Machine Learning Algorithms for Neuroimaging-based Clinical Trials in Preclinical Alzheimer's Disease

Vamsi K. Ithapu

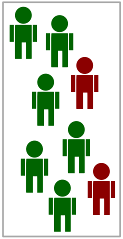
Wisconsin Alzheimer's Disease Research Center
University of Wisconsin-Madison

April 2, 2017

A Clinical Trial – *The work flow*

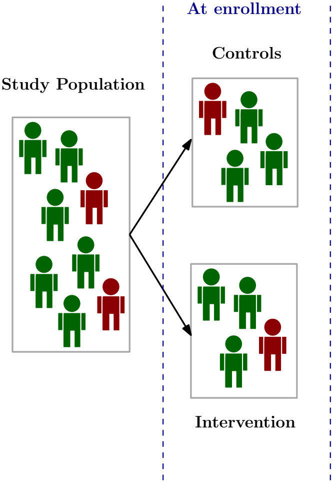
Randomized Controlled Trial

Study Population



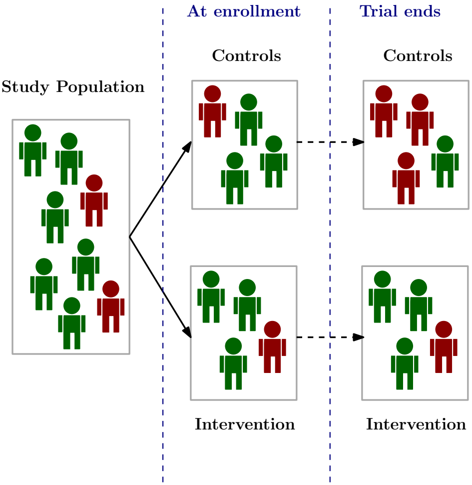
A Clinical Trial – *The work flow*

Randomized Controlled Trial



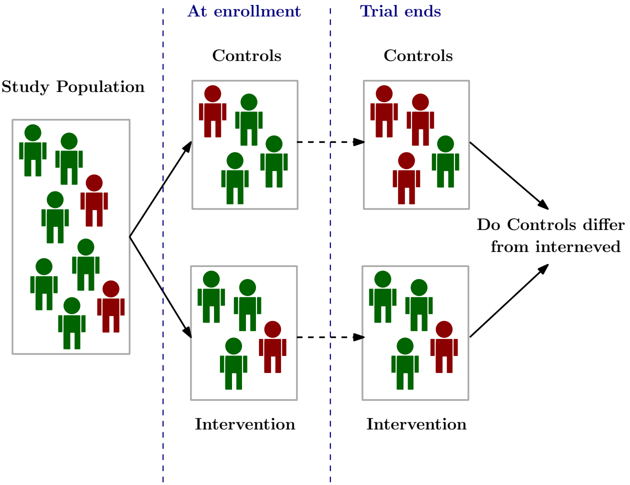
A Clinical Trial – *The work flow*

Randomized Controlled Trial



A Clinical Trial – *The work flow*

Randomized Controlled Trial



Setting up a clinical trial – *My work*

Who is participating in the trial?

Clinical Trial Enrichment

How to differentiate control from intervened?

Trial Outcome Design

Setting up a clinical trial – *My work*

Who is participating in the trial?

Clinical Trial Enrichment

How to differentiate control from intervened?

Trial Outcome Design

Setting up a clinical trial – *My work*

Who is participating in the trial?

Clinical Trial Enrichment

How to differentiate control from intervened?

Trial Outcome Design

trials aimed for **Alzheimer's Disease**

Alzheimer's Disease

Destroys memory and cognition

Irreversible. Strongest risk factor is age

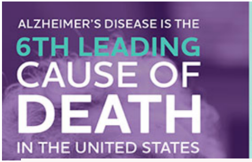
Diagnosis $\leftarrow$ { Age, Family History, Cognitive/Neuropsych/Physical Exams, Brain Scans }

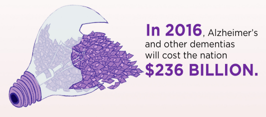
Alzheimer's Disease

Destroys memory and cognition

Irreversible. Strongest risk factor is age

Diagnosis ← { Age, Family History, Cognitive/Neuropsych/Physical Exams, Brain Scans }





Alzheimer's is a growing epidemic.

