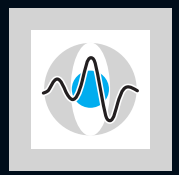
Temporal Development of the True Positive Rate of MRI Volumetric and Thickness Features During Alzheimer's Disease

Maya Angelova, Supervisor: Klaus Sembritzki
21 October 2013
Pattern Recognition Lab (CS 5)



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Motivation

- Alzheimer's Disease is the most common form of dementia
- Affects almost 50% of those over the age of 85
- The sixth leading cause of death in the US
- Early detection and intervention is desirable

Mild Cognitive Impairment[1]

- Transitional state between normal aging and early dementia
- Variably defined - distinction between normal ageing and MCI and very early dementia can be quite subtle and challenging
- Core clinical criteria for the diagnosis of MCI[2]
  - Concern regarding a change in cognition in comparison with the person's previous level
  - Impairment in one or more cognitive domains such as memory, language, etc.
  - Preservation of independence in functional abilities in everyday life
  - Not demented

[2] The diagnosis of mild cognitive impairment due to Alzheimers' disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Albert, 2011

[1] Mild cognitive impairment as a diagnostic entity, Petersen, 2004

Alzheimer's Disease[2]

- Core clinical criteria for probable AD
  - Interfere with everyday life
  - Represent a decline from previous levels of functioning and performing with an insidious onset
  - Are not explained by delirium or major psychiatric disorder
  - Cognitive impairment is detected in
    - Amnestic presentation
    - Nonamnestic presentation - language, visualspatial, executive
  - Evidence for other dementia such as VaD, FTD is ruled out

[2] The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, McKhan, 2011

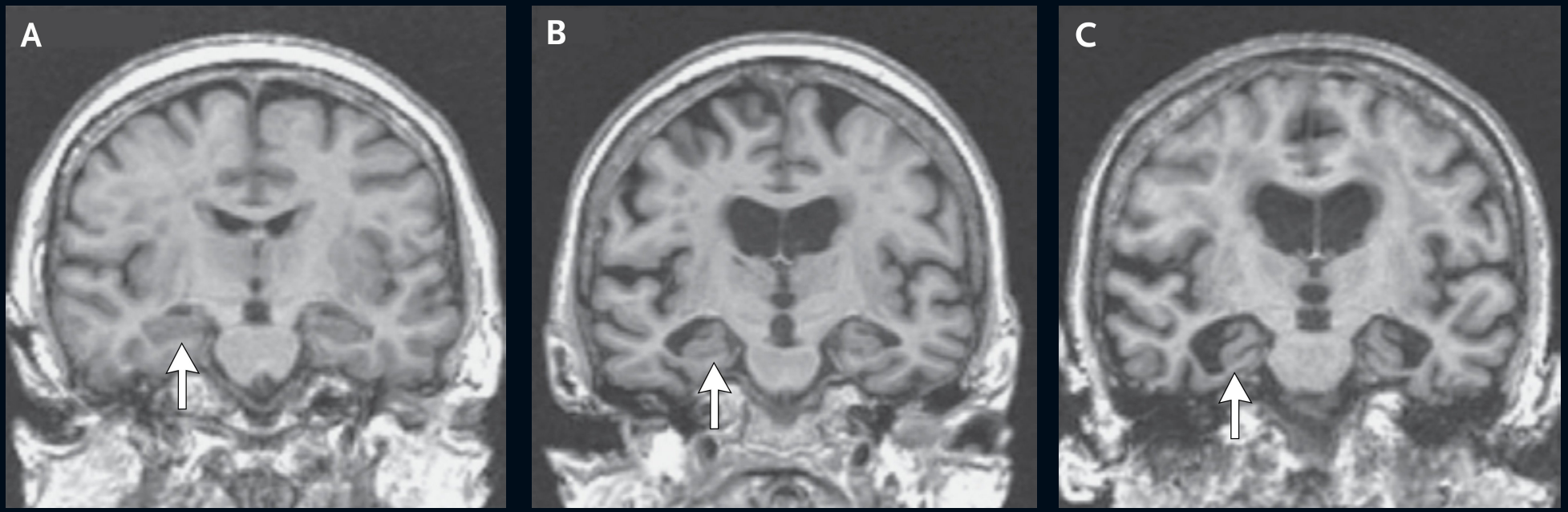
Biomarkers of Alzheimer's Disease[3]

- A physiologic, biochemical, or anatomic parameter that can be objectively measured as an indicator of biologic processes

- **Biomarkers of Aβ accumulation**
  - Abnormal radiotracer retention on amyloid PET images and
  - Low levels of CSF Aβ42
- **Biomarkers of neuronal degeneration or injury**
  - Elevated levels of CSF tau
  - Decreased FDG uptake on PET images
  - Atrophy on structural MR images

[3] Alzheimer disease: new concepts on its neurobiology and the clinical role imaging will play, Jack, 2012

