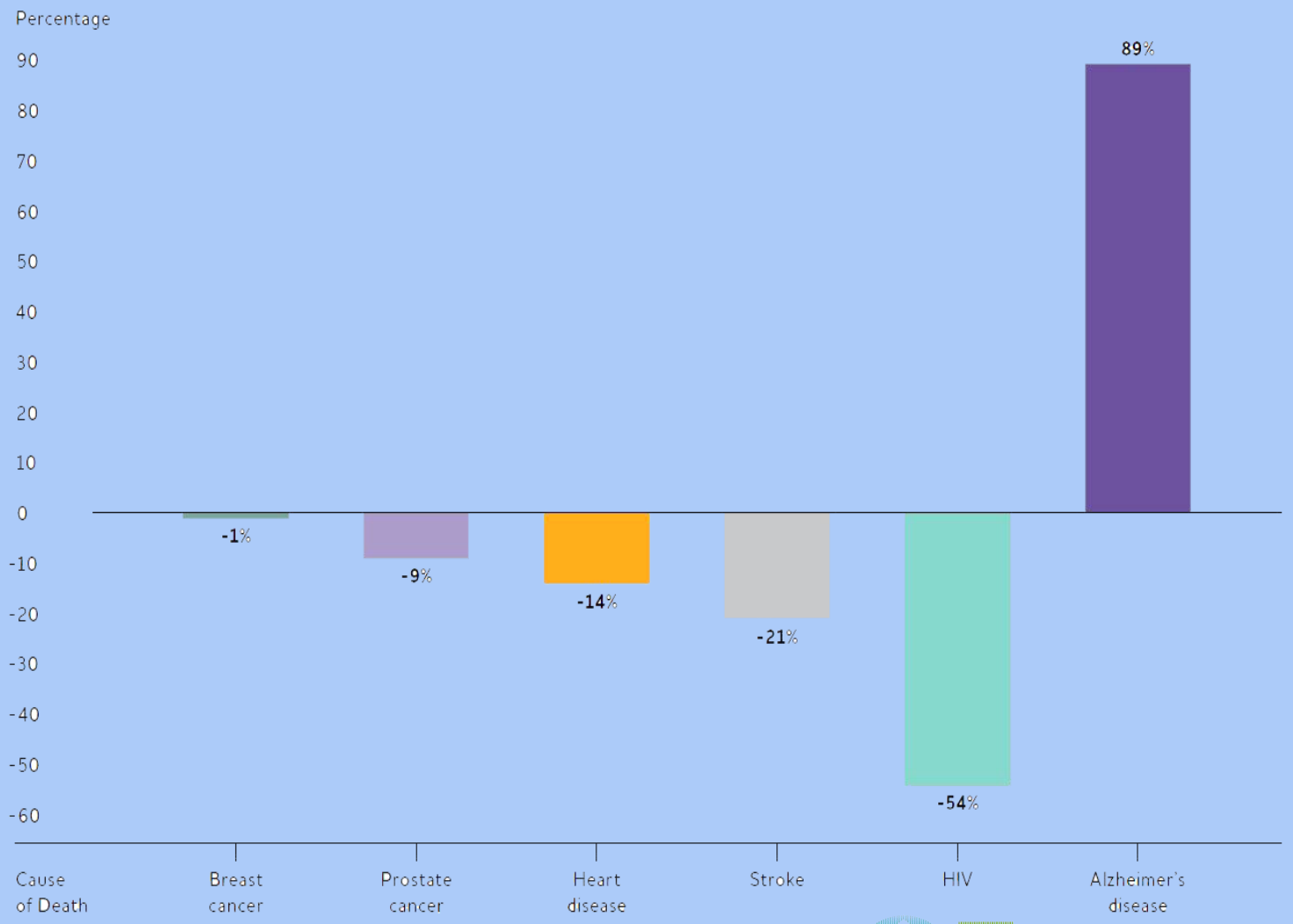
MOTIVATION
- Current status (1/2)

Estimated lifetime risk for Alzheimer's dementia, by sex, at age 45 and age 65

(2017 ALZHEIMER'S DISEASE FACTS AND FIGURES, Alzheimer's Association, USA)

Percentage **changes in selected causes of death** (all ages) between 2000 and 2014.

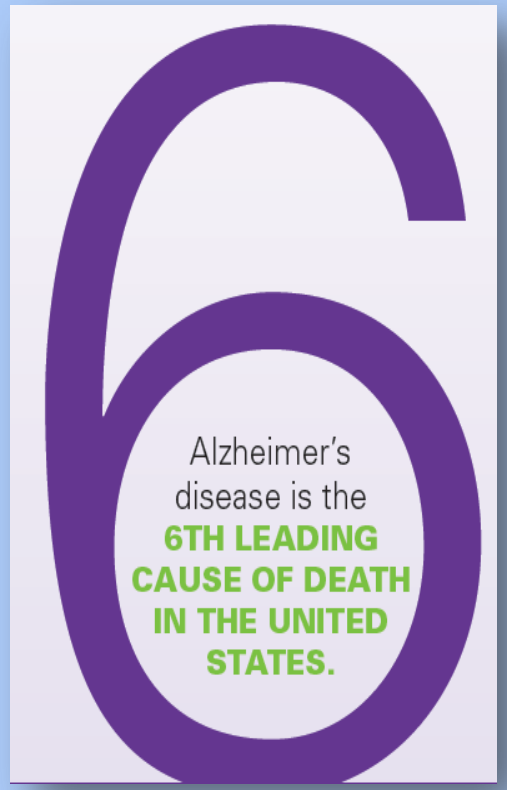
(2017 ALZHEIMER'S DISEASE FACTS AND FIGURES, Alzheimer's Association, USA)

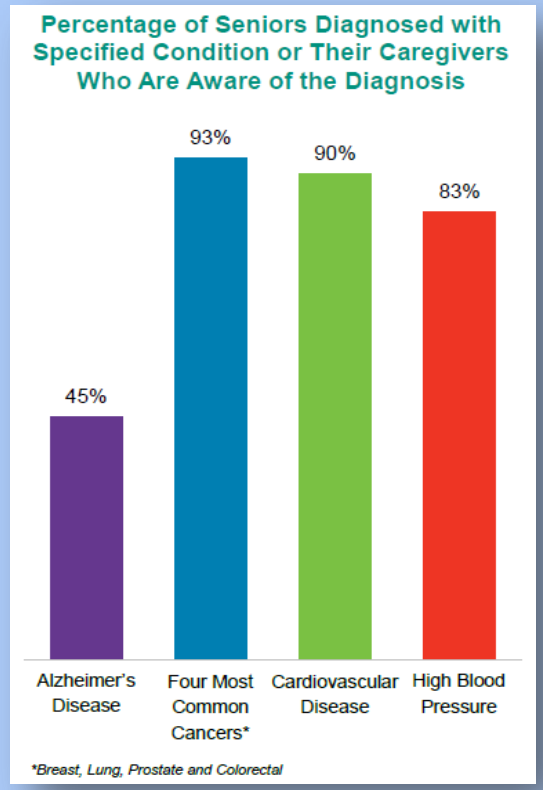


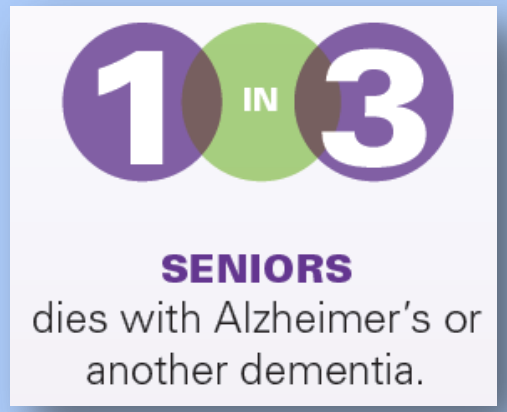
MOTIVATION
- Current status (2/2)

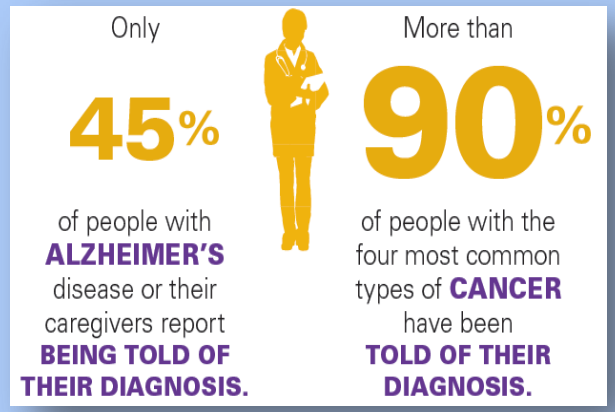
Only **45%** of AD patients or their caregivers **are told of the diagnosis.**

(2017 ALZHEIMER'S DISEASE FACTS AND FIGURES, Alzheimer's Association, USA)









MOTIVATION
- How is the diagnosis made?

- **No single test** to diagnose
- Definite diagnosis **can only be made postmortem** (Cuingnet et al, 2011)
  - *amyloid plaques*
  - *neurofibrillary tangles*
- **Medical history, physical exams, laboratory tests, neurobiological and mental assessment**
  - *A **laborious** process*
  - *Not allowing **early-detection***
  - *Lacking **objectivity***

- Brain imaging to **eliminate other possibilities**
  - *Tumors*
  - *Strokes*
  - *Hemorrhage*
  - *Fluid collection*
  - *Traumatic injuries*
- **No CAD system** has yet been a part of the clinical routine

OBJECTIVES

To develop a procedure which improves the **early detection of AD** using **structural brain MRI** volumes.

- **No new, unfamiliar requirements** – sMRI is already a part of the clinical practice.
- **Objectivity** – Procedure does not depend on manual assessment of the data.
- **Performance** – High accuracy, especially high sensitivity is aimed.

To build a foundation for a **fully-automated computer-assisted diagnostic system**.

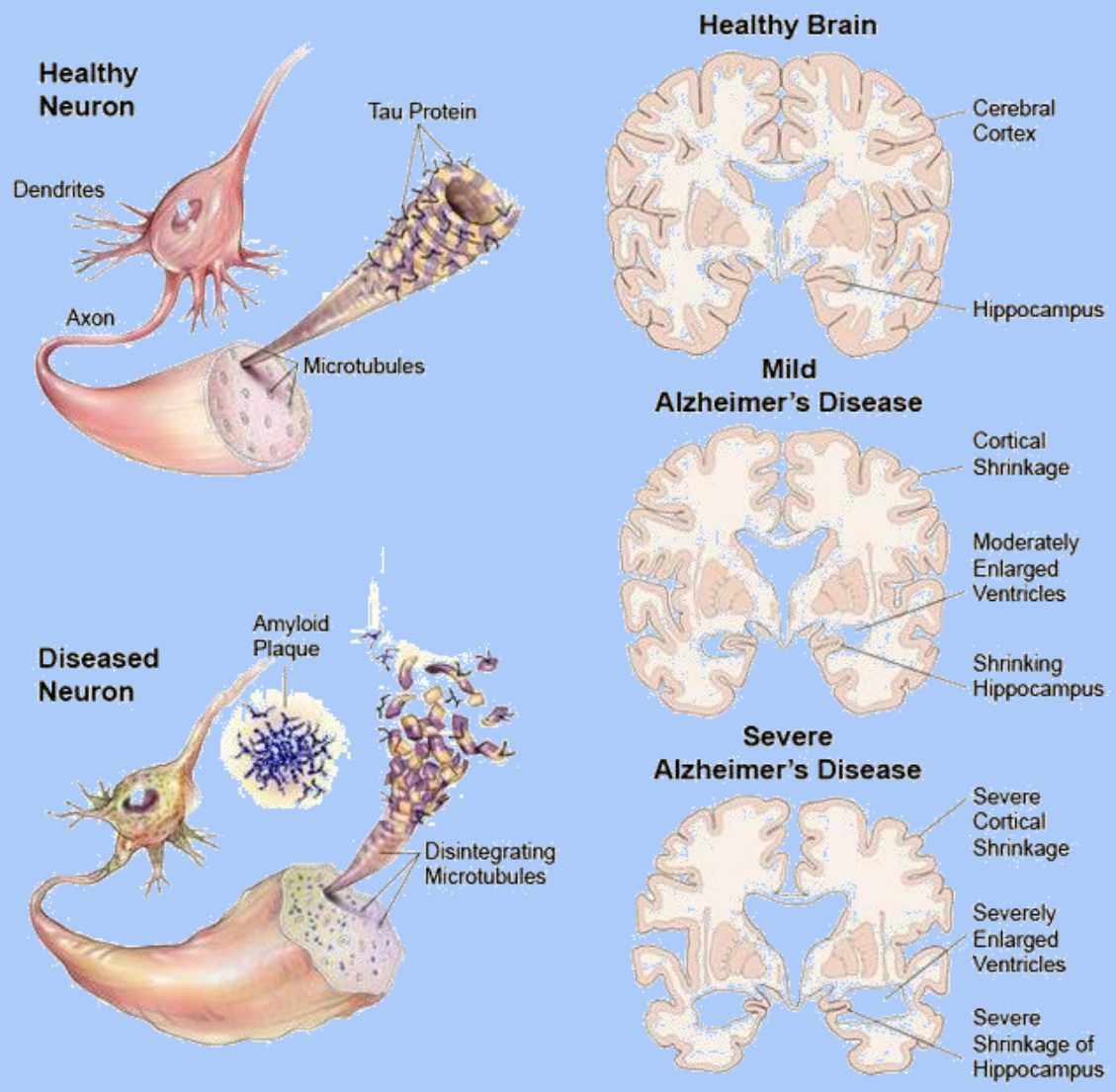
BACKGROUND

- Neuroimaging Biomarkers
- Classification with MARS/CMARS

STRUCTURAL ALTERATIONS in AD

An illustration showing the structural alterations developed in mild and severe Alzheimer's Disease.