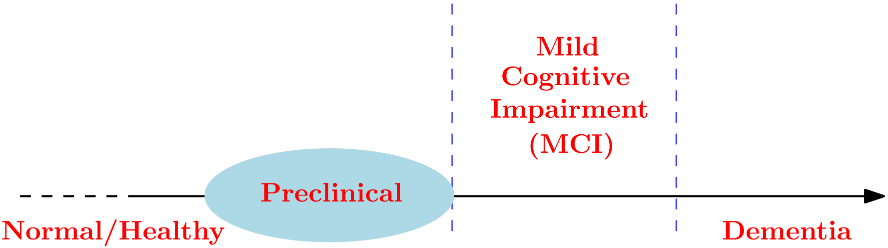
- More than 5 million Americans now have Alzheimer's disease. By 2050, nearly 14 million (13.8 million) Americans over age 65 could be living with the disease, unless scientists develop new approaches to prevent or cure it.[1]

Alzheimer's Disease

Destroys memory and cognition

Irreversible. Strongest risk factor is age

Diagnosis $\leftarrow$ { Age, Family History, Cognitive/Neuropsych/Physical Exams, Brain Scans }



Landscape of AD Clinical Trials

ClinicalTrials.gov lists 485 recruiting studies

225 in US; 147 in Europe;
68 are in Phase III and IV

Landscape of AD Clinical Trials

ClinicalTrials.gov lists 485 recruiting studies

225 in US; 147 in Europe;
68 are in Phase III and IV

Very little success . . . more than 550 trials since 2002 (Cummings 2014)

Landscape of AD Clinical Trials

CLINICAL

225

68 a

Very little

mings 2014)

AD diagnosis itself is messy

→ Early diagnosis is **much harder**

→ CN vs. MCI ≈ 70%

Landscape of AD Clinical Trials

CLINICA[L]

225

68 a[...]

Very littl[e] [...]mings 2014)

AD diagnosis itself is messy

$\rightarrow$ Early diagnosis is **much harder**

$\rightarrow$ CN vs. MCI $\approx$ 70%

$<$ 20% of MCIs convert to AD

$\Longrightarrow$ 8 out of 10 trial subjects are **not-eligible!!**

. . . but there is light

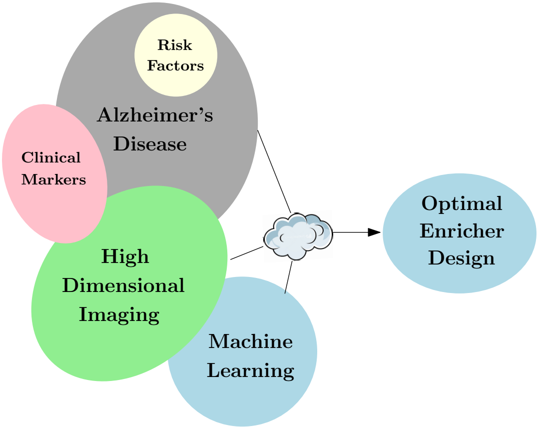
Imaging to the rescue

Cognitive decline *follows* atypical brain scans

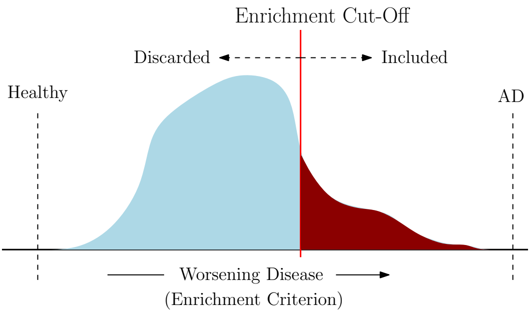
. . . but there is light

Imaging to the rescue

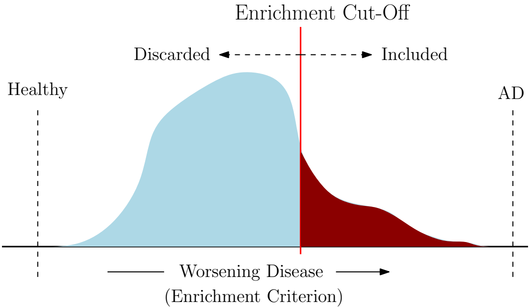
Cognitive decline *follows* atypical brain scans