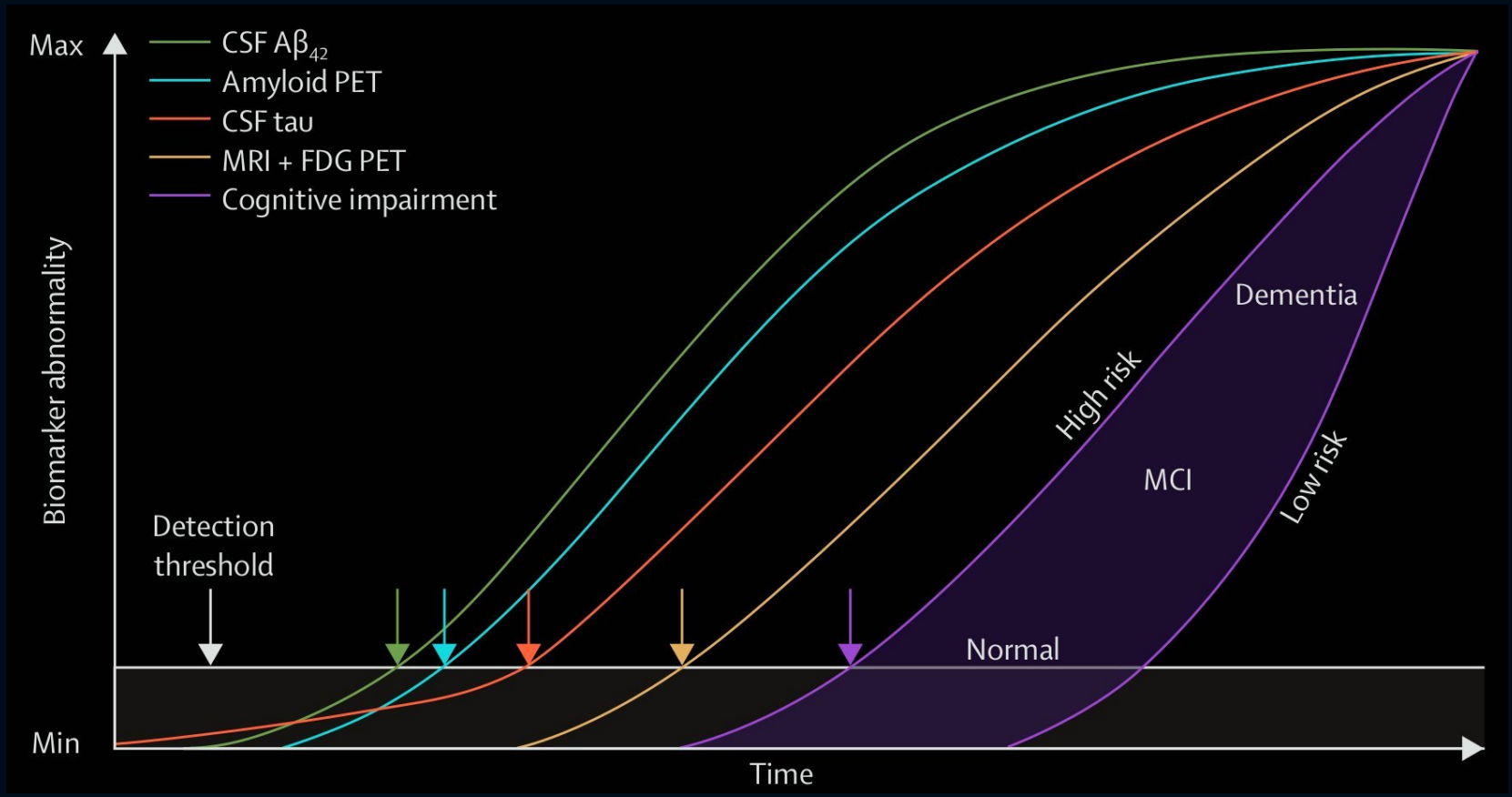
MR Image as Biomarker for Alzheimer's Disease



- Hippocampal formation and its progressive atrophy from normal cognition (Panel A) to MCI (Panel B) to Alzheimer's Disease (Panel C)[4]

