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

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

Conclusions
• Conclusions
• Future Outlook

Performance
• Results
• Discussion
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)
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 (Cuinnet 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
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
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 \(t = 0.5\)

Model function:
\[
\hat{f}(x) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(x)
\]

Basis functions:
\[
B_i(x) = \prod_{j=1}^{K_m} (s_{k_{jm}}(x_{k_{jm}} - \tau_{k_{jm}}))_+
\]

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


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.

\[
\begin{align*}
0.541 & \quad \text{ACC = 0.6689} \\
-0.317 \times \max(0, x[2561] - 0.497) & \quad \text{1} \\
-1.29 \times \max(0, 0.497 - x[2561]) & \quad \text{2} \\
+0.247 \times \max(0, x[994] - 0.0838) \times \max(0, 0.497 - x[2561]) & \quad \text{3} \\
+0.691 \times \max(0, 0.0838 - x[994]) \times \max(0, 0.497 - x[2561]) & \quad \text{4} \\
-4.72 \times \max(0, 0.497 - x[2561]) \times \max(0, x[2567] - 0.0715) & \quad \text{5} \\
+0.11 \times \max(0, x[912]) \times x[2567] & \quad \text{6} \\
+1.88 \times \max(0, x[2542] - 0.226) \times \max(0, 0.497 - x[2561]) & \quad \text{7} \\
+0.198 \times \max(0, x[392] - 0.0626) \times \max(0, 0.497 - x[2561]) & \quad \text{8} \\
-0.626 \times \max(0, x[912] - 1.03) & \quad \text{9} \\
+0.0952 \times \max(0, 1.03 - x[912]) & \quad \text{10} \\
+0.201 \times \max(0, x[2155] - 0.574) & \quad \text{11} \\
+0.281 \times \max(0, -0.574 - x[2155]) & \quad \text{12} \\
+0.269 \times \max(0, x[2840] - 0.165) & \quad \text{13} \\
+0.604 \times \max(0, 0.165 - x[2840]) & \quad \text{14}
\end{align*}
\]

\[
\begin{align*}
0.467 & \quad \text{ACC = 0.6757} \\
+0.325 \times \max(0, x[2561] - 0.497) & \quad \text{0} \\
-1.14 \times \max(0, 0.497 - x[2561]) & \quad \text{1} \\
+0.213 \times \max(0, x[994] - 0.0838) \times \max(0, 0.497 - x[2561]) & \quad \text{2} \\
+0.625 \times \max(0, 0.0838 - x[994]) \times \max(0, 0.497 - x[2561]) & \quad \text{3} \\
+4.67 \times \max(0, 0.497 - x[2561]) \times \max(0, x[2567] - 0.0715) & \quad \text{4} \\
+1.81 \times \max(0, x[2542] - 0.226) \times \max(0, 0.497 - x[2561]) & \quad \text{5} \\
+0.579 \times \max(0, x[912] - 1.03) & \quad \text{6} \\
+0.107 \times \max(0, 1.03 - x[912]) & \quad \text{7} \\
+0.208 \times \max(0, x[2155] - 0.574) & \quad \text{8} \\
+0.258 \times \max(0, -0.574 - x[2155]) & \quad \text{9} \\
+0.295 \times \max(0, x[2840] - 0.165) & \quad \text{10} \\
+0.72 \times \max(0, 0.165 - x[2840]) & \quad \text{11}
\end{align*}
\]
CMARS - Conic Multivariate Adaptive Regression Splines

MARS - Backward Algorithm

- \( GCV := \frac{1}{N} \sum_{i=1}^{N} \left( y_i - f_\alpha(x_i) \right)^2 \)

CMARS

- \( PRSS := \sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \sum_{m=1}^{n_{\max}} \sum_{|\alpha|=1}^{2} \int_{r<s} \beta_m^2 \left[ D_{r,s}^\alpha B_m(t_m^r) \right]^2 dt_m \)

- \( PRSS \approx \| y - B(d) \beta \|_2^2 + \lambda \| \beta \|_2^2 \)

- Tikhonov Regularization
CMARS - Optimization

CQP Problem

minimize \( t, \beta \)

subject to

\[ \| y - B(\tilde{a})\beta \|_2 < t \]

\[ \| L\beta \|_2 \leq \sqrt{M}. \]

\[ y_1 - B(x_1)\beta = \beta_{M+1}, \]
\[ y_2 - B(x_2)\beta = \beta_{M+2}, \]
\[ \vdots \]
\[ y_N - B(x_N)\beta = \beta_{M+N}, \]