Taken from: https://www.brightfocus.org/alzheimers/infographic/progression-alzheimers-disease



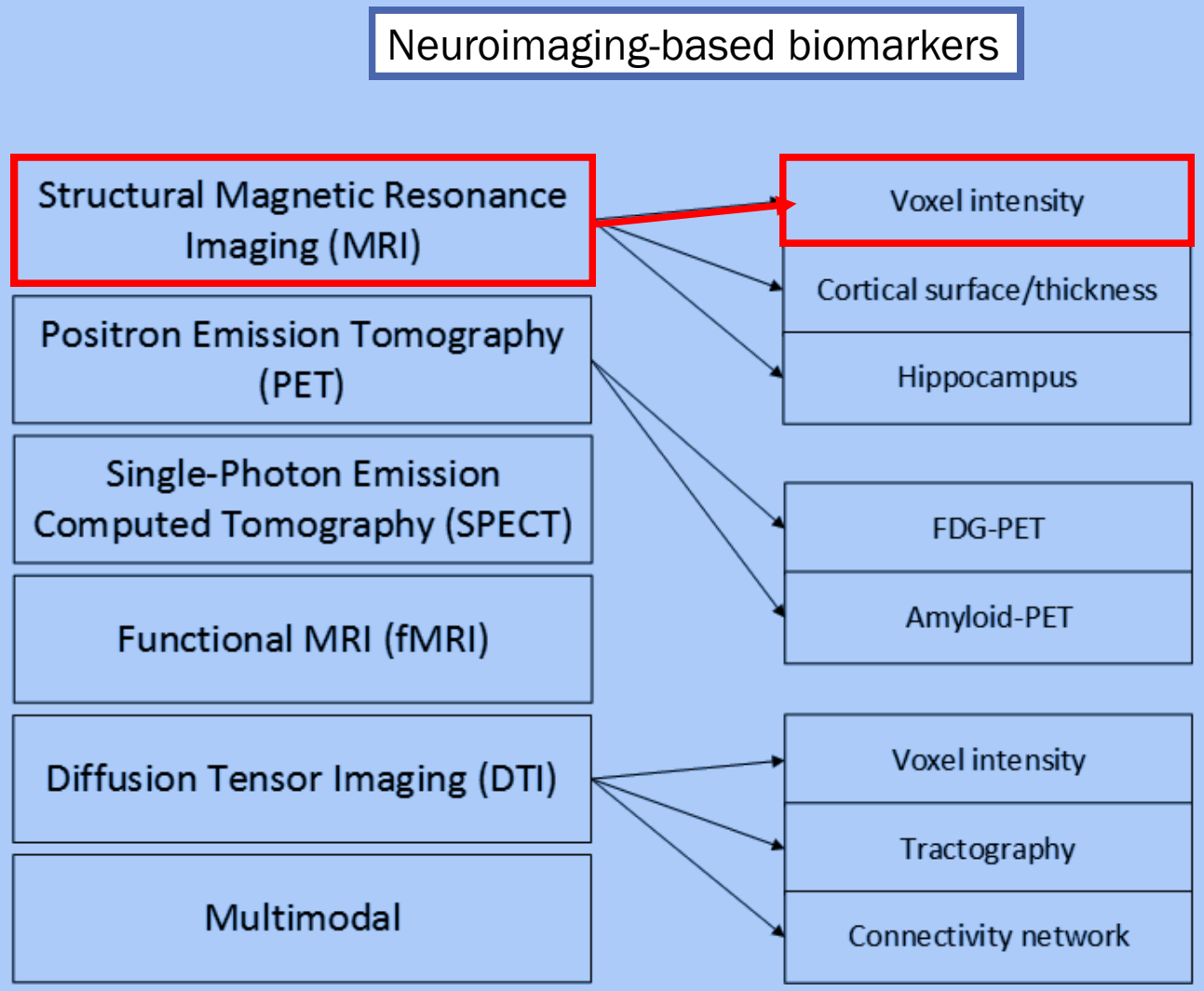
NEUROIMAGING BIOMARKERS

ROI-based methods suffer from (Zhang and Wang, 2015):

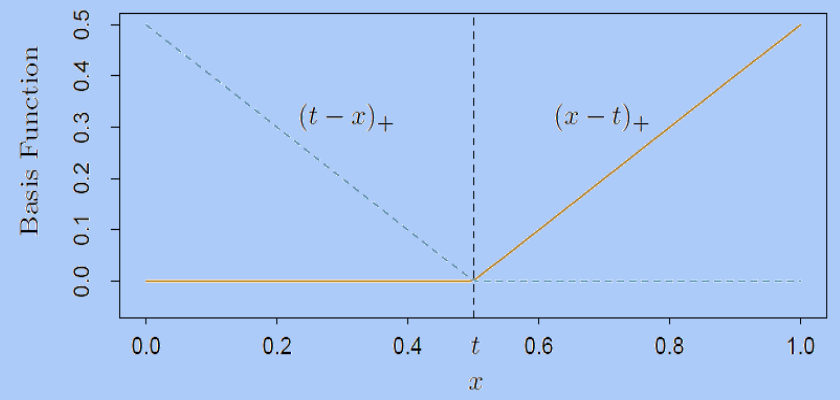
- Requirement of the **expert knowledge**
- Dependency to the **experience level**
- Difficulty in implementation of **mutual information** b/w voxels
- Potential existence of **other relevant regions**
- Examiners' tendency to **manual segmentation**

Voxel intensity-based «whole brain» approach provides:

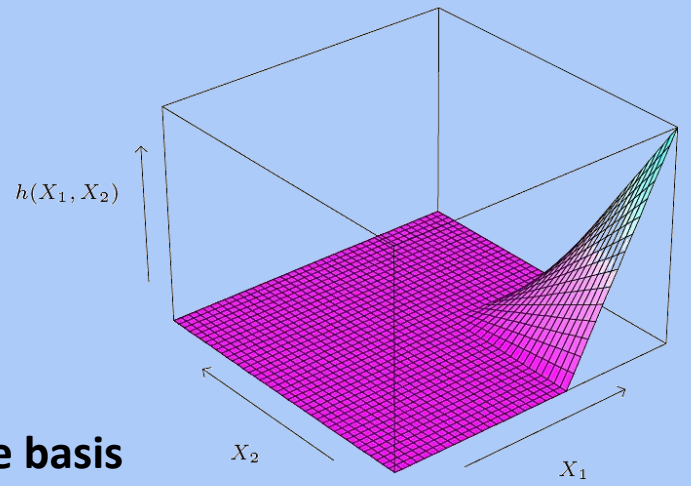
- **Largest set** of initial baseline features
- Possibility of expanding the procedure for **other neurodegenerative diseases**



MARS
- Multivariate Adaptive Regression Splines



A reflected pair of two hinge functions: The basis functions $(x-t)_+$ (solid orange) and $(t-x)_+$ (broken blue) used by MARS where t = 0.5*



Function $h(X_1, X_2)$, resulting from multiplication of two piecewise linear MARS basis functions*

(*) T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2 ed., Springer, 2009

Model function:

$$\hat{f}(\boldsymbol{x}) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(\boldsymbol{x})$$

Basis functions:

$$B_i(\boldsymbol{x}) = \prod_{j=1}^{K_m} \left(s_{\kappa_j^m} (x_{\kappa_j^m} - \tau_{\kappa_j^m}) \right)_+$$

A reflected pair:

$$(x-t)_+ = \begin{cases} x-t, & \text{if } x > t \\ 0, & \text{otherwise} \end{cases}$$

$$(t-x)_+ = \begin{cases} t-x, & \text{if } x < t \\ 0, & \text{otherwise} \end{cases}$$

MARS

- Avoiding Overfit through the Backward Step

Backward step cancels out some basis functions and update coefficients accordingly.

ACC = 0.6689

```
0.541                                                          // 0
-0.317 * max(0, x[2561] - 0.497)                               // 1
-1.29 * max(0, 0.497 - x[2561])                                // 2
+0.247 * max(0, x[994] - 0.0838) * max(0, 0.497 - x[2561])     // 3
+0.691 * max(0, 0.0838 - x[994]) * max(0, 0.497 - x[2561])     // 4
-4.72 * max(0, 0.497 - x[2561]) * max(0, x[2967] - 0.0715)     // 5
+0.117 * max(0, 0.497 - x[2561]) * max(0, 0.0715 - x[2967])    // 6
+1.88 * max(0, x[2542] - -0.226) * max(0, 0.497 - x[2561])     // 7
+0.138 * max(0, -0.226 - x[2542]) * max(0, 0.497 - x[2561])    // 8
-0.626 * max(0, x[912] -  1.03)                                // 9
+0.0952   * max(0, 1.03 - x[912])                              // 10
+0.201 * max(0, x[2155] - -0.574)                              // 11
+0.251 * max(0, -0.574 - x[2155])                              // 12
+0.269 * max(0, x[2840] - 0.165)                               // 13
+0.604 * max(0, 0.165 - x[2840])                               // 14
```

BACKWARD STEP

ACC = 0.6757

```
0.467                                                          // 0
-0.325 * max(0, x[2561] - 0.497)                               // 1
-1.14 * max(0, 0.497 - x[2561])                                // 2
+0.213 * max(0, x[994] - 0.0838) * max(0, 0.497 - x[2561])     // 3
+0.625 * max(0, 0.0838 - x[994]) * max(0, 0.497 - x[2561])     // 4
-4.67 * max(0, 0.497 - x[2561]) * max(0, x[2967] - 0.0715)     // 5
+1.81 * max(0, x[2542] - -0.226) * max(0, 0.497 - x[2561])     // 6
-0.579 * max(0, x[912] - 1.03)                                 // 7
+0.107 * max(0, 1.03 - x[912])                                 // 8
+0.208 * max(0, x[2155] - -0.574)                              // 9
+0.258 * max(0, -0.574 - x[2155])                              // 10
+0.295 * max(0, x[2840] - 0.165)                               // 11
+0.72 * max(0, 0.165 - x[2840])                                // 12
```

CMARS
- Conic Multivariate Adaptive Regression Splines

MARS - Backward Algorithm

- $GCV := \frac{1}{N}\frac{\sum_{i=1}^{N}\left(y_i - \hat{f}_\alpha(\boldsymbol{x}_i)\right)^2}{\left(1 - {}^{C(\alpha)}/_{N}\right)^2}$

RSS

Penalty Parameter

CMARS

- $PRSS := \sum_{i=1}^{N}(y_i - f(\boldsymbol{x}_i))^2 + \sum_{m=1}^{M_{max}} \lambda_m \sum^2_{\substack{|\boldsymbol{\alpha}|=1 \\ \boldsymbol{\alpha}=(\alpha_1,\alpha_2)^T}} \sum_{\substack{r<s \\ r,s\epsilon V_m}} \int \beta_m^2 \left[D_{r,s}^{\alpha} B_m(\boldsymbol{t}^m)\right]^2 d\boldsymbol{t}^m$

- $PRSS \approx \left\|\boldsymbol{y} - \boldsymbol{B}(\widetilde{\boldsymbol{d}})\boldsymbol{\beta}\right\|_2^2 + \lambda\|\boldsymbol{L}\boldsymbol{\beta}\|_2^2$

Tikhonov Regularization

CMARS
- Optimization

CQP Problem

$$\underset{t,\boldsymbol{\beta}}{\text{minimize}} \quad t,$$

$$\text{subject to } \|\boldsymbol{y} - \boldsymbol{B}(\tilde{\boldsymbol{d}})\boldsymbol{\beta}\|_2 < t,$$

$$\|\boldsymbol{L}\boldsymbol{\beta}\|_2 \leq \sqrt{\tilde{M}}.$$

$$y_1 - \boldsymbol{B}(\boldsymbol{x}_1)\boldsymbol{\beta} = \beta_{M+1},$$
$$y_2 - \boldsymbol{B}(\boldsymbol{x}_2)\boldsymbol{\beta} = \beta_{M+2},$$
$$\vdots$$
$$y_N - \boldsymbol{B}(\boldsymbol{x}_N)\boldsymbol{\beta} = \beta_{M+N},$$

$$\left(\sum_{i=M+1}^{M+N} \beta_i^{\,2}\right)^{\frac{1}{2}} < t.$$

$$L_1\beta_1 = \beta_{M+N+1},$$
$$L_2\beta_2 = \beta_{M+N+2},$$
$$\vdots$$
$$L_M\beta_M = \beta_{M+N+M},$$

$$\left(\sum_{i=M+N+1}^{M+N+M} \beta_i^{\,2}\right)^{\frac{1}{2}} \leq \sqrt{\tilde{M}}.$$

N Linear constraints

1 Conic constraint

M Linear constraints

1 Conic constraint

M O S E K

CMARS
- Conic Quadratic Optimization

CMARS replaces the backward step, does not remove basis functions.
Instead, update coefficients.