[4] Mild cognitive impairment, Petersen, 2011

An updated hypothetical model of dynamic biomarkers, Jack, 2013



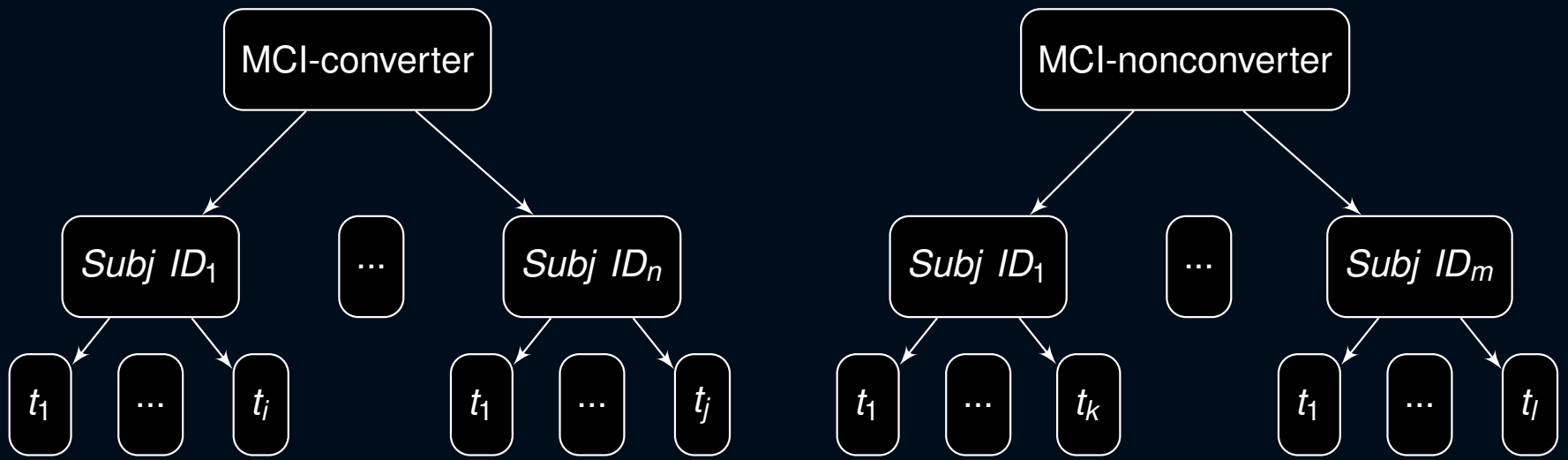
Goals of This Thesis

- Use of longitudinal subset from the ADNI[1] cohort comprising MR images from MCI, AD and Healthy Controls (HC) patients
- Choice of one or more structural MRI features suitable for an early diagnosis of Alzheimer's disease
- Temporal development of the sensitivity of MRI volumetric and thickness features
- Visual representation of volumetric changes over time

[1] http://www.adni-info.org/

Hierarchical data set

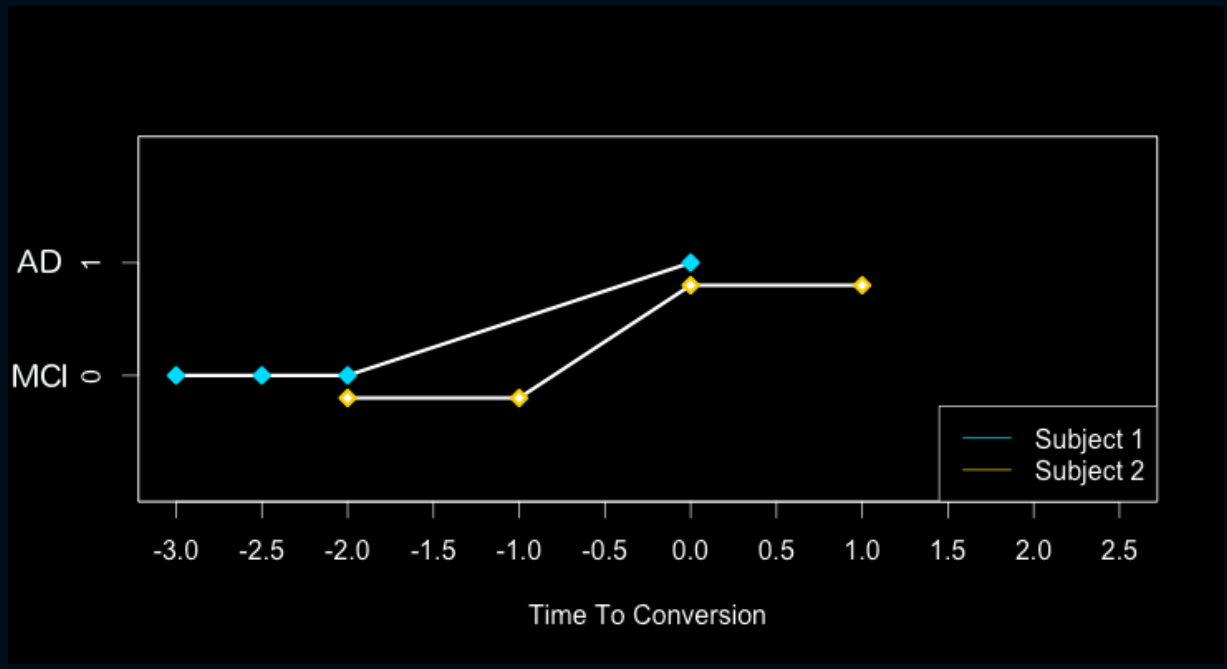
- Longitudinal subset from the ADNI[2] cohort, comprising MR images from 139 HC, 84 AD, 160 MCI patients