Population enrichment



Population enrichment



Good enrichment criterion $\iff$ High correlation with disease

Practical enrichment criterion $\iff$ High *predictive* power

Designing a *good* enricher

Given some marker
δ : **Longitudinal change**
σ : **Pooled Variance**

Designing a *good* enricher

Given some marker

δ : **Longitudinal change**

σ : **Pooled Variance**

Optimal Enricher

Small σ

$+$

Large δ

Designing a *good* enricher

Given some marker
δ : **Longitudinal change**
σ : **Pooled Variance**

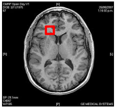
Optimal Enricher

Low-Variance
$+$
Un-Biased

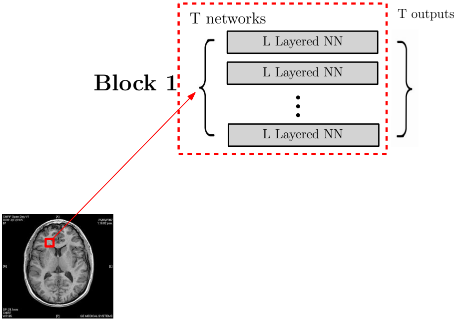
$\longrightarrow$

An Ensemble
$+$
Neural Networks

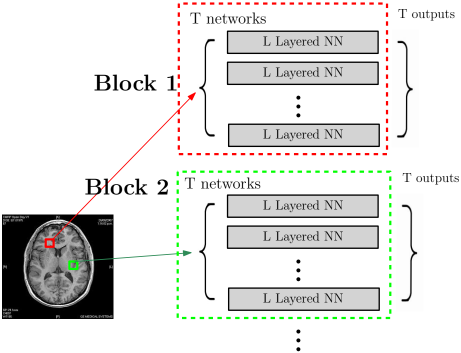
Randomized deep networks for enrichment



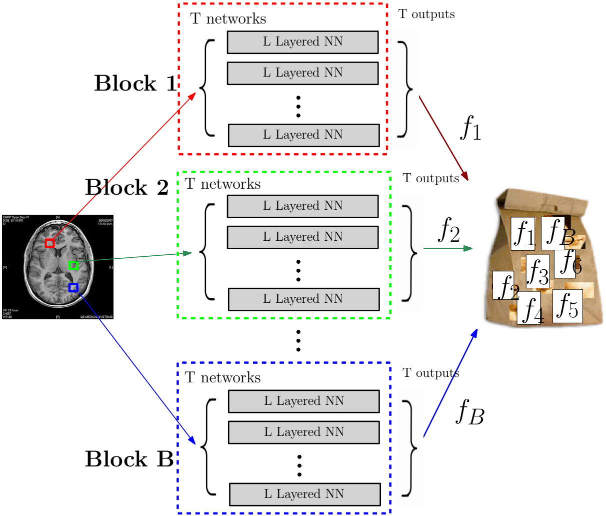
Randomized deep networks for enrichment



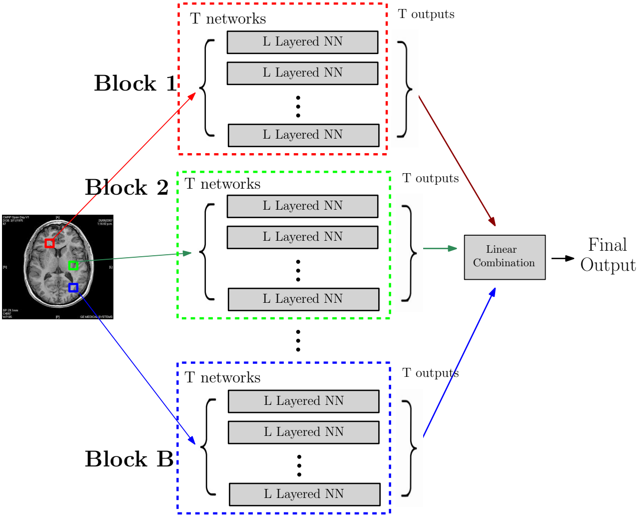
Randomized deep networks for enrichment



Randomized deep networks for enrichment



Randomized deep networks for enrichment



Randomized deep network *Markers* – rDm

Training baseline rDm

Inputs

→ MRI and PET Images

Labels

→ AD – 0, healthy – 1

Randomized deep network *Markers* – rDm

Training baseline rDm

Inputs

→ MRI and PET Images

Labels

→ AD – 0, healthy – 1

rDm at test time

Predict on MCI

Randomized deep network *Markers* – rDm

Training baseline rDm

Inputs
→ MRI and PET Images
Labels
→ AD – 0, healthy – 1

rDm at test time

Predict on MCI

Choose a cut-off $t \in [0, 1]$ & filter out subjects with rDm prediction $> t$

Predictive power of baseline rDm

Baseline rDm versus *change* (12 and 24 months) in outcomes

Spearman correlation coefficient (and *p*-value)