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

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

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

\( N \) Linear constraints

1 Conic constraint

\( M \) Linear constraints

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

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

ACC = 0.6689

```
0.541  
-0.317 * max(0, x[2561] - 0.497)  // 0
-1.29 * max(0, 0.497 - x[2561])  // 1
+0.247 * max(0, x[994] - 0.0838) * max(0, 0.497 - x[2561])  // 2
+0.691 * max(0, 0.0838 - x[994]) * max(0, 0.497 - x[2561])  // 3
-4.72 * max(0, 0.497 - x[2561]) * max(0, x[2967] - 0.0715)  // 4
+0.117 * max(0, 0.497 - x[2561]) * max(0, 0.0715 - x[2967])  // 5
+1.86 * max(0, x[2542] - 0.226) * max(0, 0.497 - x[2561])  // 6
+0.138 * max(0, -0.226 - x[2542]) * max(0, 0.497 - x[2561])  // 7
-0.626 * max(0, x[912] - 1.03)  // 8
+0.0952 * max(0, 1.03 - x[912])  // 9
+0.201 * max(0, x[2155] - 0.574)  // 10
+0.251 * max(0, -0.574 - x[2155])  // 11
+0.269 * max(0, x[2840] - 0.165)  // 12
+0.604 * max(0, 0.165 - x[2840])  // 13
```

CMARS

ACC = 0.7027

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

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

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

\[
P(\mathbf{y} | \mu, \sigma, \gamma) = \prod_{i=1}^{l} \left( \sum_{k=1}^{K} \frac{\gamma_k}{(2\pi \sigma_k^2)^{1/2}} \exp \left( - \frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right) \right)
\]

\[
- \log P(\mathbf{y} | \mu, \sigma, \gamma, \beta, \alpha) = - \sum_{i=1}^{l} \log \left( \frac{\rho_i(\beta)}{\sum_{k=1}^{K} \gamma_k b_{i,k}(\alpha)} \sum_{k=1}^{K} \frac{\gamma_k b_{i,k}(\alpha)}{(2\pi \sigma_k^2)^{1/2}} \exp \left( - \frac{(\rho_i(\beta)y_i - \mu_k)^2}{2\sigma_k^2} \right) \right)
\]
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, sagittal, 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*
Diffeomorphic Anatomical Registration Through Exponentiated Lie 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

\[ n = \frac{p}{3} = 2,122,945 \]
GM tissue probabilities

\[ n = \frac{p}{3} = 2,122,945 \]
WM tissue probabilities

\[ n = \frac{p}{3} = 2,122,945 \]
CSF tissue probabilities

\[ N=150 \]
class labels

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

\[ N = 150 \]
Instances

\[ p = 6,368,835 \] variables
(7.12 GB !)

Reduced by background elimination to:

\[ 2,112,054 \] variables
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.

\[ y = Xb + u \]

- \( y \): the vector of independent observations
- \( X \): contains the tissue probabilities as features
- \( b \): the vector of unknown parameters
- \( u \): 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. \( P_{GM}(i, j) + P_{WM}(i, j) < \tau_1 \), \( \forall i \in [1, 2, \ldots, n] \)
   
   *sum of probabilities of being gray matter and white matter is smaller than the first threshold, \( \tau_1 \)*

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, \( \tau_2 \)*

Optimum thresholds:

\( \tau_1 = 0.5 \) and \( \tau_2 = 0.7 \)
DIMENSIONALITY REDUCTION
- Step III: Within-Class Norm Thresholding

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

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

\[ \|P(j)\|_2 < \epsilon \mu_c \]

\[ \epsilon = 0.9 \]
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
DOI: 10.1007/s10479-017-2402-7

Online First

Abstract

Neuroscience is of emerging importance along with the contributions of Operative Research to the practice of diagnosing neurodegenerative diseases with computer-aided systems based on brain image analysis. Although 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 those diagnosed with AD and MCI. It was observed that our method outperformed all of the others in sensitivity (83.58%) in ADCN and 78.38% in ADMM.

Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (www.adni.loni.usc.edu). As such, 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.pdf.