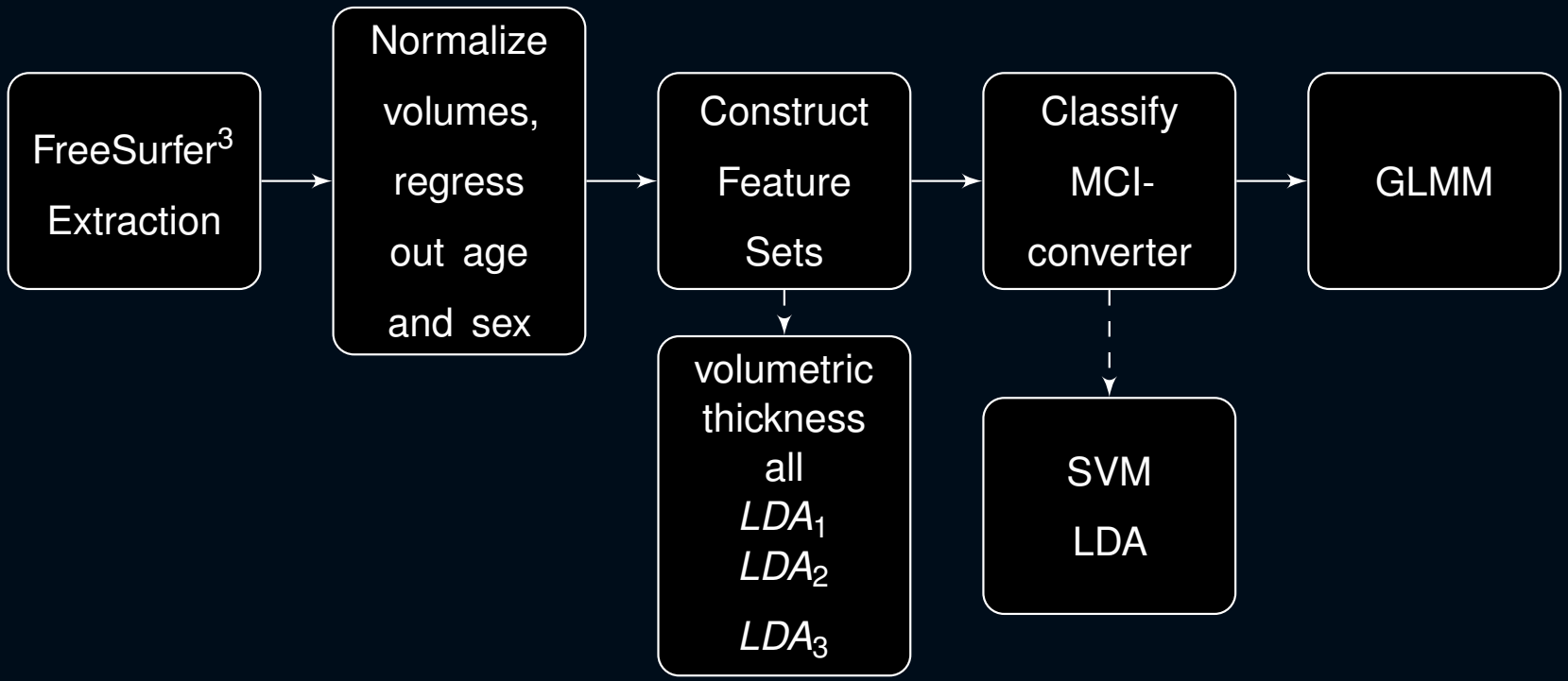
[2] http://www.adni-info.org/

Data

- Each subject has at least two and at most five MRI measurements over time – one at baseline, one at six months, at one year, at two and three years
- X-axis represents Time To Conversion and y-axis – Diagnosis



Overview



[3] https://surfer.nmr.mgh.harvard.edu/fswiki

LDA Feature Sets

- LDA_1 – features gained and used by McEvoy and colleagues [4]
  - Feature selection performed on *58 chosen ROIs* with a stepwise LDA using leave-one-out cross validation (CV) on data from HC and AD subjects
- LDA_2
  - Feature selection performed on *the 58 chosen ROIs from McEvoy and colleagues* with a stepwise LDA using leave-one-out CV on data from HC and AD subjects
- LDA_3
  - Feature selection performed on *all extracted FreeSurfer measurements* with a stepwise LDA using leave-one-out CV on data from HC and AD subjects

[4] Quantitative Structural Neuroimaging for Detection and Prediction of Clinical and Structural Changes in Mild Cognitive Impairment, McEvoy, 2009

Generalized Linear Mixed Models (GLMM)

- General Linear Model

$$y = \mu = \beta_0 + \beta_1 x_1 + ... + \beta_q x_q + \varepsilon \tag{1}$$

y – outcome, dependent variable, $y \sim \mathcal{N}(\mu, \sigma^2)$;
$x_1 ... x_q$ – predictors, independent variables;
β_0 – intercept, overall mean;
$\beta_1 ... \beta_q$ – regression coefficients;

ε – residual, $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

Generalized Linear Mixed Models (GLMM)

- Generalized Linear Model

$$g(\mu) = \ell = \beta_0 + \beta_1 x_1 + ... + \beta_q x_q + \varepsilon$$
$$y = h(\ell) + \varepsilon \tag{2}$$

$\ell = \sum_{j=0}^{q} \beta_j x_j$ – the systematic or linear component of predictors;
the outcome y – the random component;
$g(\cdot)$ – link function – determines the relationship between the systematic and the random component, allows for modeling non-gaussian data;
$h(\cdot) = g(\cdot)^{-1}$ – inverse link function;