Marker	12m	24m
MMSE	0.2123, $p = 0.0008$	0.3311, $p = 0.0003$
ADAS	0.2139, $p = 0.0007$	-0.5300, $p < 10^{-4}$
MOCA	0.0568, $p > 0.1$	0.5952, $p = 10^{-4}$
RAVLT	0.1285, $p = 0.04$	0.5702, $p = 0.0008$
PsyMEM	0.2811, $p < 10^{-4}$	0.4207, $p = 0.001$
HippoVol	0.3262, $p \ll 10^{-4}$	0.4744, $p \ll 10^{-4}$
CDR-SB	-0.3643, $p \ll 10^{-4}$	-0.5344, $p \ll 10^{-4}$
DXConv[1]	> 20, $p \ll 10^{-4}$	> 20, $p \ll 10^{-4}$

[1] ANOVA test results are reported since this variable is categorical

Predictive power of baseline rDm

Baseline rDm versus *change* (12 and 24 months) in outcomes

Spearman correlation coefficient (and *p*-value)

Marker	12m	24m
MMSE	0.2123, $p = 0.0008$	0.3311, $p = 0.0003$
ADAS	0.2139, $p = 0.0007$	–0.5300, $p < 10^{-4}$
MOCA	0.0568, $p > 0.1$	0.5952, $p = 10^{-4}$
RAVLT	0.1285, $p = 0.04$	0.5702, $p = 0.0008$
PsyMEM	0.2811, $p < 10^{-4}$	0.4207, $p = 0.001$
HippoVol	0.3262, $p \ll 10^{-4}$	0.4744, $p \ll 10^{-4}$
CDR-SB	–0.3643, $p \ll 10^{-4}$	–0.5344, $p \ll 10^{-4}$
DXConv[1]	> 20, $p \ll 10^{-4}$	> 20, $p \ll 10^{-4}$

[1] ANOVA test results are reported since this variable is categorical

Predictive power of baseline rDm

Baseline rDm versus *change* (12 and 24 months) in outcomes

Spearman correlation coefficient (and *p*-value)

Marker	12m	24m
MMSE	$\boxed{0.2123, p = 0.0008}$	0.3311, $p = 0.0003$
ADAS	$\boxed{0.2139, p = 0.0007}$	–0.5300, $p < 10^{-4}$
MOCA	0.0568, $p > 0.1$	0.5952, $p = 10^{-4}$
RAVLT	0.1285, $p = 0.04$	0.5702, $p = 0.0008$
PsyMEM	$\boxed{0.2811, p < 10^{-4}}$	0.4207, $p = 0.001$
HippoVol	$\boxed{0.3262, p \ll 10^{-4}}$	0.4744, $p \ll 10^{-4}$
CDR-SB	$\boxed{-0.3643, p \ll 10^{-4}}$	–0.5344, $p \ll 10^{-4}$
DXConv[1]	$\boxed{> 20, p \ll 10^{-4}}$	> 20, $p \ll 10^{-4}$

[1] ANOVA test results are reported since this variable is categorical

Predictive power of baseline rDm

Baseline rDm versus *change* (12 and 24 months) in outcomes

Spearman correlation coefficient (and *p*-value)

Marker	12m	24m
MMSE	0.2123, $p = 0.0008$	0.3311, $p = 0.0003$
ADAS	0.2139, $p = 0.0007$	–0.5300, $p < 10^{-4}$
MOCA	0.0568, $p > 0.1$	0.5952, $p = 10^{-4}$
RAVLT	0.1285, $p = 0.04$	0.5702, $p = 0.0008$
PsyMEM	0.2811, $p < 10^{-4}$	0.4207, $p = 0.001$
HippoVol	0.3262, $p \ll 10^{-4}$	0.4744, $p \ll 10^{-4}$
CDR-SB	–0.3643, $p \ll 10^{-4}$	–0.5344, $p \ll 10^{-4}$
DXConv[1]	> 20, $p \ll 10^{-4}$	> 20, $p \ll 10^{-4}$

[1] ANOVA test results are reported since this variable is categorical

Predictive power of baseline rDm

Baseline rDm versus *change* (12 and 24 months) in outcomes

Spearman correlation coefficient (and *p*-value)