ACC = 0.6689

```
0.541                                                          // 0
-0.317 * max(0, x[2561] - 0.497)                               // 1
-1.29 * max(0, 0.497 - x[2561])                                // 2
+0.247 * max(0, x[994] - 0.0838) * max(0, 0.497 - x[2561])     // 3
+0.691 * max(0, 0.0838 - x[994]) * max(0, 0.497 - x[2561])     // 4
-4.72 * max(0, 0.497 - x[2561]) * max(0, x[2967] - 0.0715)     // 5
+0.117 * max(0, 0.497 - x[2561]) * max(0, 0.0715 - x[2967])    // 6
+1.88 * max(0, x[2542] - -0.226) * max(0, 0.497 - x[2561])     // 7
+0.138 * max(0, -0.226 - x[2542]) * max(0, 0.497 - x[2561])    // 8
-0.626 * max(0, x[912] -  1.03)                                // 9
+0.0952   * max(0, 1.03 - x[912])                              // 10
+0.201 * max(0, x[2155] - -0.574)                              // 11
+0.251 * max(0, -0.574 - x[2155])                              // 12
+0.269 * max(0, x[2840] - 0.165)                               // 13
+0.604 * max(0, 0.165 - x[2840])                               // 14
```

CMARS

ACC = 0.7027

```
0.541                                                          // 0
-0.328 * max(0, x[2561] - 0.497)                               // 1
-0.772 * max(0, 0.497 - x[2561])                               // 2
+0.164 * max(0, x[994] - 0.0838) * max(0, 0.497 - x[2561])     // 3
+0.399 * max(0, 0.0838 - x[994]) * max(0, 0.497 - x[2561])     // 4
-2.8 * max(0, 0.497 - x[2561]) * max(0, x[2967] - 0.0715)      // 5
+0.00296  * max(0, 0.497 - x[2561]) * max(0, 0.0715 - x[2967]  // 6
+1.94 * max(0, x[2542] - -0.226) * max(0, 0.497 - x[2561])     // 7
+0.0239 * max(0, -0.226 - x[2542]) * max(0, 0.497 - x[2561])   // 8
-0.451 * max(0, x[912] -  1.03)                                // 9
+0.257 * max(0, 1.03 - x[912])                                 // 10
+0.297 * max(0, x[2155] - -0.574)                              // 11
+0.32 * max(0, -0.574 - x[2155])                               // 12
+0.345 * max(0, x[2840] - 0.165)                               // 13
+0.575 * max(0, 0.165 - x[2840])                               // 14
```

WHY MARS/CMARS?

MARS

- ***Nonparametric:*** *No prior assumption of any parametric form on the data*
- ***Adaptive:*** *Learning from the data*
- *Enables* ***nonlinear*** *models: Modelling interactions and dependencies between variables*
- *Enables* ***flexible, complex*** *models: Superposition of linearly independent BFs*
- ***Global:*** *Assessing the data as a whole, determining significance of the variables.*

CMARS

- *Preserves* ***information:*** *Re-weighting the relevancy of variables*
- *Mathematically more* ***integrated:*** *Involves utilization of regularization and modern optimization tools*

PROPOSED METHODOLOGY

- Study Data
- Feature Extraction Method
- Dimensionality Reduction Procedure

STUDY DATA
- Subject Groups in (Cuingnet et al., 2011)

- Performances of **10 different approaches** of various researchers have been evaluated using 509 subjects from the ADNI database (**Cuingnet et al., 2011**):
  - *5 **voxel-based** methods,*
  - *3 methods based on **cortical thickness**, and,*
  - *2 methods based on the **hippocampus***
- 3 classification experiments were performed:
  - ***CN/AD, CN/MCIc, and MCIc/MCInc***

- Image properties:
  - ***T1-weighted MR** images,*
  - *ADNI acquisition protocol,*
  - *When available, **baseline scan**, otherwise visiting scan,*
  - *Pre-processed images with some **post-acquisition corrections:***
    - image geometry corrections,
    - magnetic field intensity non-uniformity corrections, and,
  - ***"Best" quality** scan is determined by the ADNI investigators.*

USC University of Southern California

ADNI Alzheimer's Disease Neuroimaging Initiative

STUDY DATA
- Demographic Characteristics

Demographic characteristics of the studied patients. (Adapted from Cuingnet et al., 2011)

Diagnosis	Number	Age	Gender	MMS	Centers
CN (Train)	81	76.1 ± 5.6 [60 - 89]	38 M/43 F	29.2 ± 1.0 [25 - 30]	35
AD (Train)	69	75.8 ± 7.5 [55 - 89]	34 M/35 F	23.3 ± 1.9 [18 - 26]	32
MCIc (Train)	39	74.7 ± 7.8 [55-88]	22M/17F	26.0 ± 1.8 [23 - 30]	21
MCInc (Train)	67	74.3 ± 7.3 [58-87]	42M/25F	27.1 ± 1.8 [24 - 30]	30
CN (Test)	81	76.5 ± 5.2 [63 - 90]	38 M/43 F	29.2 ± 0.9 [26 - 30]	35
AD (Test)*	67	76.0 ± 7.1 [57 - 91]	32 M/35 F	23.2 ± 2.1 [20 - 27]	33
MCIc (Test)	37	74.9 ± 7.0 [57 - 87]	21 M/16 F	26.9 ± 1.8 [24 - 30]	24
MCInc (Test)	67	74.7 ± 7.3 [58 - 88]	42 M/25 F	27.3 ± 1.7 [24 - 30]	31

FEATURE EXTRACTION

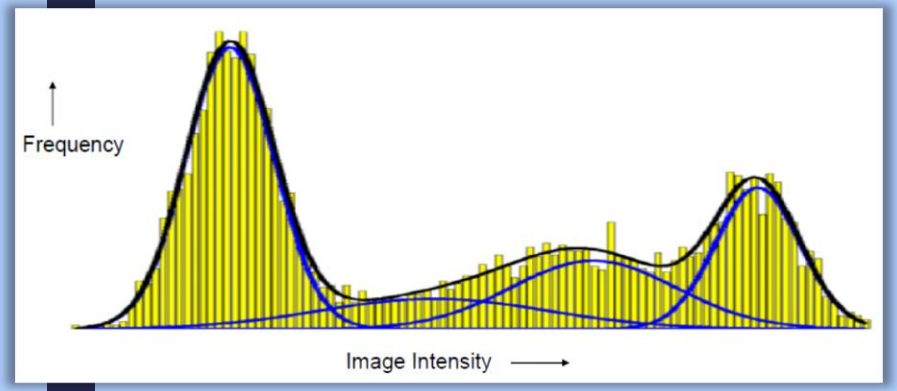
- Voxel-Based Morphometry (VBM) using SPM

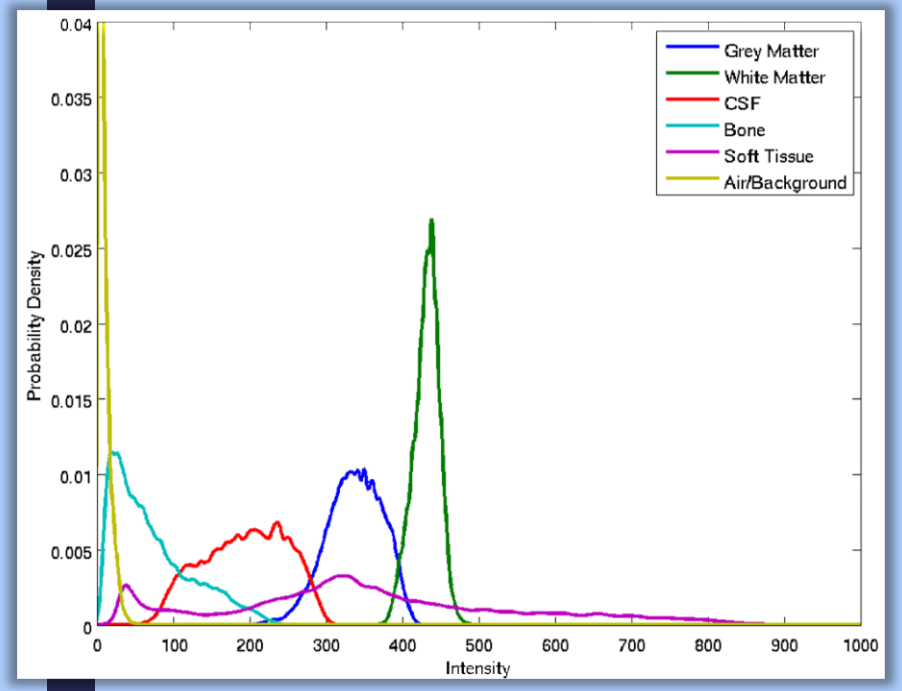
- SPM
  - *a MATLAB suite to **organize** and **interpret** neuroimaging data*
  - *analysis of **brain imaging data sequences***
  - *a series of images from **different cohorts**, or **time-series from the same subject***
  - ***MRI**, fMRI, PET, SPECT, EEG, and MEG*

- VBM
  - *sensitive to the **differences b/w local composition of brain tissues**, while discounting **positional and other large-scale variations** in gross anatomy*
  - ***mass-univariate** (analysis of each voxel separately)*
  - *independent of the a priori assumption that abnormalities are contained within **specific anatomical regions***

UNIFIED SEGMENTATION

- MoG --> The Objective Function

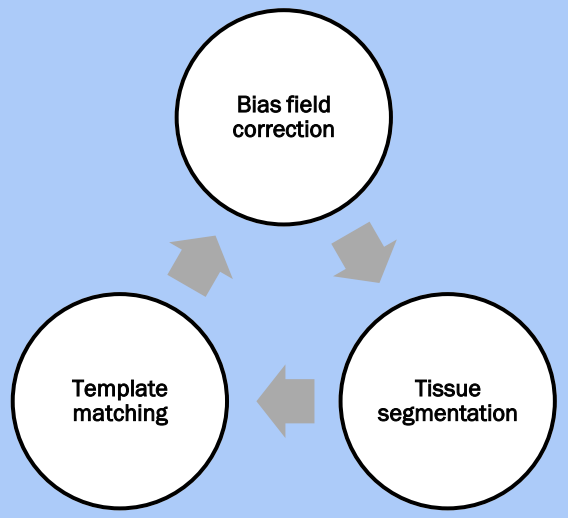




(Top) An example MoG model visualization*.

(Bottom) T1 MRI Tissue intensity distributions*.