ε – error term

GLMM

Linear Mixed-Effect Model

- a.k.a. multilevel regression models, hierarchical linear models, random coefficient models

$$y = \mathbf{X}\beta + \mathbf{Z}\boldsymbol{u} + \varepsilon \tag{3}$$

y – the outcome variable;
$\mathbf{X}$ – matrix of predictor variables; β – fixed-effects regression coefficients;
$\mathbf{Z}$ – design matrix for the random effects; $\boldsymbol{u}$ – random-effects coefficients, $\boldsymbol{u} \sim \mathcal{N}(0, \mathbf{G})$
ε – residual term, $\varepsilon \sim \mathcal{N}(0, \mathbf{I}\sigma_\varepsilon^2)$

  - Fixed effect
    - Influence only the mean of the outcome
    - To be estimated from the data
  - Random effect
    - Impact only on the variance of the outcome
    - To be predicted from the population from where the sample is drawn

GLMM

Linear predictor

$$\ell = \mathbf{X}\beta + \mathbf{Z}\boldsymbol{u} + \varepsilon \tag{4}$$

Link function

$$g(\cdot) = g(E(y)) = \ell \tag{5}$$

General equation

$$g(\mu) = \ell = \mathbf{X}\beta + \mathbf{Z}\boldsymbol{u} + \varepsilon \tag{6}$$

Inverse link function

$$h(\cdot) = g^{-1}(\cdot) = E(y) = h(\ell) = \mu \tag{7}$$

Outcome y of the linear predictor

$$y = h(\ell) + \varepsilon \tag{8}$$

GLMM

GLMM defined for individual *i* at time point *j*

$$g(\mu) = \ell_i = \beta_0 + \beta_1 t_{ij} + u_i + \varepsilon_{ij} \tag{9}$$

β_0 – intercept
β_1 – fixed effect coefficient
t_{ij} – fixed effect, Time To Conversion
$u_i \sim \mathcal{N}(0, \sigma_u^2)$ – random effect, subjects
$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ – residual

Logit link function

$$g(\mu) = log_e(\frac{\mu}{1-\mu}) \tag{10}$$

Inverse logit

$$h(\cdot) = \frac{e^{\cdot}}{1+e^{\cdot}} \tag{11}$$

GLMM

- Models variation at different levels of a data set
- Appropriate when the data are not independent with equal variances, as in repeated measurements taken on one individual
- Models non-gaussian data
- Handles unbalanced data

Fitting GLMM with MCMCglmm R package[5]

Bayesian inference

- Frequentist vs Bayesian approach
- Goal: estimation of parameter (set) θ
- Combines quantifying existing knowledge (using a prior) with evidence from new data (using the likelihood) by applying Bayes' theorem and updating the posterior

$$p(\theta \mid X) \propto \ell(\theta \mid X)p(\theta) \tag{12}$$

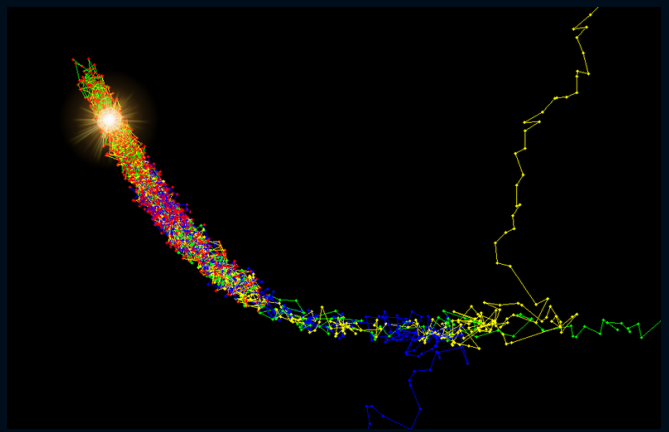
- Posterior distribution is then used to obtain inference about θ
- Informative vs diffuse prior
- High posterior density interval – shortest interval calculated to have i.e. 95% probability content

[5] http://cran.r-project.org/web/packages/MCMCglmm/index.html

Fitting GLMM with MCMCglmm R package[6]

Markov chain Monte Carlo (MCMC)

- Used to determine parameters and their posterior distributions
- Generate posterior distributions by sampling likelihood function, computed based on data, in parameter space
- Future state is dependent only on its current state; markov chain property