Marker	12m	24m
MMSE	0.2123, $p = 0.0008$	0.3311, $p = 0.0003$
ADAS	0.2139, $p = 0.0007$	–0.5300, $p < 10^{-4}$
MOCA	0.0568, $p > 0.1$	0.5952, $p = 10^{-4}$
RA[illegible]	[illegible]	[illegible]08
Psy[illegible]	[illegible]	[illegible]1
HippoVol	0.3262, $p \ll 10^{-4}$	0.4744, $p \ll 10^{-4}$
CDR-SB	–0.3643, $p \ll 10^{-4}$	–0.5344, $p \ll 10^{-4}$
DXConv[1]	> 20, $p \ll 10^{-4}$	> 20, $p \ll 10^{-4}$

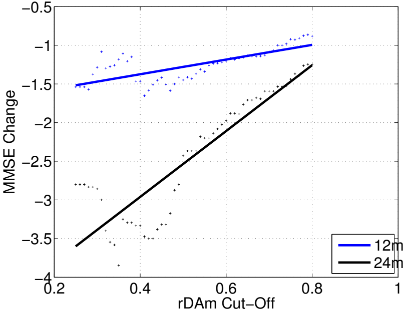
Very strong correlations across all markers

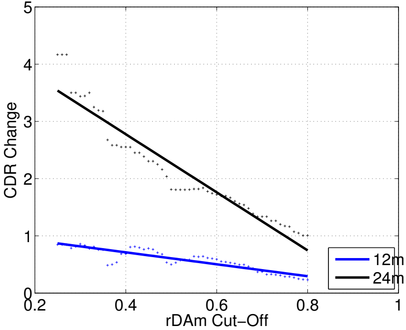
[1] ANOVA test results are reported since this variable is categorical

Predictive power of baseline rDm

Mean longitudinal change in MMSE & CDR

Important trial outcomes





Baseline rDm vs. alternate enrichers

Sample sizes per arm

80% power, 25% improvement from treatment

Sample enricher	Outcome measure							
	MMSE	ADAS	MOCA	RAVLT	PsyMEM	HipVol	CDR-SB	DxConv
HipVol	500	>2000	1005	1606	1009	>2000	389	420
FDG	384	1954	579	>2000	832	752	415	371
AV45	224	>2000	875	>2000	826	698	382	443
FAH	296	>2000	705	>2000	826	722	397	402
MKLm [2]	228	874	827	896	487	877	295	284
rDm	200	775	449	591	420	543	281	230

[2] MKLm is the current state-of-the-art based on SVMs

Baseline rDm vs. alternate enrichers

Sample sizes per arm

80% power, 25% improvement from treatment

Sample enricher	Outcome measure							
	MMSE	ADAS	MOCA	RAVLT	PsyMEM	HipVol	CDR-SB	DxConv
HipVol	500	>2000	1005	1606	1009	>2000	389	420
FDG	384	1954	579	>2000	832	752	415	371
AV45	224	>2000	875	>2000	826	698	382	443
FAH	296	>2000	705	>2000	826	722	397	402
MKLm [2]	228	874	827	896	487	877	295	284
rDm	200	775	449	591	420	543	281	230

[2] MKLm is the current state-of-the-art based on SVMs

Baseline rDm vs. alternate enrichers

Sample sizes per arm

80% power, 25% improvement from treatment

Sample enricher	Outcome measure MMSE	ADAS	MOCA	RAVLT	PsyMEM	HipVol	CDR-SB	DxConv
HipVol	500	>2000	1005	1606	1009	>2000	389	420
FDG	384	1954	579	>2000	832	752	415	371
AV45	224	>2000	875	>2000	826	698	382	443
FAH	296	>2000	705	>2000	826	722	397	402
MKLm [3]	228	874	827	896	487	877	295	284
rDm	**200**	**775**	**449**	**591**	**420**	**543**	**281**	**230**

rDm has smallest estimates across all outcomes

[3] MKLm is the current state-of-the-art based on SVMs

Baseline rDm vs. alternate enrichers

Sample sizes per arm

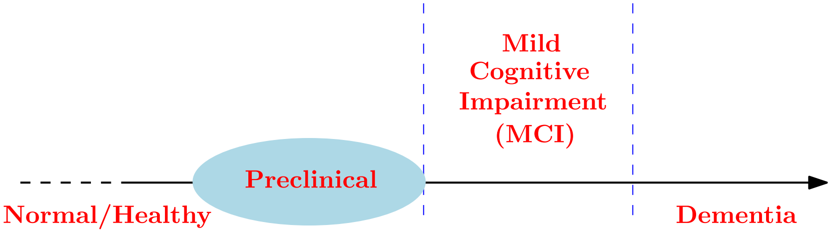
80% power, 25% improvement from treatment