* Taken from Ashburner's SPM course notes



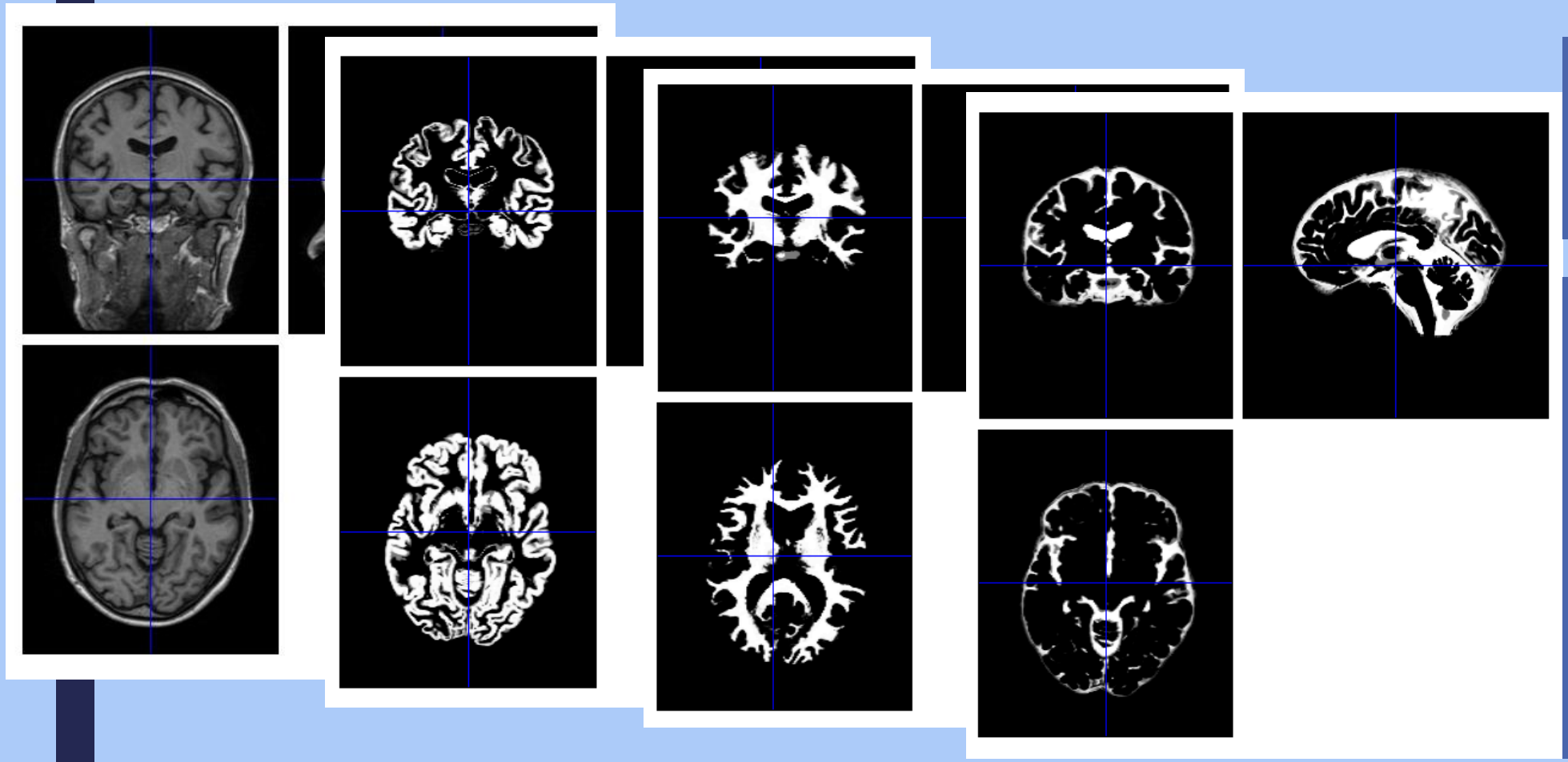
$$P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\gamma}) = \prod_{i=1}^{I}\left(\sum_{k=1}^{K_i}\frac{\gamma_k}{(2\pi\sigma_k^2)^{\frac{1}{2}}}\exp\left(-\frac{(y_i-\mu_k)^2}{2\sigma_k^2}\right)\right)$$

$$-\log P(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})$$

$$= -\sum_{i=1}^{I}\log\left(\frac{\rho_i(\boldsymbol{\beta})}{\sum_{k=1}^{K_i}\gamma_k b_{i,k}(\boldsymbol{\alpha})}\sum_{k=1}^{K_i}\frac{\gamma_k b_{i,k}(\boldsymbol{\alpha})}{(2\pi\sigma_k^2)^{\frac{1}{2}}}\exp\left(-\frac{(\rho_i(\boldsymbol{\beta})y_i-\mu_k)^2}{2\sigma_k^2}\right)\right)$$

UNIFIED SEGMENTATION

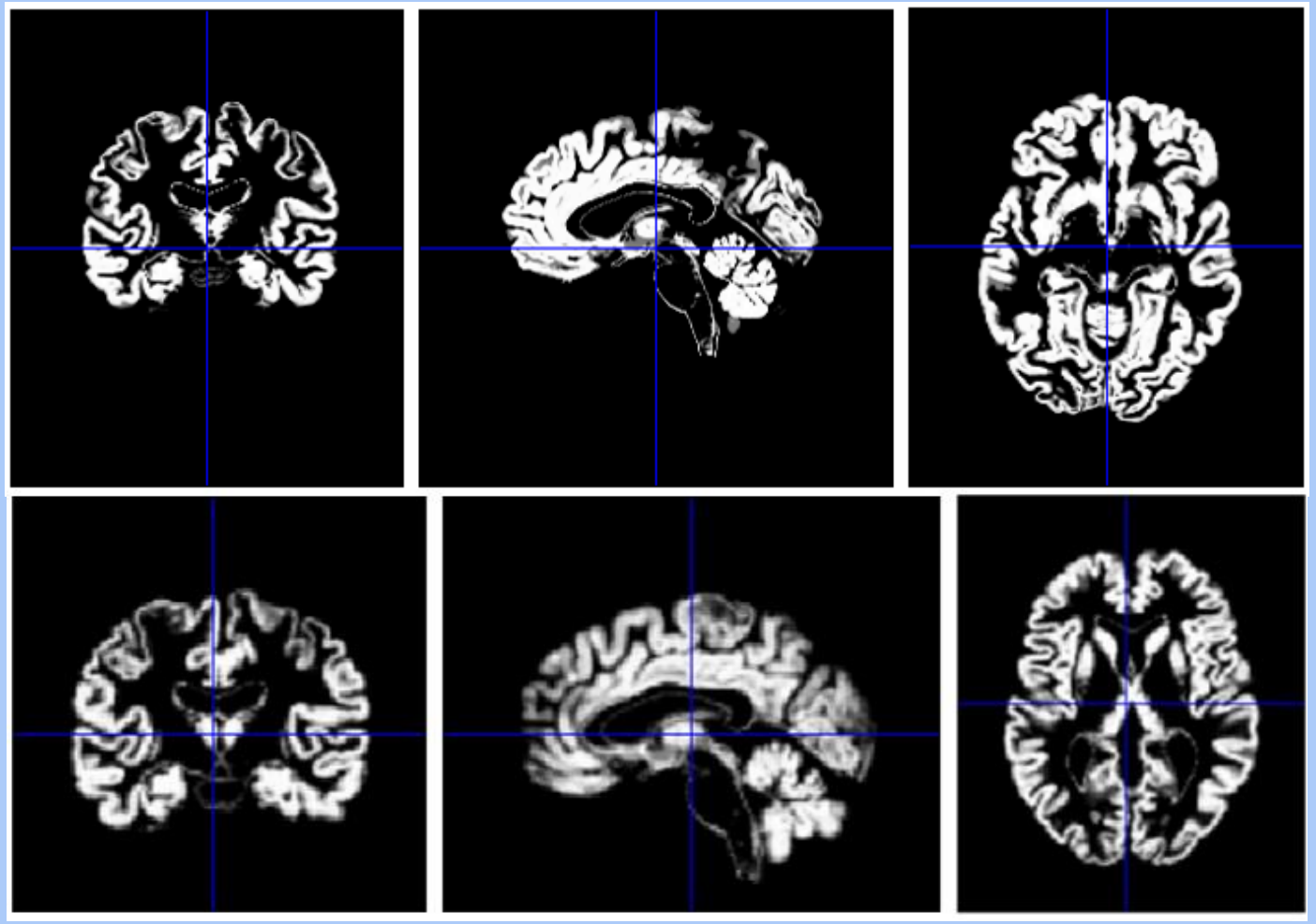
- Native Space-Aligned *Tissue Probability Maps* (*TPM*)



- Original Image
- Gray Matter TPM
- White Matter TPM
- CSF TPM

- Native-space aligned
- 256 x 256 x 166
- 0.938 x 0.938 x 1.2 mm anisotropic resolution
- # of Gaussians:
  - 3 for GM
  - 2 for WM
  - 2 for CSF
  - 5 for others

SPATIAL NORMALIZATION
- MNI* Space



- Rigid body registration + Zoom
- (*Top*) Coronal, sagittal, axial views of an example GM native space-aligned TPM
- (*Bottom*) Axial, sagital, coronal views after normalization to the MNI space

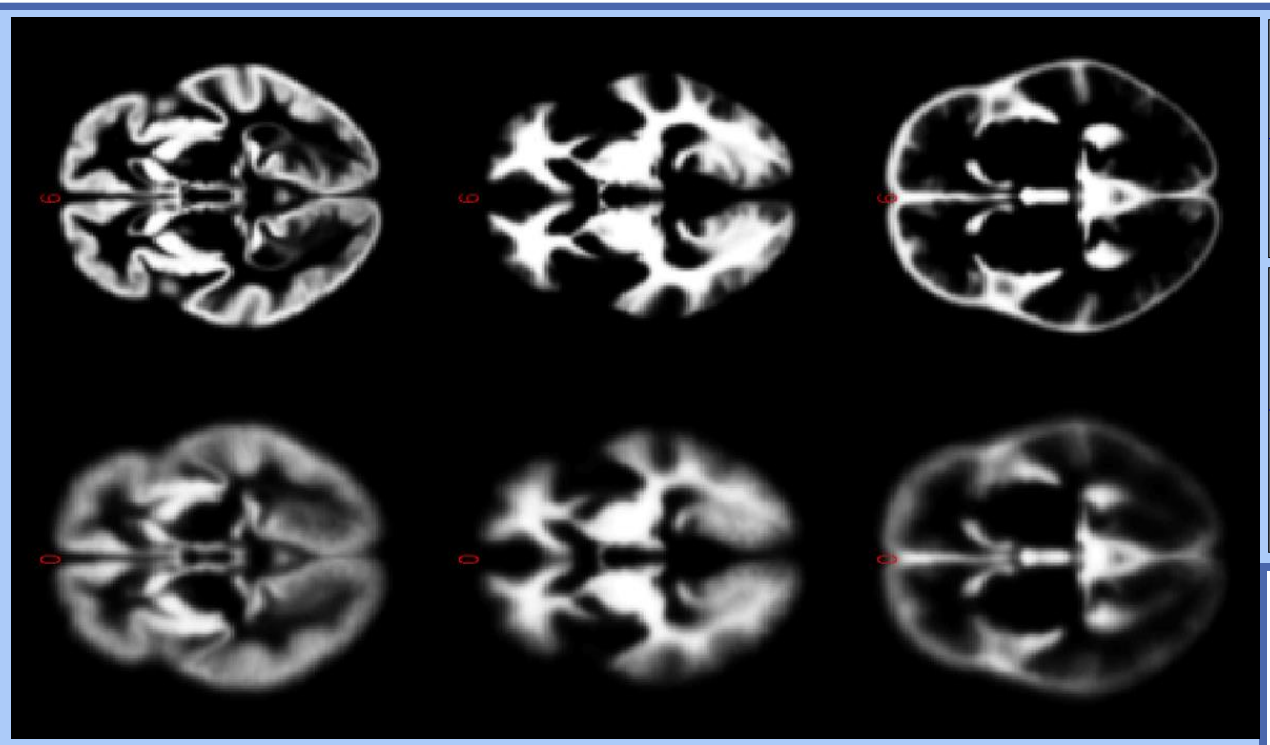
- 121 x 145 x 121
- 1.5 x 1.5 x 1.5 mm isotropic resolution
- Still **group-wise unregistered!**

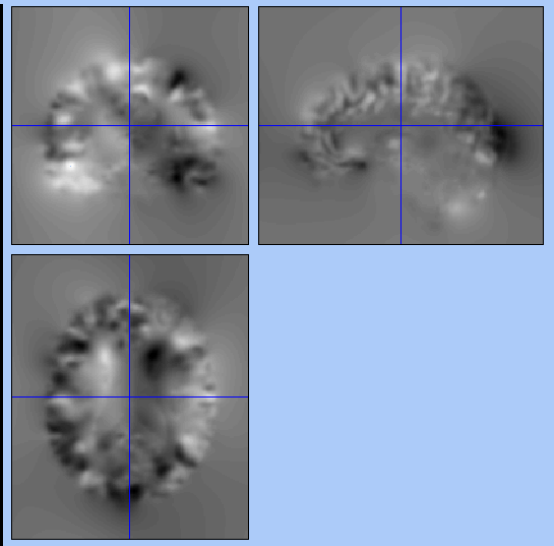
*MNI: Montreal Neurological Institute

GROUPWISE NONLINEAR REGISTRATION
- DARTEL Templates & Flow Fields

Diffeomorphic **A**natomical **R**egistration **T**hrough **E**xponentiated **L**ie Algebra

- A diffeomorphism is a **differentiable mapping** with a **differentiable inverse**, i.e., non-zero Jacobian determinant.
- Modeling transformations with diffeomorphisms ensures **certain unique and desirable topological properties**.
- The **exponential mapping** maintains only a **single vector field** that is constant in time.