[6] http://cran.r-project.org/web/packages/MCMCglmm/index.html

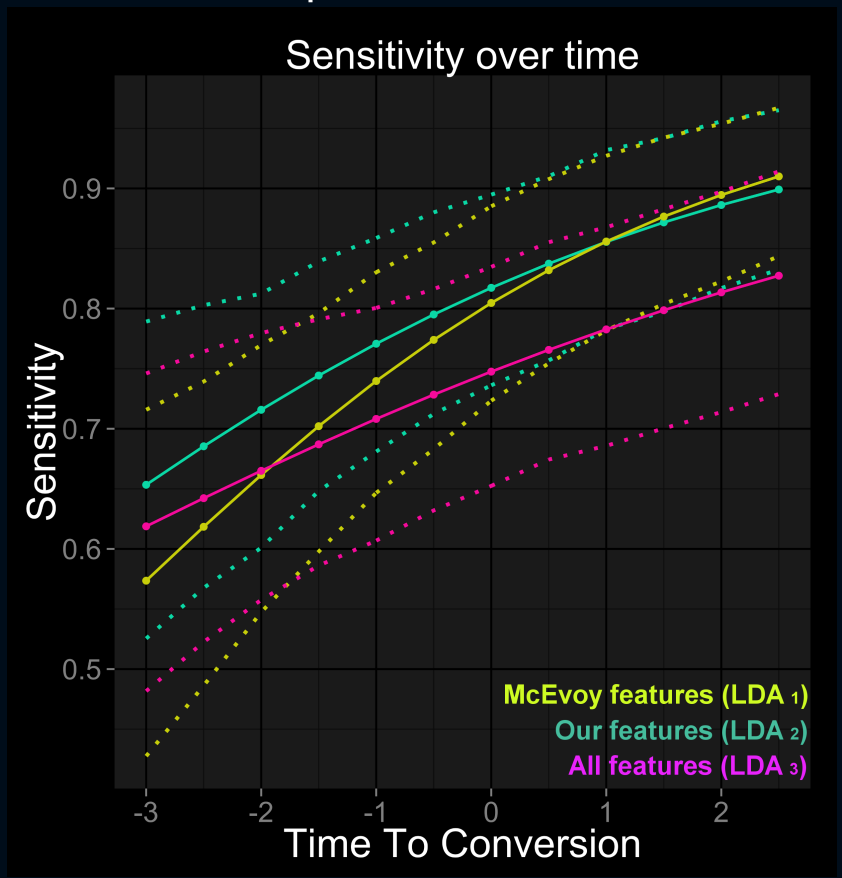
Questions

- Comparison with McEvoy and colleagues [7] when using LDA classifier
- Is feature selection performed with LDA criterion necessary?
- Is there a difference in the performance of the SVM classifier trained on AD/HC patients in using the three different kinds of measurements – volumetric, thickness, all (volumetric and thickness)?
- When trained on MCI baseline measurements if there is any observable difference in the performance of the SVM classifier using the three different kinds of feature sets from above?
- Are there differences across the four subjects' groups in hippocampal temporal deterioration – AD, HC, MCI-converter and MCI-nonconverter?

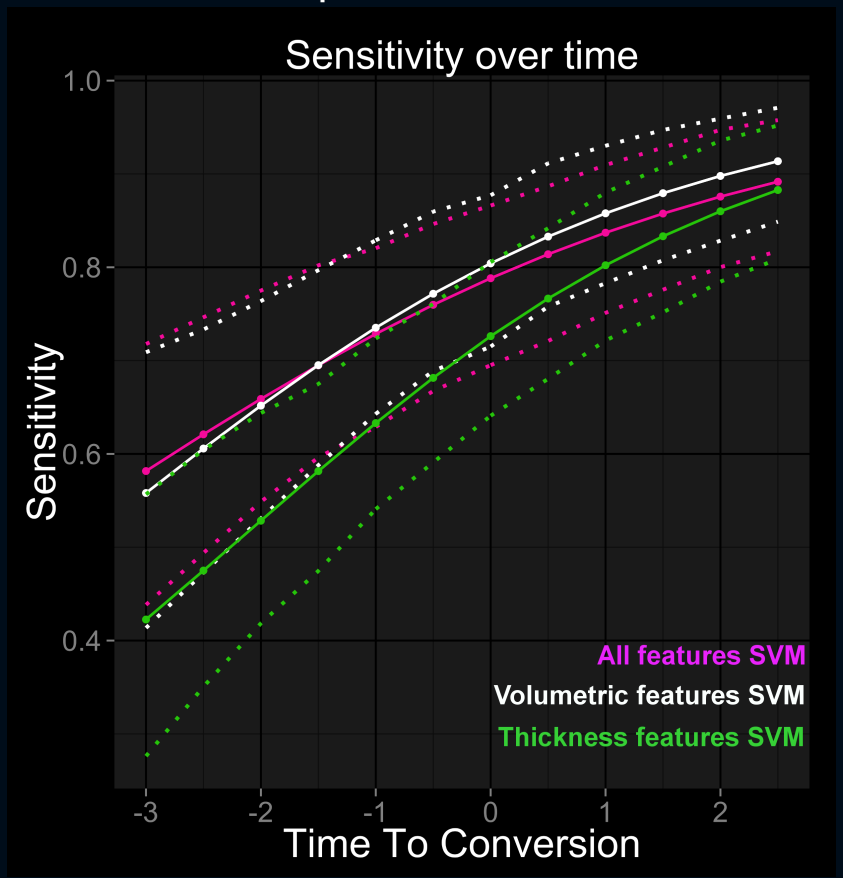
[7] Quantitative Structural Neuroimaging for Detection and Prediction of Clinical and Structural Changes in Mild Cognitive Impairment, McEvoy, 2009