Sample enricher	MMSE	ADAS	MOCA	RAVLT	PsyMEM	HipVol	CDR-SB	DxConv
HipVol								420
FDG								371
AV45								443
FAH	296	>2000	705	>2000	826	722	397	402
MKLm [3]	228	874	827	896	487	877	295	284
rDm	**200**	**775**	**449**	**591**	**420**	**543**	**281**	**230**

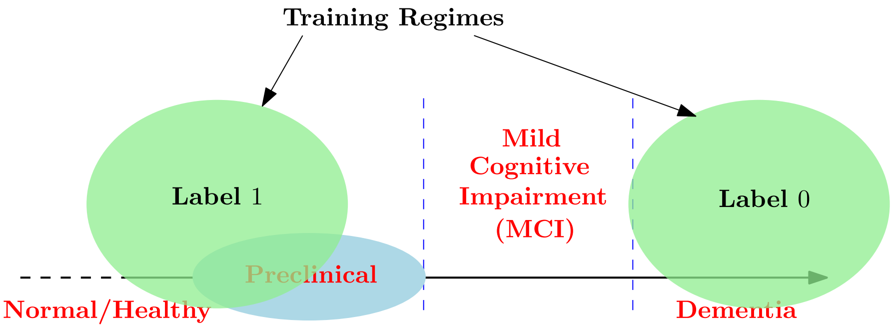
Baseline rDm can detect weak treatment effects