Alper Civril

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

Selected terms and phrases in the text:
- Alzheimer’s Disease
- Magnetic Resonance Imaging (MRI)
- Multivariate Adaptive Regression Splines (MARS)
- Control Normal (CN)
- Alzheimer’s Disease (AD)
- Mild Cognitive Impairment (MCI)
- Feature Extraction
- Classification

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

<table>
<thead>
<tr>
<th>Case</th>
<th>Metric</th>
<th>Other Methods Used-in-Research</th>
<th>Difference (%)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHCN</td>
<td>SEN</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
<tr>
<td>MPN</td>
<td>PPV</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
<tr>
<td>PNG</td>
<td>NPV</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
<tr>
<td>AHCN</td>
<td>MCI</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
<tr>
<td>MPN</td>
<td>PPV</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
<tr>
<td>PNG</td>
<td>NPV</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7: Our method compared with five other methods in ADNCN classification

<table>
<thead>
<tr>
<th>Case</th>
<th>Metric</th>
<th>Other Methods Used-in-Research</th>
<th>Difference (%)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHCN</td>
<td>SEN</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
<tr>
<td>MPN</td>
<td>PPV</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
<tr>
<td>PNG</td>
<td>NPV</td>
<td>89.50 (55-65)</td>
<td>89.50</td>
<td>1</td>
</tr>
</tbody>
</table>

An overall 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 (90% confidence interval) [dynamic range (90%)]." In both of the ADCN and ADMM classification cases, all 12 methods produced reasonable results, whereas in MCI/MMC classification case, only 13 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.
### PERFORMANCE MEASURES

<table>
<thead>
<tr>
<th>Actual Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITIVE (DISEASED)</td>
</tr>
<tr>
<td>NEGATIVE (HEALTHY)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITIVE (DISEASED)</td>
</tr>
<tr>
<td>True Positive (TP)</td>
</tr>
<tr>
<td>False Positive (FP)</td>
</tr>
<tr>
<td>NEGATIVE (HEALTHY)</td>
</tr>
<tr>
<td>False Negative (FN)</td>
</tr>
<tr>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

- **ACCURACY:**
  \[ \frac{(TP + TN)}{(TP + TN + FP + FN)} \]

- **SENSITIVITY (RECALL):**
  \[ \frac{TP}{(TP + FN)} \]

- **SPECIFICITY:**
  \[ \frac{TN}{(TN + FP)} \]

- **PPV (PRECISION):**
  \[ \frac{TP}{(TP + FP)} \]

- **NPV:**
  \[ \frac{TN}{(TN + FN)} \]

- **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:

\[ \text{SAMPLE SIZE} \times (k - 1)/k \approx 100. \]

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:**
\( M_{\text{max}} \in \{11, 21, ..., 101\} \)
\( K_{\text{max}} \in \{1, 2, 3\} \)

**Fine:**
\( M_{\text{max}} \in \{M_{\text{max}1} - 8, ..., M_{\text{max}1} - 2, M_{\text{max}1}, M_{\text{max}1} + 2, M_{\text{max}1} + 8\} \)
\( K_{\text{max}} \in \{1, 2, 3\} \)
Dimensionality Reduction
- Results

- 6,368,835 variables
- Zero-Voxel Elimination
- 2,112,054 variables
- Feature Selection
- 3,320 variables
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.

<table>
<thead>
<tr>
<th>Method</th>
<th>SEN (%)</th>
<th>SPE (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voxel-MARS</td>
<td>83.58</td>
<td>86.42</td>
<td>83.58</td>
<td>86.42</td>
</tr>
<tr>
<td>PCA</td>
<td>82.09</td>
<td>71.60</td>
<td>70.51</td>
<td>82.86</td>
</tr>
<tr>
<td>MDS</td>
<td>82.09</td>
<td>71.60</td>
<td>70.51</td>
<td>82.86</td>
</tr>
<tr>
<td>Laplacian Eigenmaps</td>
<td>79.10</td>
<td>74.07</td>
<td>71.62</td>
<td>81.08</td>
</tr>
<tr>
<td>Kernel PCA</td>
<td>2.99</td>
<td>95.06</td>
<td>33.33</td>
<td>54.23</td>
</tr>
<tr>
<td>Diffusion Maps</td>
<td>91.04</td>
<td>1.24</td>
<td>43.26</td>
<td>14.29</td>
</tr>
<tr>
<td>GDA</td>
<td>0</td>
<td>100</td>
<td>-</td>
<td>54.73</td>
</tr>
<tr>
<td>None</td>
<td>67.16</td>
<td>81.48</td>
<td>75.00</td>
<td>75.00</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Method ID</th>
<th>Method Name</th>
<th>SEN</th>
<th>SPE</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Voxel-MARS</td>
<td>83.58%</td>
<td>86.42%</td>
<td>83.58%</td>
<td>86.42%</td>
</tr>
<tr>
<td>1.5.1 a</td>
<td>Voxel-COMPARE-D-gm</td>
<td>82%</td>
<td>89%</td>
<td>86%</td>
<td>86%</td>
</tr>
<tr>
<td>1.1.1 a</td>
<td>Voxel-Direct-D-gm</td>
<td>81%</td>
<td>95%</td>
<td>93%</td>
<td>86%</td>
</tr>
<tr>
<td>1.4.1 b</td>
<td>Voxel-Atlas-D-all</td>
<td>81%</td>
<td>90%</td>
<td>87%</td>
<td>85%</td>
</tr>
<tr>
<td>2.2</td>
<td>Thickness-Atlas</td>
<td>79%</td>
<td>90%</td>
<td>87%</td>
<td>84%</td>
</tr>
<tr>
<td>1.4.1 a</td>
<td>Voxel-Atlas-D-gm</td>
<td>78%</td>
<td>93%</td>
<td>90%</td>
<td>83%</td>
</tr>
</tbody>
</table>
### Classification with MARS
- A Comparison to the Average Success Rates