Results LDA vs SVM

LDA classifier performance over time

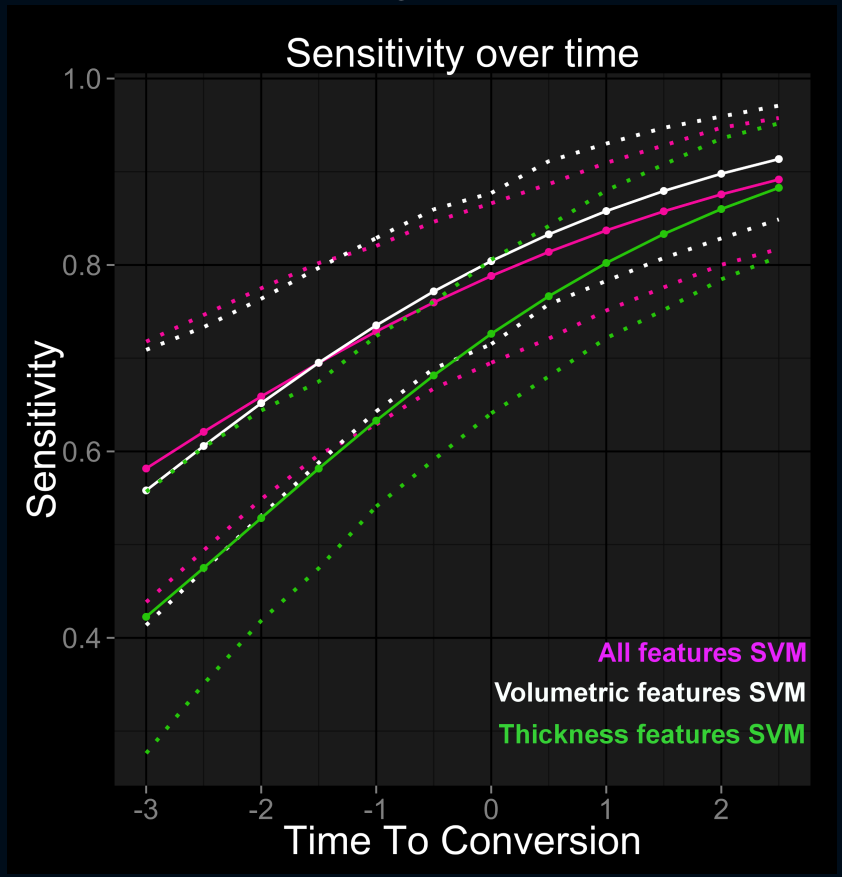


SVM classifier performance over time

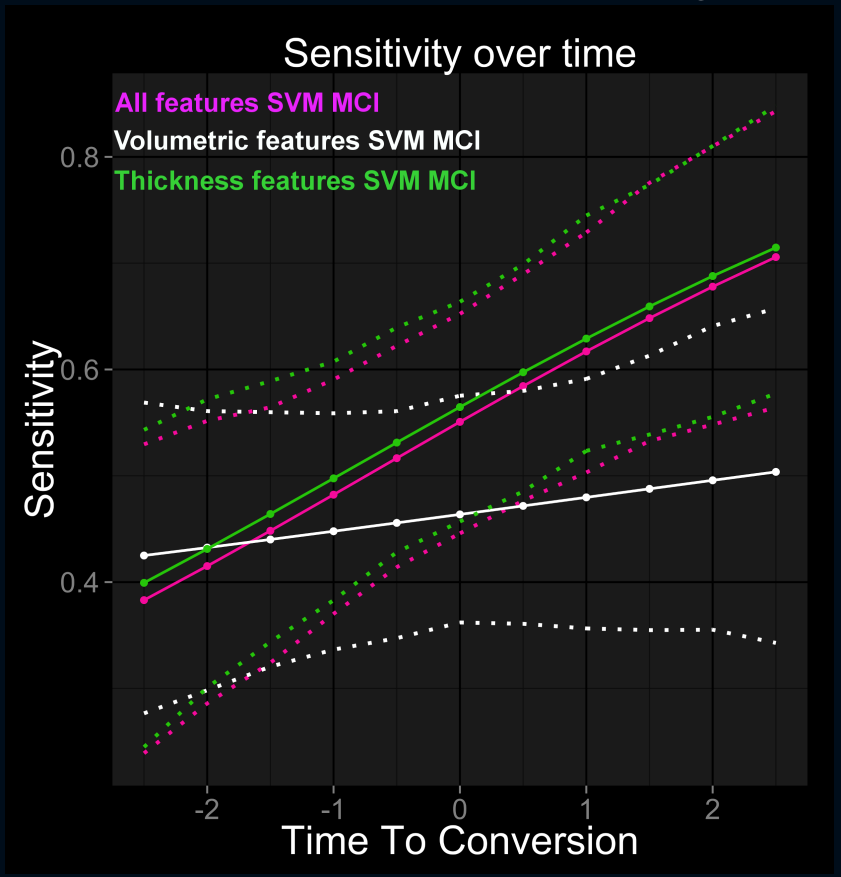


Results SVM AD/HC vs MCI Baseline training set

SVM classifier performance over time with AD/HC training set



SVM classifier performance over time with MCI Baseline traning set



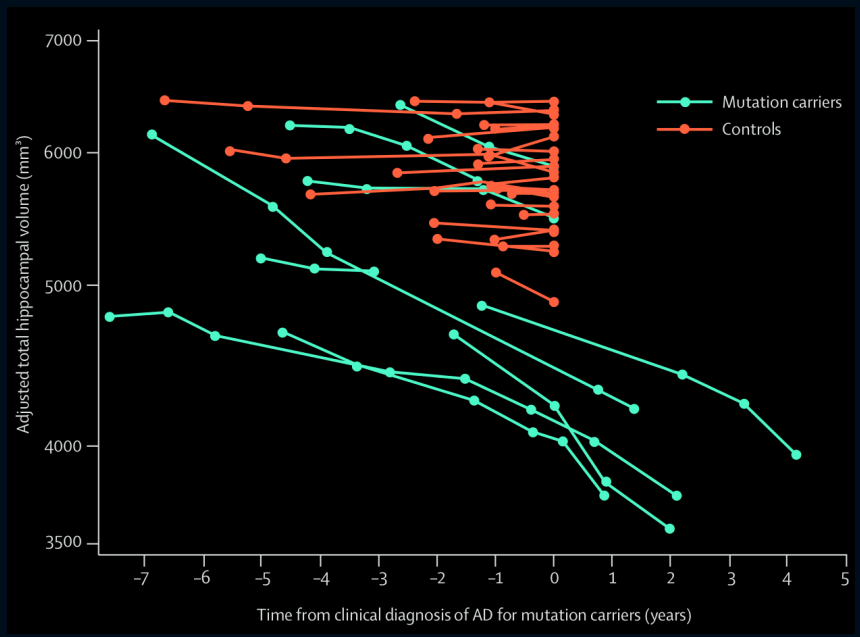
Previous work and comparison

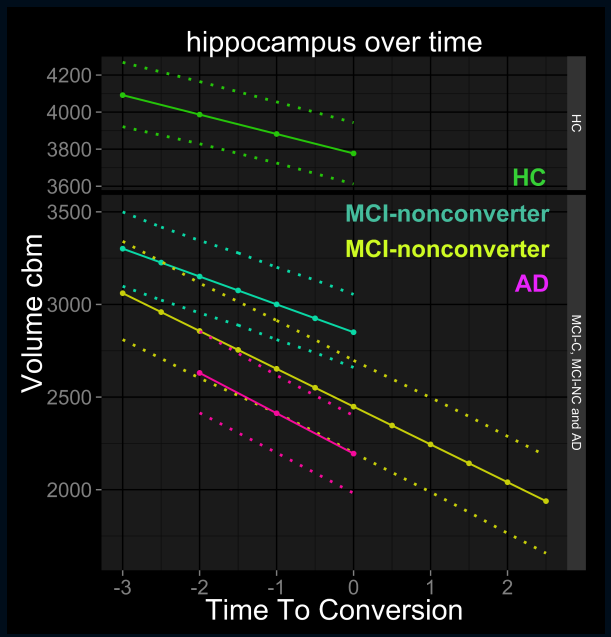
Quantitative Structural Neuroimaging for Detection and Prediction of Clinical and Structural Changes in Mild Cognitive Impairment, McEvoy, 2009

- Source of comparison
- Data used: ADNI cohort
- No editing performed, but same classification results
- LDA as feature selection proved unnecessary
- No temporal development of the analyzed features by McEvoy and colleagues
- No evaluation of classifier performance over time

Previous work and comparison

Tracking atrophy progression in familial Alzheimer's disease: a serial MRI study, Ridha, 2006





Summary

- Longitudinal MRI data set, automatically segmented
- The extracted features were normalized and age and sex effects regressed out