[3] MKLm is the current state-of-the-art based on SVMs

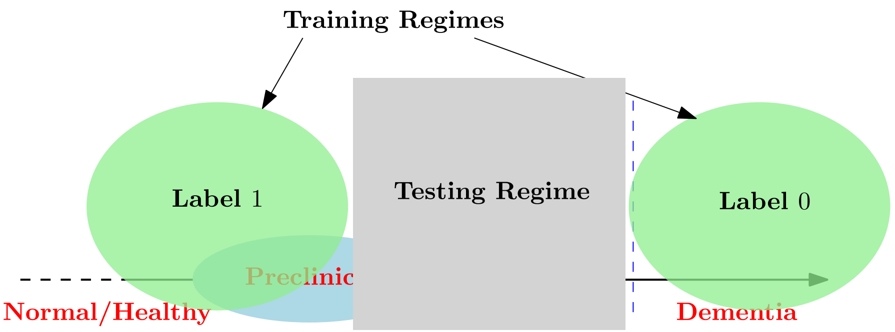
The proposed enricher – AD Spectrum



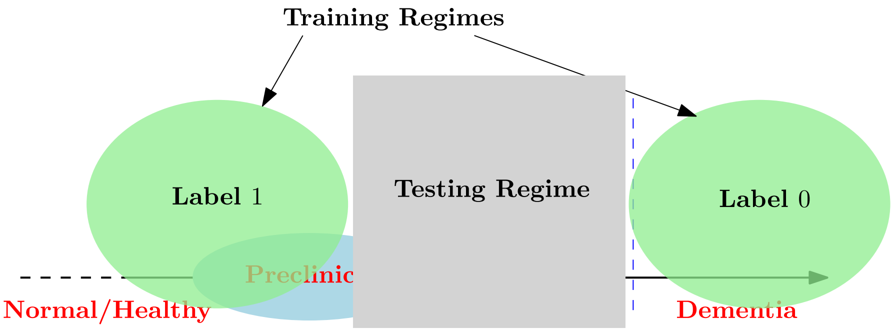
The proposed enricher – AD Spectrum



The proposed enricher – AD Spectrum

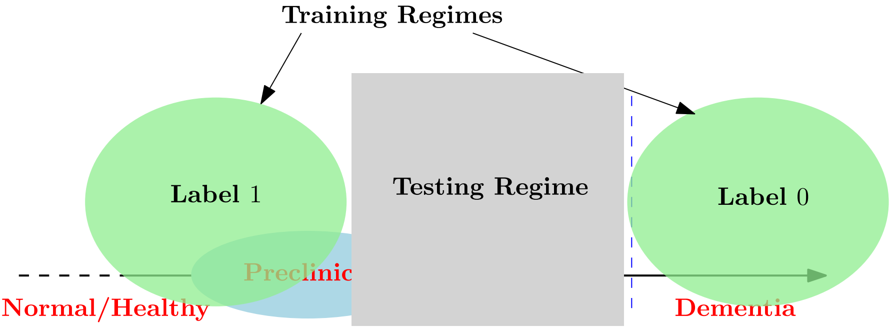


The proposed enricher – AD Spectrum



Two Issues

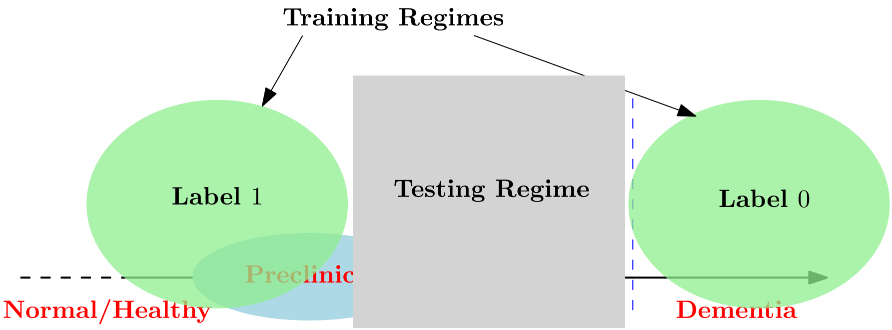
The proposed enricher – AD Spectrum



Two Issues

Disease spectrum is continuous
$\rightarrow$ Labels *somewhat* artificial – Supervised models are sensitive

The proposed enricher – AD Spectrum



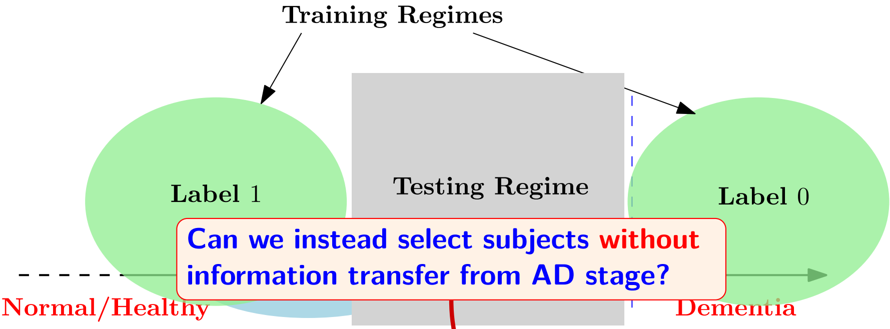
Two Issues

Disease spectrum is continuous

$\rightarrow$ Labels *somewhat* artificial – Supervised models are sensitive

Bio-markers interact differently in preclinical vs. AD

The proposed enricher – AD Spectrum



Can we instead select subjects without information transfer from AD stage?

Two Issues

Disease spectrum is continuous
$\rightarrow$ Labels *somewhat* artificial – Supervised models are sensitive
Bio-markers interact differently in preclinical vs. AD

An alternate View – *Sampling*

Select atypical subjects

The more *unique* a subject is

$\rightarrow$ ... the more information they contribute to trial

Some typical points also needed

An alternate View – *Sampling*

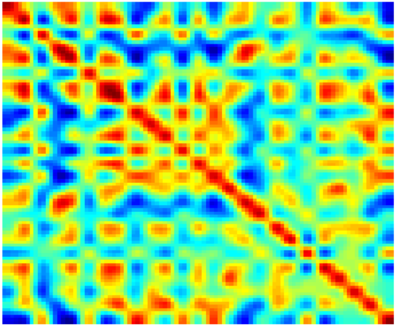
Select atypical subjects

The more *unique* a subject is

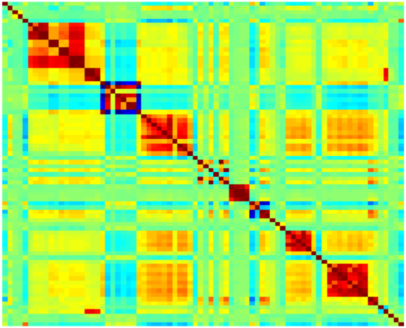
$\rightarrow$ ... the more information they contribute to trial