(Left) Initial vs final averages of GM, WM, and CSF

(Top) Flow-field example

FINAL STEPS
- Smoothing & Intensity Modulation

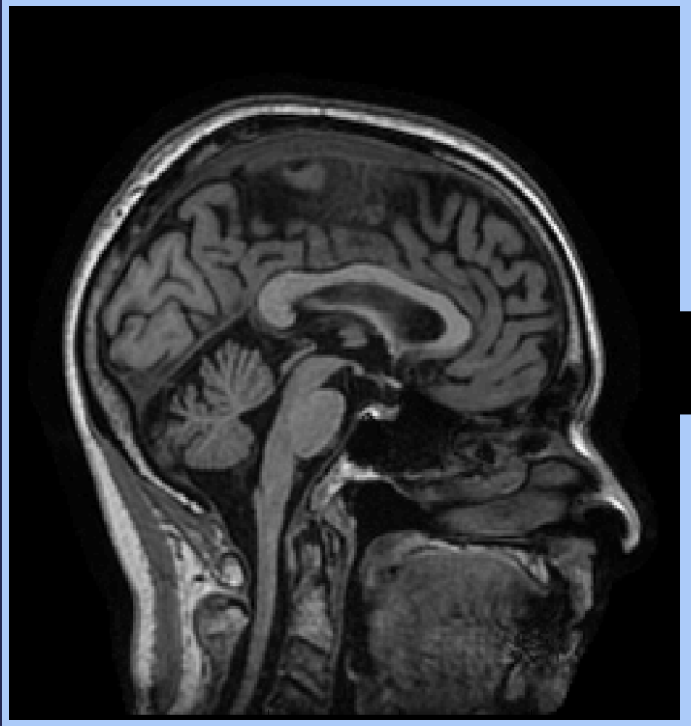
Spatial Smoothing

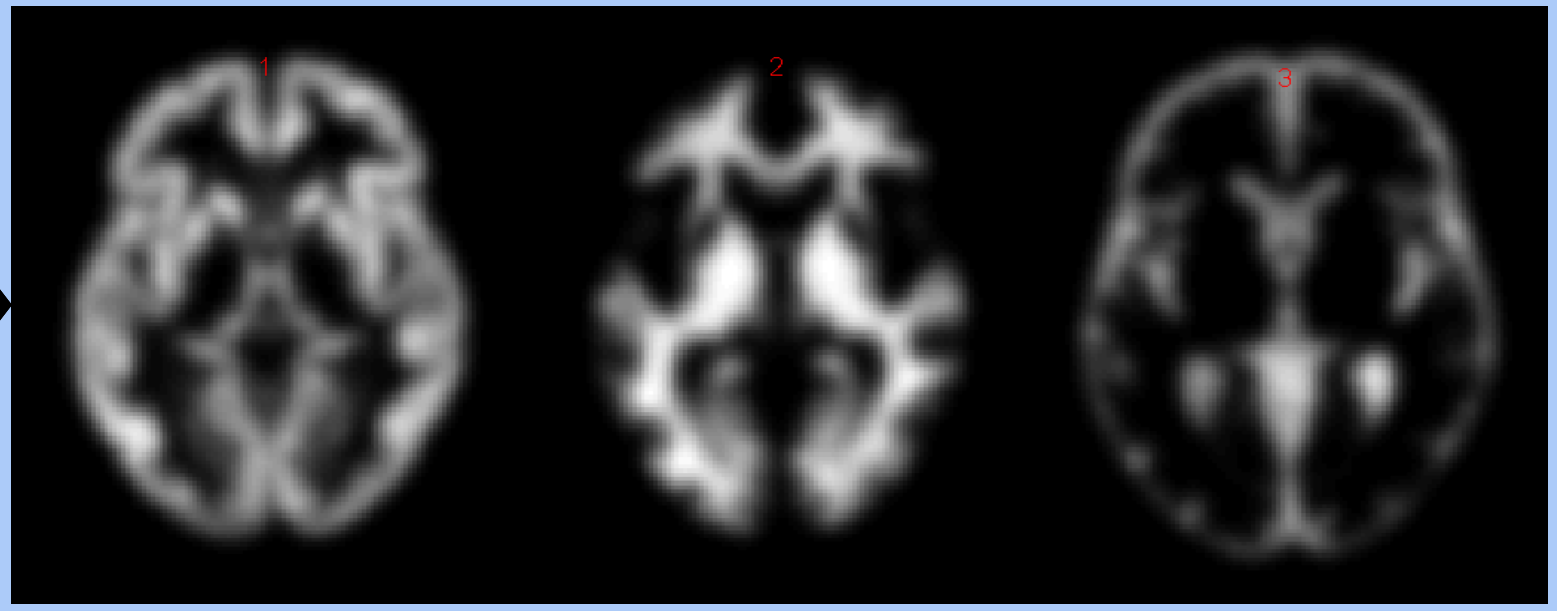
- 8 mm FWHM,
- Improvement of the **signal to noise ratio (SNR)**, thus, sensitivity – The Matched Filter Theorem,
- Improving **validity of the statistical tests** by making the error distribution more normal,
- **Accommodation of anatomical and functional variations** between subjects,
- Reduction of **spatial resolution** of the data.

Intensity Modulation

- Correction of the volumetric differences which are inevitably introduced by warping,
- Normalized tissue volume is adjusted by multiplying by its relative volume before and after warping,
- Total amount of tissues are preserved.

BASELINE FEATURES
- Tissue Probability Maps (TPM)





DESIGN MATRIX
- AD/CN Group

$n = p/3 = 2{,}122{,}945$ GM tissue probabilities

$n = p/3 = 2{,}122{,}945$ WM tissue probabilities

$n = p/3 = 2{,}122{,}945$ CSF tissue probabilities

N=150 class labels

$N = 150$ Instances

$$\begin{bmatrix} GM_1^1 & \cdots & GM_n^1 & WM_1^1 & \cdots & WM_n^1 & CSF_1^1 & \cdots & CSF_n^1 & L_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ GM_1^N & \cdots & GM_n^N & WM_1^N & \cdots & WM_n^N & CSF_1^N & \cdots & CSF_n^N & L_N \end{bmatrix}$$

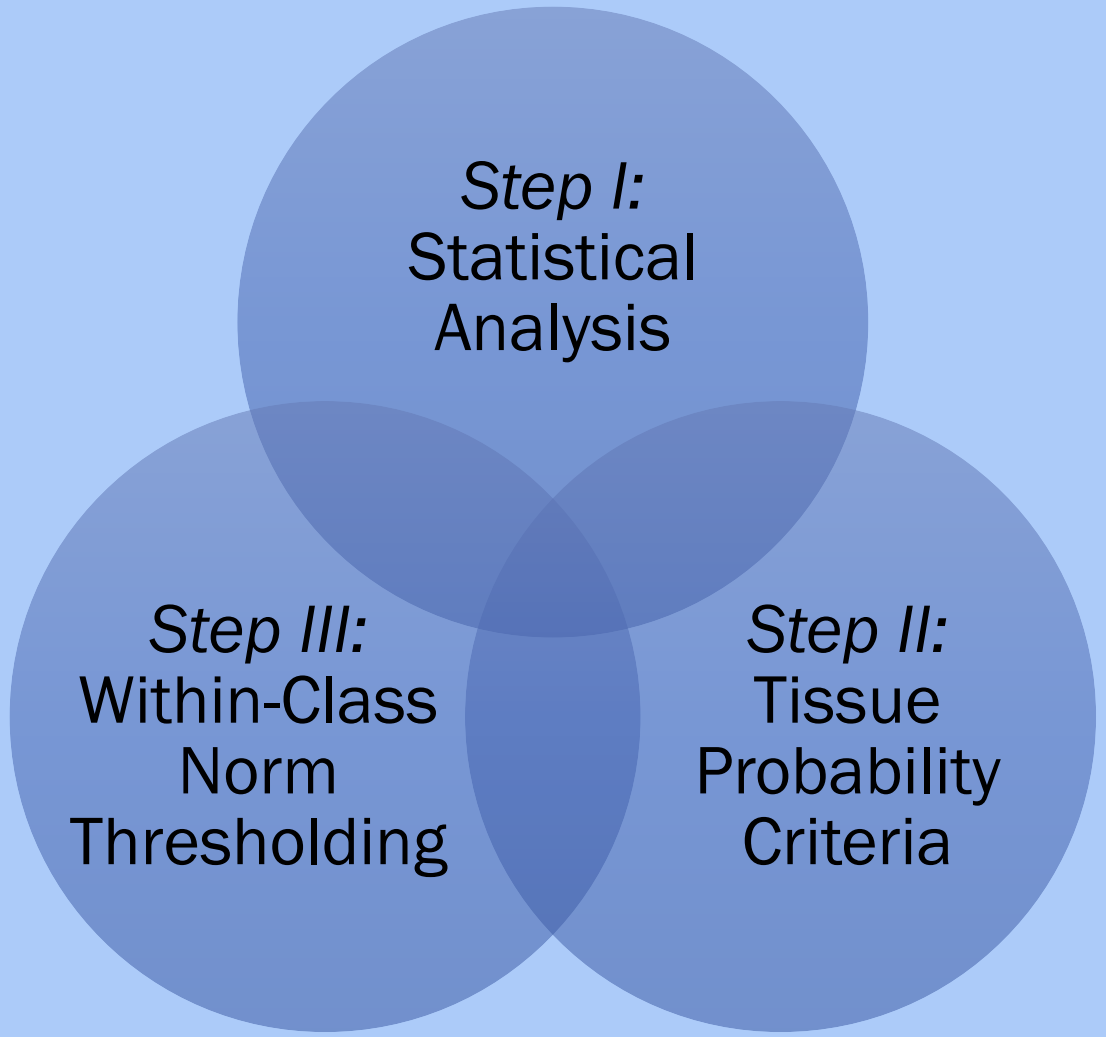
$p = 6{,}368{,}835$ variables (7.12 GB !)

Reduced by background elimination to:

2,112,054 variables

DIMENSIONALITY REDUCTION

a 3-Step, Hybrid Procedure for Feature Selection



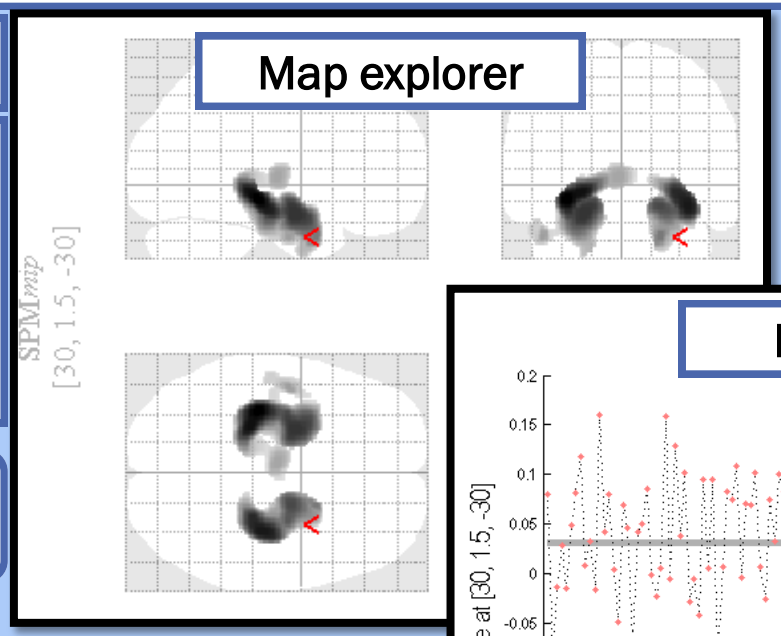
DIMENSIONALITY REDUCTION

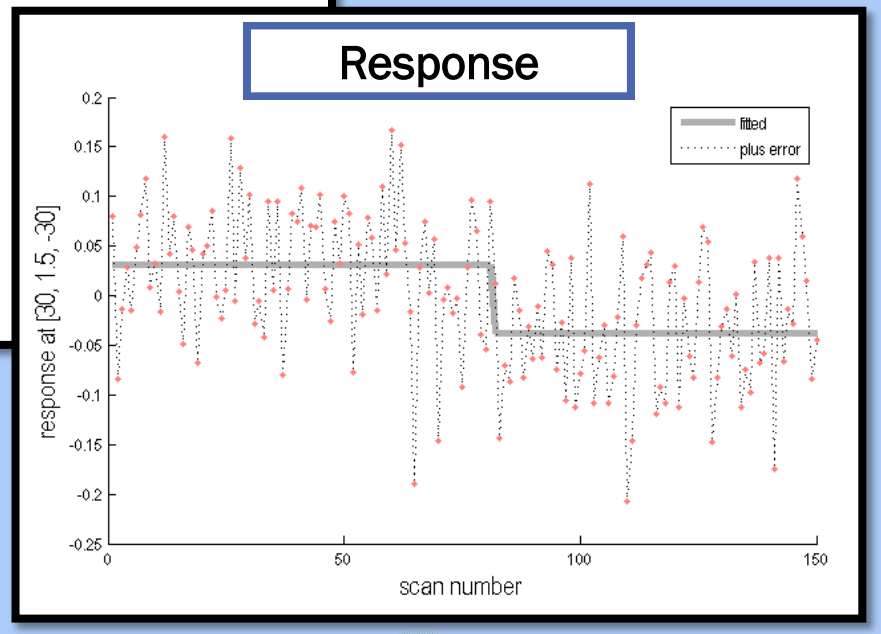
- *Step I:* Statistical Analysis

GLM: General Linear Model is a generalization of multiple linear regression model to the case of more than one dependent variable.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{u}$$