<table>
<thead>
<tr>
<th></th>
<th>AD/CN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEN</td>
</tr>
<tr>
<td><strong>AVERAGE</strong></td>
<td>71.46%</td>
</tr>
<tr>
<td><strong>STD. DEV.</strong></td>
<td>5.65%</td>
</tr>
<tr>
<td><strong>MARS</strong></td>
<td>83.58%</td>
</tr>
<tr>
<td><strong>DIFF.</strong></td>
<td>12.12%</td>
</tr>
</tbody>
</table>
### Feature Extraction Method: VBM
- A Comparison with Other Feature Descriptors

<table>
<thead>
<tr>
<th>ACCURACY</th>
<th>SIFT*</th>
<th>HOG**</th>
<th>Voxel-MARS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Implementation</td>
<td>0.71</td>
<td>0.74</td>
<td>0.85</td>
</tr>
<tr>
<td>(Toews et al., 2010)</td>
<td>0.71</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Daliri, 2012)</td>
<td>0.75</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Chen et al., 2014)</td>
<td>0.74</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Cattell et al., 2016)</td>
<td>-</td>
<td>0.90*</td>
<td>-</td>
</tr>
<tr>
<td>(Ameer et al., 2017)</td>
<td>-</td>
<td>0.68</td>
<td>-</td>
</tr>
<tr>
<td>(Unay and Ekin, 2011)</td>
<td>-</td>
<td>0.74</td>
<td>-</td>
</tr>
</tbody>
</table>

*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. ($M_{\text{max}} \in \{11, 21, \ldots, 101\}; K_{\text{max}} \in \{1, 2, 3\}; \tilde{M} = 1$).

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.

![Confusion matrices](image)

### MARS

<table>
<thead>
<tr>
<th>Predictions</th>
<th>True Classes</th>
<th>AD/CN</th>
<th>MCI/CN</th>
<th>MCIc/MCInc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>D</td>
<td>H</td>
</tr>
<tr>
<td>Healthy</td>
<td></td>
<td>70</td>
<td>11</td>
<td>72</td>
</tr>
<tr>
<td>Diseased</td>
<td></td>
<td>11</td>
<td>56</td>
<td>9</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td>81</td>
<td>67</td>
<td>81</td>
</tr>
</tbody>
</table>

### CMARS

<table>
<thead>
<tr>
<th>Predictions</th>
<th>True Classes</th>
<th>AD/CN</th>
<th>MCI/CN</th>
<th>MCIc/MCInc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>D</td>
<td>H</td>
</tr>
<tr>
<td>Healthy</td>
<td></td>
<td>71</td>
<td>20</td>
<td>67</td>
</tr>
<tr>
<td>Diseased</td>
<td></td>
<td>10</td>
<td>47</td>
<td>14</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td>81</td>
<td>67</td>
<td>81</td>
</tr>
</tbody>
</table>

### SEN and SPE outcomes

<table>
<thead>
<tr>
<th>Classification Group</th>
<th>MARS</th>
<th>CMARS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEN (%)</td>
<td>SPE (%)</td>
</tr>
<tr>
<td>AD/CN</td>
<td>83.58</td>
<td>86.42</td>
</tr>
<tr>
<td>MCI/CN</td>
<td>78.38</td>
<td>88.89</td>
</tr>
<tr>
<td>MCIc/MCInc</td>
<td>62.16</td>
<td>59.70</td>
</tr>
</tbody>
</table>

*MCIC: Mild Cognitive Impairment converting to AD, **MCInc: MCI not converting to AD.
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 possibility of utilizing Deep Learning methods.
REFERENCES


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