Some typical points also needed

AD Imaging Features



AD Clinical/Neuropsych Scores



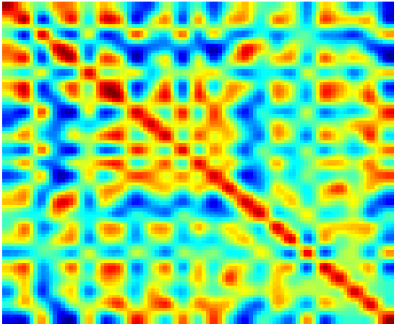
An alternate View – *Sampling*

Select atypical subjects

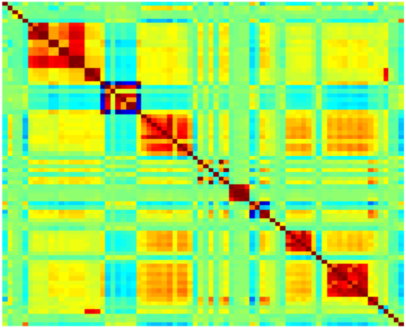
The more *unique* a subject is
$\rightarrow$ ... the more information they contribute to trial
Some typical points also needed

Very Rich Block (Hierarchical) Structure

AD Imaging Features

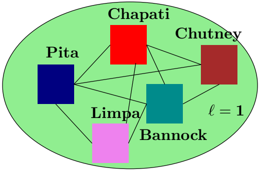


AD Clinical/Neuropsych Scores

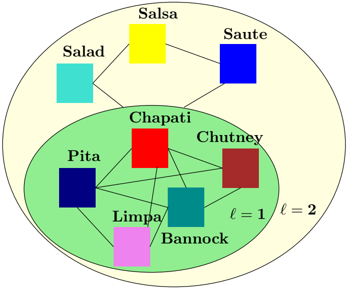


Multiresolution Matrix Factorization

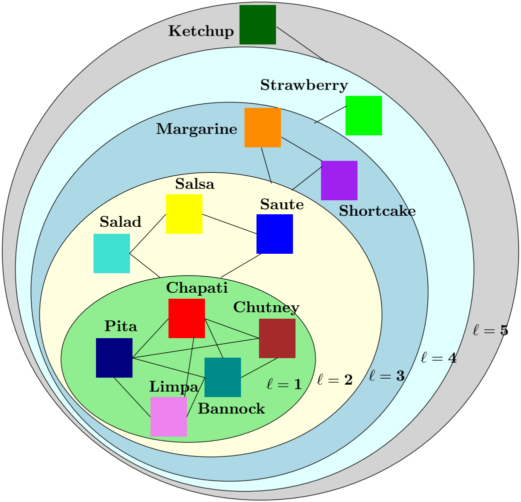
Multiresolution Matrix Factorization



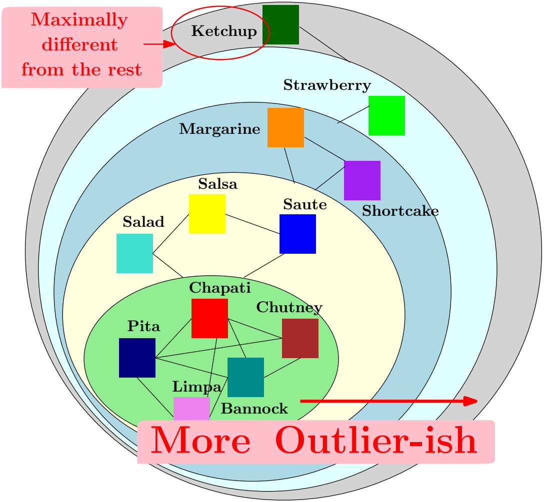
Multiresolution Matrix Factorization



Multiresolution Matrix Factorization



Multiresolution Matrix Factorization



Thank you . . . Questions?

I., V. Singh, O. C. Okonkwo, S. C. Johnson, A predictive multi-modal imaging marker for designing efficient and robust AD clinical trials, Clinical Trials on Alzheimer's Disease (CTAD), 2014

I., V. Singh, S. C. Johnson, Randomized deep learning methods for clinical trial enrichment and design in Alzheimer's disease, Deep Learning for Medical Image Analysis (1st Edition) ISBN: 9780128104088; Chapter 15

I., V. Singh, O. C. Okonkwo, R. J. Chappell, N. M. Dowling, S. C. Johnson, Imaging based enrichment criteria using deep learning algorithms for efficient clinical trials in MCI, Alzheimer's and Dementia, 2015

I., R. Kondor, S. C. Johnson, V. Singh, The Incremental Multiresolution Matrix Factorization Algorithm, Computer Vision and Pattern Recognition (CVPR), 2017

http://pages.cs.wisc.edu/~vamsi/publications.html

Acknowledgements:

Vikas Singh, **Sterling Johnson**, Chris Hinrichs, Risi Kondor, Barbara Bendlin, Ozioma Okonkwa

NIH AG040396, NSF CAREER 1252725, NSF CCF 1320755, UW ADRC AG033514