- None of the designated features did come up to the criteria of a good predictor for early detection of AD
- No feature selection with LDA needed
- When trained with AD/HC the SVM classifier with all kinds of measurements performs poorly years before conversion, achieving up to 90% sensitivity several years after conversion
- When trained on MCI patients the SVM classifier performed poorly with all kinds of measurements

Outlook

Random Intercept vs Random Slope Intercept GLMM

- Random Intercept model
  - Equal covariances across individuals

$$g(\mu) = \ell_i = \beta_0 + \beta_1 t_{ij} + u_i + \varepsilon_{ij} \tag{13}$$

- Random Intercept and Slope Model
  - Extension to Random Intercept Model
  - Non equal covariances across individuals

$$g(\mu) = \ell_i = \beta_0 + \beta_1 t_{ij} + u_{0i} + u_{1j} t_{ij} + \varepsilon_{ij} \tag{14}$$

where the slope of t is explicitly modeled as $\beta_{1j} = \beta_1 + u_{1j}$ with $u_{0i} \sim \mathcal{N}(0, \sigma^2_{u0})$ and $u_{1i} \sim \mathcal{N}(0, \sigma^2_{u1})$, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon})$ and $cov(u_{0i}, u_{1i}) = \sigma_{u01}$.

Thank you for your attention!