$\boldsymbol{y}$: the vector of independent observations
$\boldsymbol{X}$: contains the tissue probabilities as features
$\boldsymbol{b}$: the vector of unknown parameters
$\boldsymbol{u}$: errors - independent and identically distributed (i.i.d.) random variables with mean value 0





Two-sample *t*-Test

- Compare the means of the two populations at each voxel
- Discriminate the statistically significant voxel positions from others

DIMENSIONALITY REDUCTION

- *Step II:* Tissue Probability Criteria

A modification to the STAND Score (Vemuri, 2008)

1. $P_{GM}(i,j) + P_{WM}(i,j) < \tau_1, \ \forall i \in [1, 2, \dots, n]$

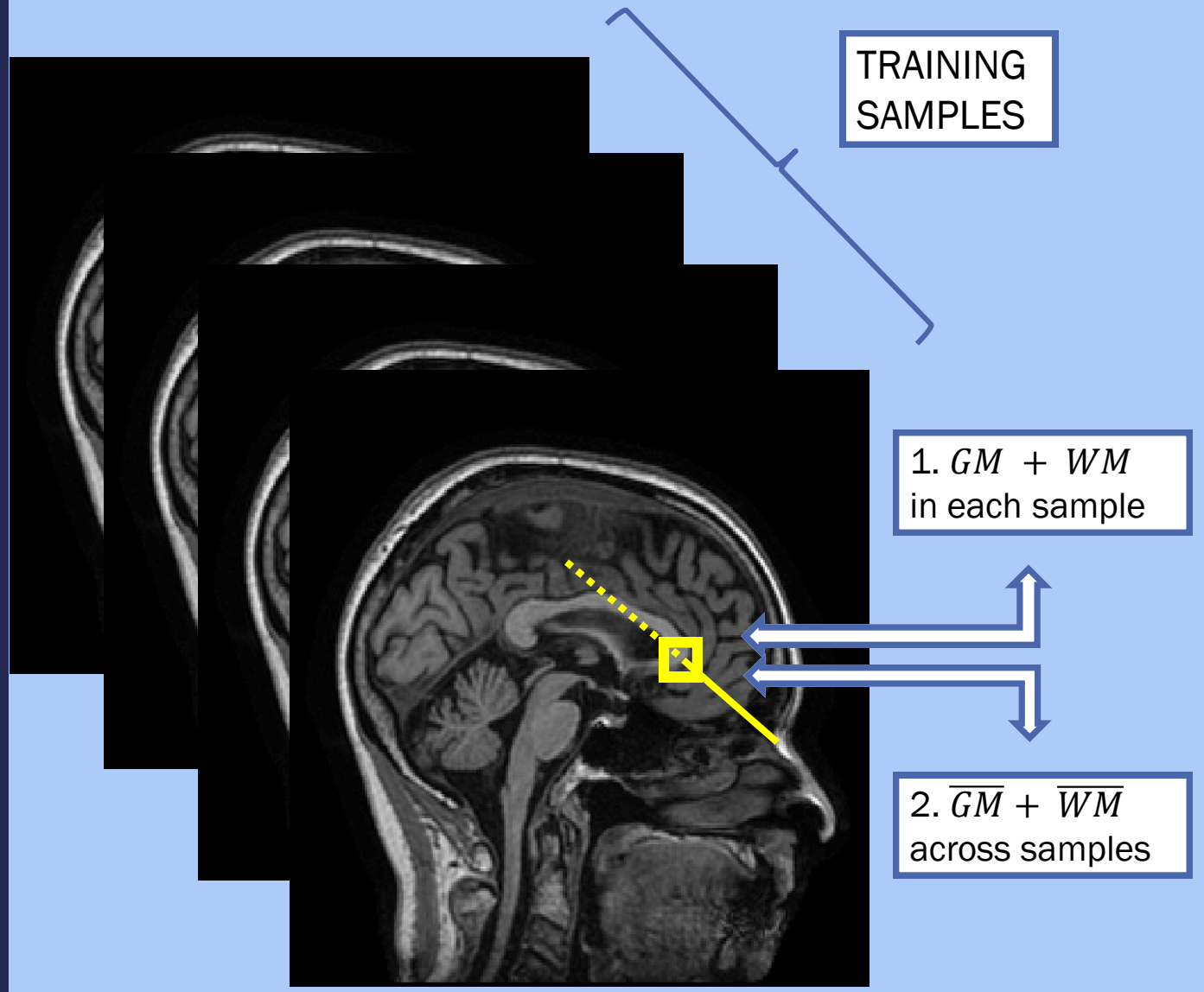
   sum of probabilities of being gray matter and white matter *is smaller than the first threshold,* τ_1

2. $\overline{P_{GM}}(j) + \overline{P_{WM}}(j) < \tau_2$

   the sum of the sample means of probabilities of being gray matter and white matter *is smaller than the second threshold,* τ_2

Optimum thresholds:

$$\boldsymbol{\tau_1 = 0.5} \ \textit{and} \ \boldsymbol{\tau_2 = 0.7}$$



DIMENSIONALITY REDUCTION

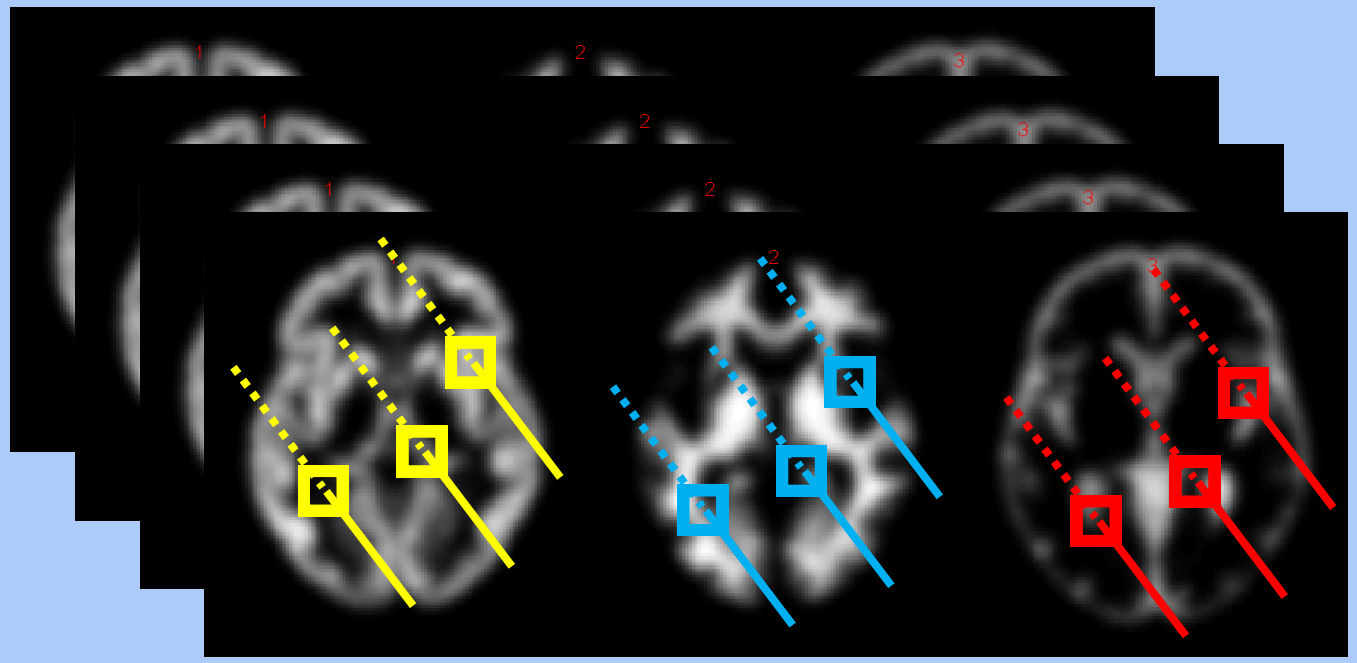
- *Step III:* Within-Class Norm Thresholding

$$\|P(j)\|_2 = \sqrt{\sum_{i=1}^{n} (P(i,j))^2}$$

$$\mu_c = \frac{\sum_{j \in c} \|P(j)\|_2}{j_c}, \quad c \in \{GM, WM, CSF\}$$

$$\|P(j)\|_2 < \epsilon \mu_c$$

ϵ = 0.9



μ_{GM} μ_{WM} μ_{CSF}

PERFORMANCE EVALUATION

- Parameter Optimization
- Results
- Discussion

Voxel-MARS (*ANOR SI: OR in Neuroscience*)

Voxel-MARS: a method for early detection of Alzheimer's disease by classification of structural brain MRI

The Alzheimer's Disease Neuroimaging Initiative

Annals of Operations Research

ISSN 0254-5330

Ann Oper Res
DOI 10.1007/s10479-017-2405-7

Springer

Ann Oper Res (2017) 258:31–57
DOI 10.1007/s10479-017-2405-7

CrossMark

OR IN NEUROSCIENCE

Voxel-MARS: a method for early detection of Alzheimer's disease by classification of structural brain MRI

Alper Çevik[1] · Gerhard-Wilhelm Weber[2] · B. Murat Eyüboğlu[3] · Kader Karlı Oğuz[4] · The Alzheimer's Disease Neuroimaging Initiative

Published online: 3 February 2017

Abstract Neuroscience is of emerging importance along with the contributions of Operational Research to the practices of diagnosing neurodegenerative diseases with computer-aided systems based on brain image analysis. Although multiple biomarkers derived from Magnetic Resonance Imaging (MRI) data have proven to be effective in diagnosing Alzheimer's disease (AD) and mild cognitive impairment (MCI), no specific system has yet been a part of routine clinical practice. This paper aims to introduce a fully-automated voxel-based procedure, Voxel-MARS, for detection of AD and MCI in early stages of progression. Performance was evaluated on a dataset of 508 MRI volumes gathered from the Alzheimer's Disease Neuroimaging Initiative database. Data were transformed into a high-dimensional space through a feature extraction process. A novel 3-step feature selection procedure was applied. Multivariate Adaptive Regression Splines method was used as a classifier for the first time in the field of brain MRI analysis. The results were compared to those presented in a previous study on 28 voxel-based methods in terms of their ability to separate control normal (CN) subjects from the ones diagnosed with AD and MCI. It was observed that our method outperformed all of the others in sensitivity (83.58% in AD/CN and 78.38%

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). 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. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.

✉ Alper Çevik
alper.cevik@metu.edu.tr

1 Biomedical Engineering Graduate Program, Graduate School of Natural and Applied Sciences, Middle East Technical University, Ankara, Turkey

2 Institute of Applied Mathematics and Biomedical Engineering Graduate Program, Middle East Technical University, Ankara, Turkey

3 Department of Electrical and Electronics Engineering and Biomedical Engineering Graduate Program, Middle East Technical University, Ankara, Turkey

4 Department of Radiology, Faculty of Medicine, Hacettepe University, Ankara, Turkey

Springer

Table 6 Performance of our method is compared with the average outcomes of others

Case	Metric	Other methods	Voxel-MARS	Difference	Rank
AD/CN	SEN	71.46 ± 5.65 [59–82]	83.58	+12.12	1
	SPE	89.39 ± 5.03 [77–98]	86.42	−2.97	22
	PPV	85.18 ± 6.23 [72–96]	83.58	−1.60	21
	NPV	78.93 ± 3.88 [70–86]	86.42	+7.49	1
MCIc/CN	SEN	54.39 ± 12.39 [22–73]	78.38	+23.99	1
	SPE	88.82 ± 7.27 [73–99]	88.89	+0.07	17
	PPV	71.50 ± 11.97 [50–89]	76.32	+4.82	12
	NPV	81.21 ± 3.60 [73–87]	90.00	+8.79	1
MCIc/MCInc	SEN	44.20 ± 15.22 [22–70]	62.16	+17.96	2
	SPE	76.47 ± 8.77 [61–91]	59.70	−16.77	16
	PPV	51.33 ± 7.25 [39–67]	46.00	−5.33	12
	NPV	68.18 ± 5.02 [66–79]	74.07	+5.89	7

Values in "other methods", "Voxel-MARS" and "difference" columns are given as percentages

Table 7 Our method compared with five methods on AD/CN classification

ID	Method name	SEN (%)	SPE (%)	PPV (%)	NPV (%)
0	Voxel-MARS	83.58	86.42	83.58	86.42
1.5.1 a	Voxel-COMPARE-D-gm	82	89	86	86
1.1.1 a	Voxel-Direct-D-gm	81	95	93	86
1.4.1 b	Voxel-Atlas-D-all	81	90	87	85
2.2	Thickness-Atlas	79	90	87	84
1.4.1 a	Voxel-Atlas-D-gm	78	93	90	83

Table 8 Our method compared with five methods on MCI/CN classification

ID	Method name	SEN (%)	SPE (%)	PPV (%)	NPV (%)
0	Voxel-MARS	78.38	88.89	76.32	90.00
1.3.1 a	Voxel-STAND-D-gm	73	85	69	87
3.1.1	Hippo-Volume-F	73	74	56	86
3.1.2	Hippo-Volume-S	70	73	54	84
1.4.2 a	Voxel-Atlas-S-gm	68	95	86	87
2.3	Thickness-ROI	65	94	83	85

An overall quantitative comparison of the performance of our method with other methods is introduced in Table 6. Performance statistics of other methods are presented in the format: "average (%) ± standard deviation (%) [range (%)]". In both of the AD/CN and AD/MCI classification cases, all 28 methods had produced reasonable results, whereas in MCIc/MCInc classification case, only 15 of them had. Therefore, for the 3rd case, 13 methods producing "zero sensitivity" were not included in computations. The column "Difference" compares the results gathered using our method and the averages of other methods as percentage.

Springer

PERFORMANCE MEASURES

		Actual Classes	
		POSITIVE (DISEASED)	NEGATIVE (HEALTHY)
Predictions	POSITIVE (DISEASED)	True Positive (TP)	False Positive (FP)
	NEGATIVE (HEALTHY)	False Negative (FN)	True Negative (TN)

- **ACCURACY:**
  - $(TP + TN)/(TP + TN + FP + FN)$
- **SENSITIVITY (RECALL):**
  - $TP/(TP + FN)$
- **SPECIFICITY:**
  - $TN/(TN + FP)$
- **PPV (PRECISION):**
  - $TP/(TP + FP)$
- **NPV:**
  - $TN/(TN + FN)$
- **AUC**

PARAMETER OPTIMIZATION
- *N*-Times Replicated *k*-Fold Cross-Validation

AD/CN: 150, **MCI/CN**: 120,

MCIc/MCInc: 104 training samples.

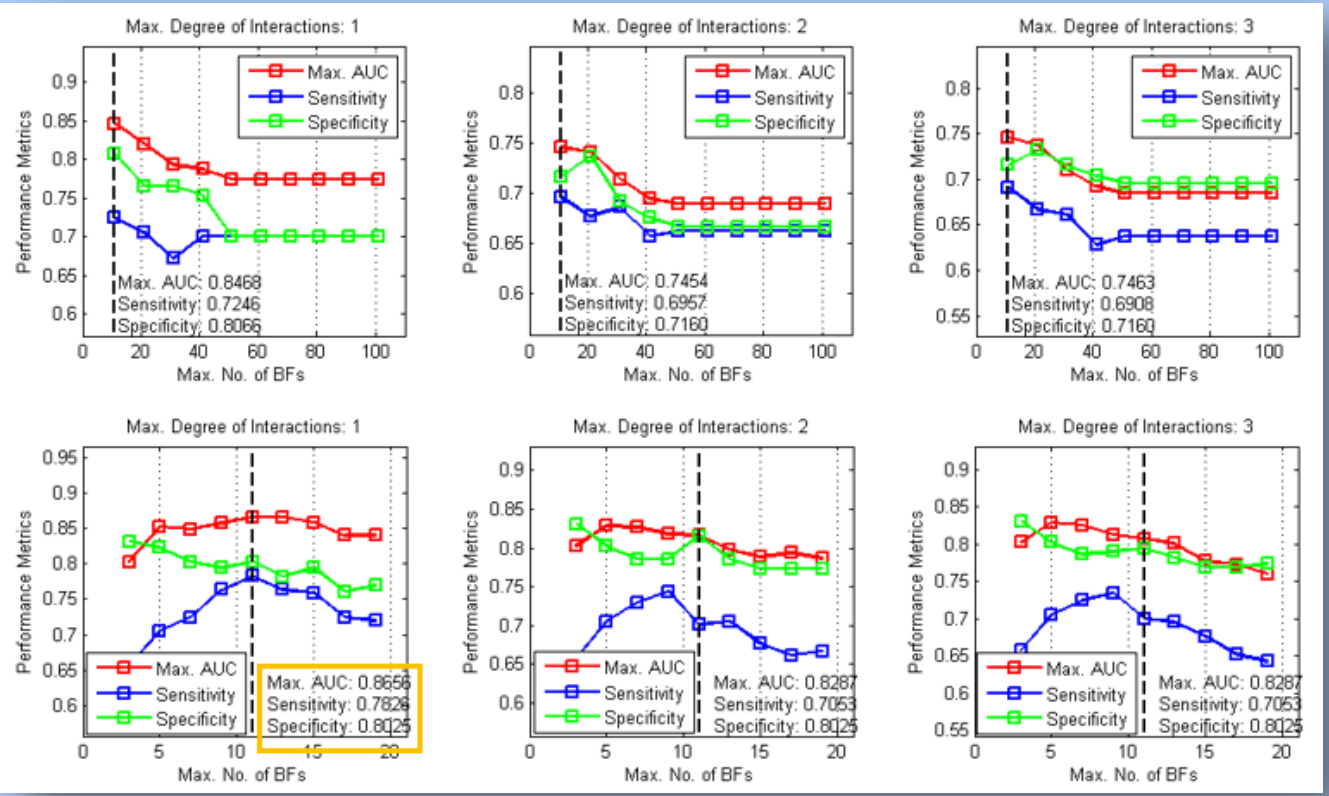
To keep training samples constant at each iteration:

$$SAMPLE\ SIZE\ \times\ (k-1)/k\ \approx\ 100.$$

To keep total repetition at each iteration constant:

Keep $n \times k$ constant.

AD/CN: n = 18, k = 3, **MCI/CN**: n = 9, k = 6, **MCIc/MCInc**: n = 3, k = 18



Coarse:

$M_{max} \in \{11, 21, \ldots, 101\}$

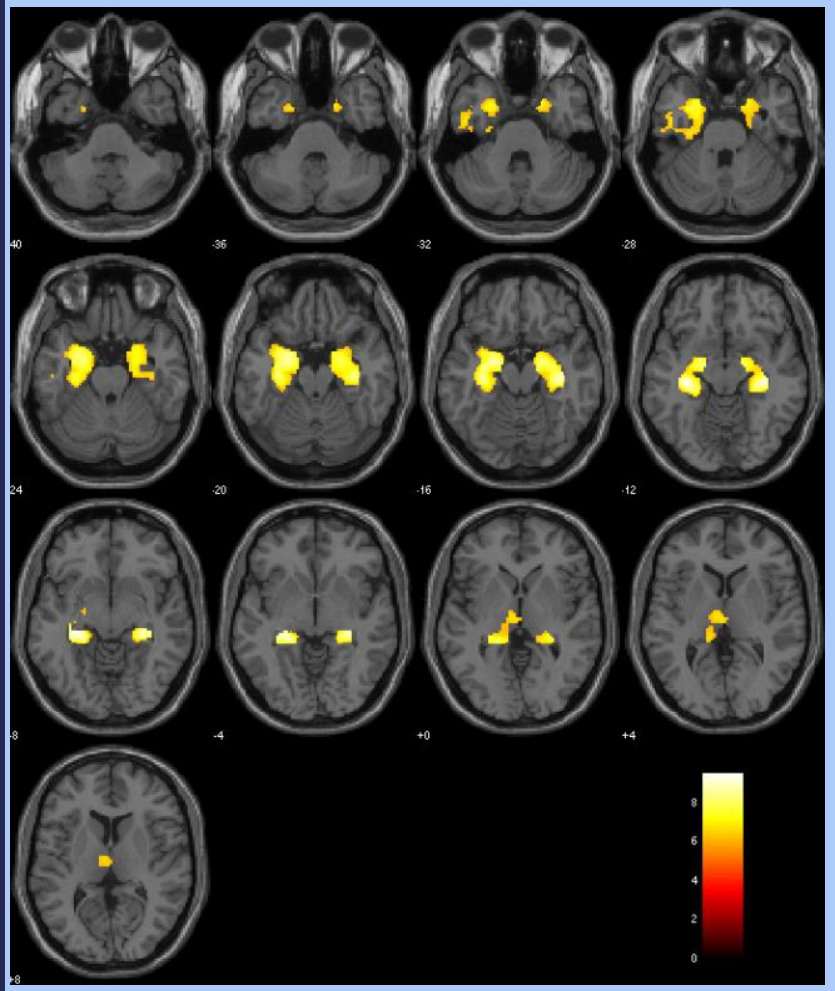
$K_{max} \in \{1, 2, 3\}$

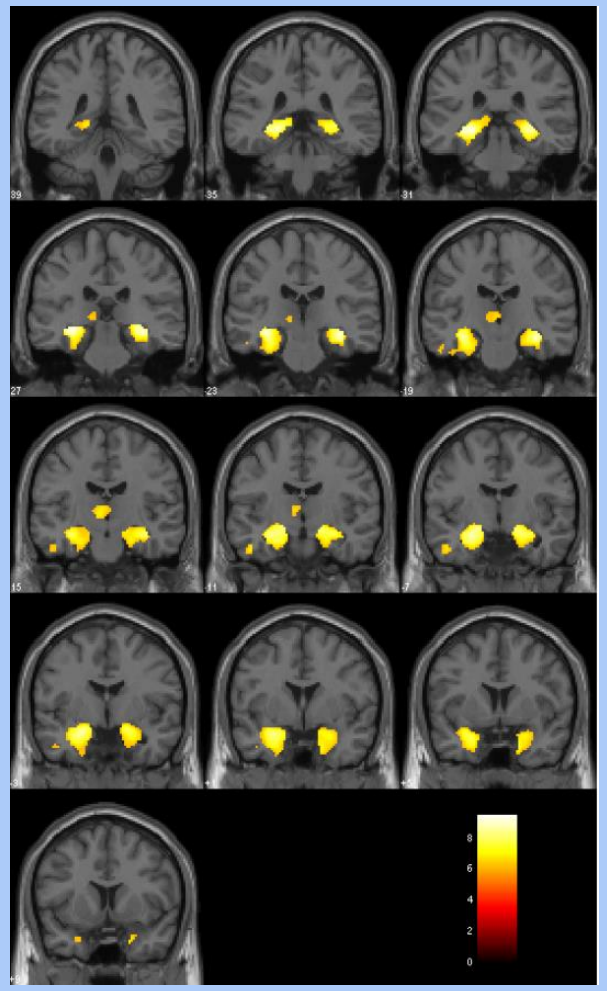
Fine:

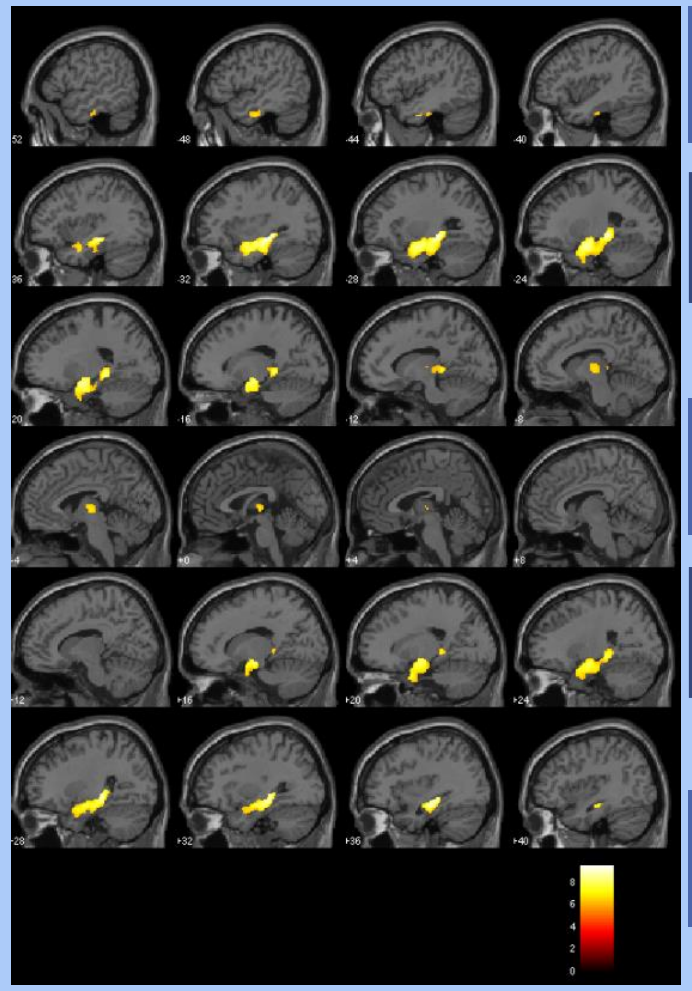
$M_{max} \in \{M_{max1} - 8, \ldots, M_{max1} - 2, M_{max1}, M_{max1} + 2, M_{max1} + 8\}$

$K_{max} \in \{1, 2, 3\}$

Dimensionality Reduction
- Results







6,368,835 variables

Zero-Voxel Elimination

2,112,054 variables

Feature Selection

3,320 variables

Dimensionality Reduction
- A Comparison with Commonly-Used Methods

Our procedure for dimensionality reduction is compared with other commonly used techniques, in terms of sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV) outcomes.

	SEN (%)	SPE (%)	PPV (%)	NPV (%)
Voxel-MARS	83.58	86.42	83.58	86.42
PCA	82.09	71.60	70.51	82.86
MDS	82.09	71.60	70.51	82.86
Laplacian Eigenmaps	79.10	74.07	71.62	81.08
Kernel PCA	2.99	95.06	33.33	54.23
Diffusion Maps	91.04	1.24	43.26	14.29
GDA	0	100	-	54.73
None	67.16	81.48	75.00	75.00

Classification with MARS (1/3)
- AD/CN Case

Performance of Voxel-MARS in **AD/CN** classification is compared to those of the 5 highest-ranking methods in terms of sensitivity. **Voxel-MARS** provides the **highest sensitivity outcome**.

Method ID	Method Name	SEN	SPE	PPV	NPV
0	Voxel-MARS	83.58%	86.42%	83.58%	86.42%
1.5.1 a	Voxel-COMPARE-D-gm	82%	89%	86%	86%
1.1.1 a	Voxel-Direct-D-gm	81%	95%	93%	86%
1.4.1 b	Voxel-Atlas-D-all	81%	90%	87%	85%
2.2	Thickness-Atlas	79%	90%	87%	84%
1.4.1 a	Voxel-Atlas-D-gm	78%	93%	90%	83%

Classification with MARS
- A Comparison to the Average Success Rates

	AD/CN			
	SEN	SPE	PPV	NPV
AVERAGE	71.46%	89.39%	85.18%	78.93%
STD. DEV.	5.65%	5.03%	6.23%	3.88%
MARS	**83.58%**	**86.42%**	**83.58%**	**86.42%**
DIFF.	12.12%	-2.97%	-1.60%	7.49%

Feature Extraction Method: VBM
- A Comparison with Other Feature Descriptors

ACCURACY	SIFT*	HOG**	Voxel-MARS
Our Implementation	0.71	0.74	0.85
(Toews et al., 2010)	0.71	-	-
(Daliri, 2012)	0.75	-	-
(Chen et al., 2014)	0.74	-	-
(Cattell et al., 2016)	-	0.90*	-
(Ameer et al., 2017)	-	0.68	-
(Unay and Ekin, 2011)	-	0.74	-

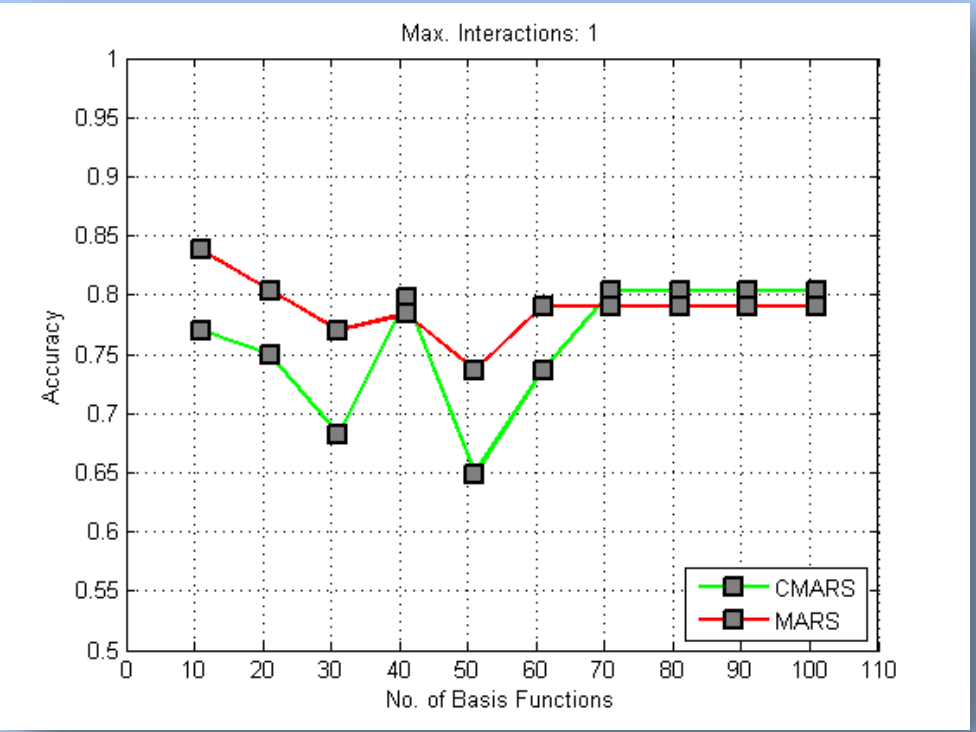
**SIFT*: Scale-Invariant Feature Transform
***HOG*: Histogram of Oriented Gradients

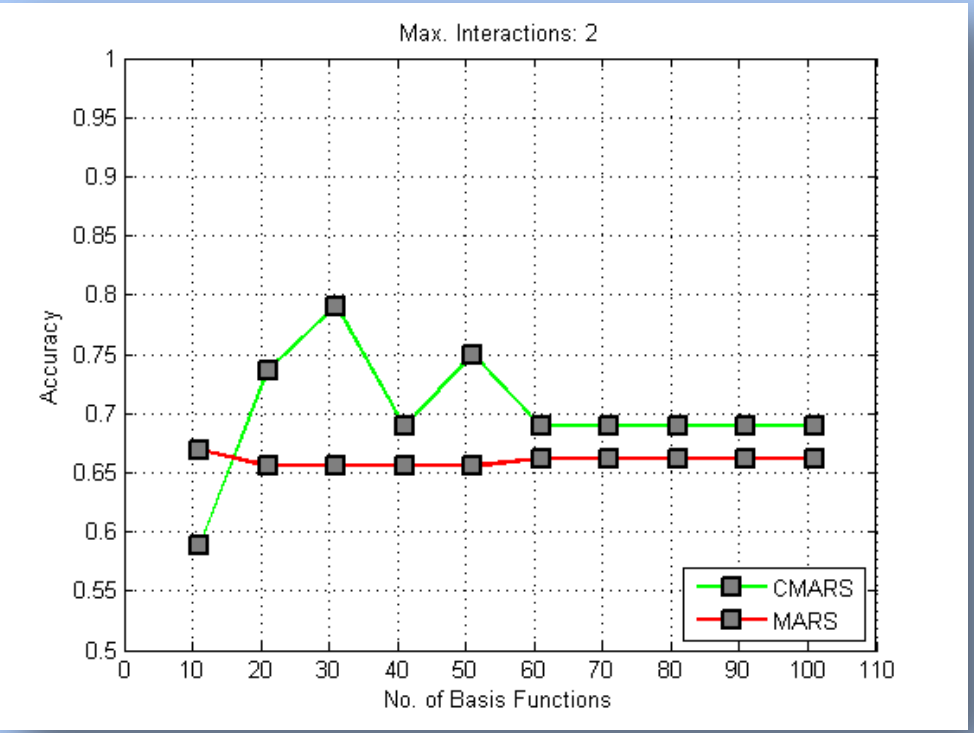
- Voxel-MARS outperforms feature descriptors in terms of accuracy.
- * (Cattell et al., 2014) does not involve early diagnosis, involves classification of amyloid status.

Classification with CMARS

Accuracy gathered by CMARS for varying parameter values. ($M_{max} \in \{11, 21, \dots, 101\}$; $K_{max} \in \{1, 2, 3\}$; $\widetilde{M} = 1$).

Higher accuracy in higher dimensions and higher degree of interactions.





MARS vs. CMARS

- Confusion matrices
- SEN and SPE outcomes

per classification group, acquired by **MARS & CMARS.**

CMARS provided higher **specificity** in:

- **AD/CN**
- **MCIc*/MCInc****

classification cases.

*MCIc: Mild Cognitive Impairment converting to AD,
**MCInc: MCI not converting to AD.

MARS		True Classes					
		AD/CN		MCI/CN		MCIc/MCInc	
		H	D	H	D	H	D
Predictions	Healthy	70	11	72	8	40	14
	Diseased	11	56	9	29	27	23
	TOTALS	81	67	81	37	67	37

CMARS		True Classes					
		AD/CN		MCI/CN		MCIc/MCInc	
		H	D	H	D	H	D
Predictions	Healthy	71	20	67	13	42	16
	Diseased	10	47	14	24	25	21
	TOTALS	81	67	81	37	67	37

		AD/CN	MCI/CN	MCIc/MCInc
MARS	SEN (%)	83.58	78.38	62.16
	SPE (%)	86.42	88.89	59.70
CMARS	SEN (%)	70.15	64.86	56.76
	SPE (%)	87.65	82.72	62.69

CONCLUSIONS

- Conclusions
- Future Outlook

CONCLUSIONS
- Contributions

- A solid basis for a **fully-automated Computer-Aided Diagnosis system** for early AD diagnosis is built.
- Qualitative and quantitative comparison between **VBM** and **Feature Descriptor-based approach** is made. VBM is shown to be **more effective in early diagnosis.**
- A novel, 3-step, hybrid **Dimensionality Reduction procedure** employing both **Statistical Analysis** and **Domain Knowledge** is developed. Proposed method **outperformed** commonly-used space-transforming methods.
- **MARS** and **CMARS** methods are utilized for classification of medical images for the **first time in the literature.** Very **successful results** are obtained, especially in terms of **sensitivity.**
- A **flexible codebase** and a **high-quality dataset** is ready for further use.
- Our paper, ***Voxel-MARS*** was published in ***ANOR, SI: OR in Neuroscience.***

A FUTURE OUTLOOK

Planned work

- The research on MARS and CMARS will be extended to include very recent variants of these methods, e.g., **RMARS, RCMARS, RCGPLM.**

Potential future directions

- Methodology may be extended to produce **probabilities rather than class labels.**
- The procedure may be extended to cover **other neurodegenerative diseases.**
- Dataset may be enriched to investigate **multimodal feature performance.**
- Dataset may be enriched to investigate possiblity of utilizing **Deep Learning** methods.

REFERENCES

- A. Çevik, G.-W. Weber, B. M. Eyüboğlu, and K. K. Oğuz. Voxel-MARS: a method for early detection of Alzheimer's disease by classification of structural brain MRI. *Annals of Operations Research*, Feb 2017.
- A. Çevik, Computer-Aided Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment With MARS/CMARS Classification Using Structural MR Images, Thesis (Ph.D.), Graduate School of Natural and Applied Sciences, METU, Sep 2017.
- R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766 – 781, 2011.

Thank you very much for your interest.

Gerhard-Wilhelm Weber, Poznan University of Technology,
gerhard.weber@put.poznan.pl
Alper Çevik, Middle East Technical University,
alper.cevik@metu.edu.tr